Defining Gravity: Effective Field Theory, Entanglement, and Cosmology

Thesis by

Grant Newton Remmen

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

CALIFORNIA INSTITUTE OF TECHNOLOGY
Pasadena, California

2017
(Defended May 17, 2017)

Grant Newton Remmen
ORCID: 0000-0001-6569-8866

To my family.

Acknowledgments

There are many people whose support and encouragement have made the research that led to this thesis possible. Isaac Newton, the father of gravitational physics, once said, "If I have seen further, it is by standing on the shoulders of giants." I am grateful for the giants, both of physics and of character, in my own life who have enabled my studies and with whom I have had the opportunity to try, in my own way, to extend humanity's knowledge of the universe.

I would like to first thank my two amazing advisors, Cliff Cheung and Sean Carroll, for their invaluable support, guidance, advice, and insights during my years at Caltech. I am incredibly grateful for the knowledge and perspective I gained from them both on all aspects of the practice of theoretical physics. Any student would be very lucky to have a single advisor as great as either Sean or Cliff, so I feel extremely fortunate to have worked with both of them. Cliff's and Sean's energy and dedication as my research collaborators have made the process of discovery in theoretical physics an always enjoyable and exhilarating undertaking.

I am appreciative of the other members of my thesis committee, Mark Wise and Alan Weinstein. Mark's physics wisdom and humor have both been a welcome addition to my experience at Caltech and Alan's interesting questions and perspective have always proven useful.

This thesis is based on a number of papers I have written during my time at Caltech with several groups of collaborators. In addition to Cliff and Sean, I particularly thank Jason Pollack and Ning Bao, as well as Brando Bellazzini, Charles Cao, Aidan Chatwin-Davies, and Nick Hunter-Jones, for productive

collaborations. I also appreciate the many stimulating physics discussions relating to work in this thesis that I have had with Nima Arkani-Hamed, Zvi Bern, Stanley Deser, Dan Harlow, Stefan Leichenauer, Matt Reece, and many others both at Caltech and elsewhere.

I would also like to acknowledge the Hertz Foundation and the National Science Foundation. The Hertz Graduate Fellowship and the NSF Graduate Research Fellowship both represented a vote of confidence in my research ability and its value to the nation and the world, for which I am very grateful. Furthermore, I am happy to have had the opportunity for thought-provoking conversations with the fascinating researchers I have been able to meet through the Hertz Fellowship Community.

Finally, I am tremendously thankful to my parents, Larry and Kellie Remmen, and my brother, Cole. As a child, my mother instilled in me a drive to discover, taking me to the local library every week and making every day an adventure; my father encouraged my learning as well, helping me build and launch model rockets and nurturing my love of mathematics. My parents have always supported me in all my endeavors, from spelling bees to musicals, in whatever I wanted to do. Though none of them are physicists, my family is always interested in what I am working on; they have bravely listened to more of my presentations than they probably care to remember, but are always encouraging. My brother Cole—my best friend—is always there for me and I have thoroughly enjoyed our music projects outside of science (including Caltech's premiere of *Boldly Go!*). I particularly thank Cole for finding Kip Thorne's *Black Holes and Time Warps* at the local library used-book sale when I was eleven years old, which our mother bought for me for the fantastic sum of one dollar! This book inspired my already burgeoning interest in fundamental physics and showed me the beauty of general relativity.

This work was supported by the Hertz Graduate Fellowship and the NSF Graduate Research Fellowship under Grant No. DGE-1144469.

Abstract

Many of the most exciting open problems in high-energy physics are related to the behavior and ultimate nature of gravity and spacetime. In this dissertation, several categories of new results in quantum and classical gravity are presented, with applications to our understanding of both quantum field theory and cosmology.

A fundamental open question in quantum field theory is related to ultraviolet completion: Which low-energy effective field theories can be consistently combined with quantum gravity? A celebrated example of the swampland program—the investigation of this question—is the weak gravity conjecture, which mandates, for a $U(1)$ gauge field coupled consistently to gravity, the existence of a state with charge-to-mass ratio greater than unity. In this thesis, we demonstrate the tension between the weak gravity conjecture and the naturalness principle in quantum field theory, generalize the weak gravity conjecture to multiple gauge fields, and exhibit a model in which the weak gravity conjecture solves the standard model hierarchy problem. Next, we demonstrate that gravitational effective field theories can be constrained by infrared physics principles alone, namely, analyticity, unitarity, and causality. In particular, we derive bounds related to the weak gravity conjecture by placing such infrared constraints on higher-dimension operators in a photon-graviton effective theory. We furthermore place bounds on higher-curvature corrections to the Einstein equations, first using analyticity of graviton scattering amplitudes and later using unitarity of an arbitrary tree-level completion, as well as constrain the couplings in models of massive gravity. Completing our treatment of perturba-

tive quantum gravity, outside of the swampland program, we also reformulate graviton perturbation theory itself, finding a field redefinition and gauge-fixing of the Einstein-Hilbert action that drastically simplifies the Feynman diagram expansion. Furthermore, our reformulation also exhibits a hidden symmetry of general relativity that corresponds to the double copy relations equating gravity amplitudes to sums of squares of gluon amplitudes in Yang-Mills theory, a surprising correspondence that yields insights into the structure of quantum field theories.

Moving beyond perturbation theory into nonperturbative questions in quantum gravity, we consider the deep relation between spacetime geometry and properties of the quantum state. In the context of holography and the anti-de Sitter/conformal field theory correspondence, we test the proposed ER=EPR correspondence equating quantum entanglement with wormholes in spacetime. In particular, we demonstrate that the no-cloning theorem in quantum mechanics and the no-go theorem for topology change of spacetime are dual under the ER=EPR correspondence. Furthermore, we prove that the presence of a wormhole is not an observable in quantum gravity, rescuing ER=EPR from potential violation of linearity of quantum mechanics. Excitingly, we also prove a new area theorem within classical general relativity for arbitrary dynamics of two collections of wormholes and black holes; this area theorem is the ER=EPR analogue of entanglement conservation. We next turn our attention to the emergence of spacetime itself, placing consistency conditions on the proposed correspondence between anti-de Sitter space and the Multiscale Entanglement Renormalization Ansatz, a special tensor network that constitutes a computational tool for finding the ground state of certain quantum systems. Further examining the role of quantum entanglement entropy in the emergence of general relativity, we ask whether there is a consistent microscopic formulation

of the entropy in theories of entropic gravity; we find that our results weaken equation-of-state proposals for entropic gravity while strengthening those more akin to holography, guiding future investigation of theories of emergent gravity.

Finally, we examine the consequences of the Hamiltonian constraint in classical gravity for the early universe. The Hamiltonian constraint allows for the Liouville measure on the phase space of cosmological parameters for homogeneous, isotropic universes to be converted into a probability distribution on trajectories, or equivalently, on initial conditions. However, this measure diverges on the set of spacetimes that are spatially flat, like the observable universe. In this thesis, we derive the unique, classical, Hamiltonian-conserved measure for the subset of flat universes. This result allows for distinction between different models of cosmic inflation with similar observable predictions; for example, we find that the measure favors models of large-scale inflation, as such potentials more naturally produce the number of e-folds necessary to match cosmological observations.

Published Content and Contributions

Below are the references to the published articles [1–13] that make up Chaps. 2 through 14 of this thesis, respectively. I was a primary author for all of these papers.

[1] C. Cheung and G. N. Remmen, "Naturalness and the Weak Gravity Conjecture," *Phys. Rev. Lett.* **113** (2014) 051601, arXiv:1402.2287 [hep-ph]

[2] C. Cheung and G. N. Remmen, "Infrared Consistency and the Weak Gravity Conjecture," *JHEP* **1412** (2014) 087, arXiv:1407.7865 [hep-th]

[3] B. Bellazzini, C. Cheung, and G. N. Remmen, "Quantum Gravity Constraints from Unitarity and Analyticity," *Phys. Rev.* **D93** (2016) 064076, arXiv:1509.00851 [hep-th]

[4] C. Cheung and G. N. Remmen, "Positive Signs in Massive Gravity," *JHEP* **04** (2016) 002, arXiv:1601.04068 [hep-th]

[5] C. Cheung and G. N. Remmen, "Positivity of Curvature-Squared Corrections in Gravity," *Phys. Rev. Lett.* **118** (2017) 051601, arXiv:1608.02942 [hep-th]

[6] C. Cheung and G. N. Remmen, "Twofold Symmetries of the Pure Gravity Action," *JHEP* **01** (2017) 104, arXiv:1612.03927 [hep-th]

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

[8] N. Bao, J. Pollack, and G. N. Remmen, "Wormhole and Entanglement (Non-)Detection in the ER=EPR Correspondence," *JHEP* **11** (2015) 126, arXiv:1509.05426 [hep-th]

[9] G. N. Remmen, N. Bao, and J. Pollack, "Entanglement Conservation, ER=EPR, and a New Classical Area Theorem for Wormholes," *JHEP* **07** (2016) 048, arXiv:1604.08217 [hep-th]

[10] N. Bao, C. Cao, S. M. Carroll, A. Chatwin-Davies, N. Hunter-Jones, J. Pollack, and G. N. Remmen, "Consistency Conditions for an AdS Multiscale Entanglement Renormalization Ansatz Correspondence," *Phys. Rev.* **D91** (2015) 125036, arXiv:1504.06632 [hep-th]

[11] S. M. Carroll and G. N. Remmen, "What is the Entropy in Entropic Gravity?," *Phys. Rev.* **D93** (2016) 124052, arXiv:1601.07558 [hep-th]

[12] G. N. Remmen and S. M. Carroll, "Attractor Solutions in Scalar-Field Cosmology," *Phys.Rev.* **D88** (2013) 083518, arXiv:1309.2611 [gr-qc]

[13] G. N. Remmen and S. M. Carroll, "How Many e-Folds Should We Expect from High-Scale Inflation?," *Phys. Rev.* **D90** (2014) 063517, arXiv:1405.5538 [hep-th]

Contents

List of Figures

Chapter 1
Introduction

The early years of the twentieth century bore witness to three revolutions in physics: *special relativity*, *general relativity*, and *quantum mechanics*. Special and general relativity indelibly reshaped our classical understanding of space and time; the special theory removed the immutable distinction between the two, mixing them for different frames of reference into a single entity, *spacetime*, and the general theory further promoted spacetime from merely a stage on which physics is performed to an active player in the dynamics of relativistic gravitation. Quantum mechanics, moreover, swept away the classical clockwork universe described by precise particle data and trajectories in phase space, replacing the description of the world with one described fundamentally by probabilities given by wave functions evolving within Hilbert space. The ramifications and interplay among the triple revolution of special and general relativity and quantum theory have, to a large extent, guided and determined the progress of physics for the last century, which has seen extraordinary and unprecedented advancement of humanity's knowledge and understanding of the world. In this thesis, we investigate the nature of gravity and spacetime and their implications for the universe, an undertaking that will require the use of various combinations of ideas from all three of these branches of twentieth-century physics, with exciting and sometimes surprising results.

The reconciliation of quantum mechanics with special relativity resulted in the development of *quantum field theory*, constituting arguably the greatest success story in the history of science. The formulation of quantum field

theory over the middle decades of the twentieth century occurred in tandem with progress in nuclear physics that led ultimately to the development of the standard model of particle physics. This wildly successful model, whose experimental verification was completed in 2012 with the discovery of the Higgs boson at the Large Hadron Collider [14, 15], describes with incredible precision virtually all observed non-gravitational physical processes, at length scales ranging from the astrophysical down to around 10^{-19} m and perhaps even smaller. The standard model encompasses the electromagnetic, weak, and strong nuclear forces and their interactions with matter, describing phenomena ranging from light, to nuclear physics, to chemistry and, when combined with classical gravity, even describes the workings of stars and predicts the abundances of elements produced at the beginning of the universe in the big bang. Of course, there are other known processes and phenomena outside the reach of the standard model, including dark matter, dark energy, neutrino masses, the dynamics of the big bang, and the ultimate nature of spacetime, some of which will be discussed and investigated in this thesis.

Just as important as its success in the standard model, however, quantum field theory's power is far greater. More than merely a single model, quantum field theory is a framework for constructing further physical theories to describe our universe or others. It is a prescription for model-building, for organizing physical laws. One writes down an action as an integral over a Lagrangian density, a functional over spacetime defined in terms of various fields, then promotes the fields to Hermitian operators obeying the canonical commutation relations, and one has a quantum field theory. If the theory is perturbative, one can immediately begin calculating useful quantities, such as *scattering amplitudes*, which describe the interaction probabilities of particles, using the traditional Feynman diagram approach [16] or, alternatively, using modern

amplitudes methods that make use of the principles of unitarity and analyticity [17, 18].

A central feature in our understanding of quantum field theory is the concept of an *effective field theory*. Rather than forming a description of a set of phenomena at all energy scales, down to arbitrarily small distances, an effective field theory is a quantum field theoretic description that is valid only over a certain range of energy scales [16, 19, 20]. Just as the classical description of a wave in water breaks down at a length scale of approximately the size of a water molecule, so, too, does a given effective field theory break down at some energy scale, at which it is replaced by a new quantum field theory, with additional quantum fields describing the new degrees of freedom in the microscopic theory. Indeed, quantum field theory is sufficiently powerful that effective field theories predict their own demise: scattering amplitudes diverge at large energies, with the characteristic energy scale at which the divergence occurs giving the cutoff of the effective field theory. Classic examples of effective field theories occur as approximations to the standard model at low energies. The four-fermi interaction among electrons, neutrinos, and quarks provides a good description of the weak nuclear force at low energies, but predicts violations of unitarity around the weak scale (246 GeV); as expected, new degrees of freedom, the W and Z bosons, enter and unitarize the theory. Similarly, one can compute the scattering of W bosons without the Higgs boson, in which case one also finds unitarity violation around the weak scale; the Higgs boson enters the model to save the theory, resulting in the standard model. Indeed, the standard model itself will ultimately prove to be an effective field theory. One can view high-energy physics experiments at the Large Hadron Collider and elsewhere as constituting the search for the energy scale at which this effective theory breaks down.

While special relativity and quantum mechanics proved amenable to unification, general relativity is another matter. It is sometimes claimed that general relativity and quantum mechanics are fundamentally incompatible, but from the correct perspective this is not so. One can treat general relativity perturbatively as an effective field theory [21, 22]. The Lagrangian is written in terms of a spin-two quantum field, the *graviton*, defined on a background spacetime and with interactions prescribed by the Einstein-Hilbert action, which is proportional to the spacetime curvature and which in its classical form gives the action formulation for the Einstein equation of general relativity. As an effective field theory, perturbative quantum gravity predicts its own breakdown at the Planck scale, $\sim 10^{19}$ GeV, or around 10^{-35} m. This is a fantastically high-energy regime, orders of magnitude outside the bounds of current experiment, but excitingly, not necessarily outside of our theoretical ability to describe physics. This energy scale can be extracted, for example, by computing graviton scattering amplitudes, a calculational exercise that will be of importance later in this thesis.

As a result of this predicted breakdown of perturbative quantum gravity at the Planck scale, over the last several decades enormous efforts have been undertaken to find the ultraviolet completion of gravity, just as the W and Z bosons are the ultraviolet completion of the four-fermi theory. In the case of gravity, however, the task of finding a high-energy completion has proved to be not nearly so straightforward. The best, and arguably unique, candidate for an ultraviolet-complete quantum theory of gravity comes from *string theory* [23, 24], in which pointlike quantum fields are replaced by extended objects. While the present discussion does not permit space for a full introduction to string theory, we briefly note that, like quantum field theory, string theory is a framework that describes a very large number of low-energy effective field theories; in the case of

string theory, the ultraviolet completion for each consistent effective field theory contains an infinite tower of massive states, which prevent graviton scattering amplitudes from diverging at high energies [25, 26]. The large collection of effective field theories spanned by stringy completions is known as the *landscape* [27–30] and it is believed that our standard model occupies a point in this wide array of possible sets of low-energy laws of physics.

A crucial question that the quantum field theorist asks of string theory is whether all effective field theories lie in the landscape. That is, can all possible sets of low-energy laws of physics be consistently incorporated into quantum gravity? Perhaps surprisingly, the answer is that they cannot [31]. Some apparently theoretically healthy effective field theories at low energies cannot be consistently combined with quantum gravity and thus are outside the landscape; they are said to lie in the *swampland* [27, 29]. A crisis immediately arises: How can one characterize the set of effective field theories allowed by quantum gravity and thus distinguish the landscape from the swampland? This *swampland program*, the focus of Chaps. 2 through 6 of this thesis, is of fundamental importance to our understanding of quantum field theories and furthermore is crucial for restricting model-building to achieve consistency with quantum gravity [1–5].

The most celebrated result of the swampland program is the *weak gravity conjecture* [30], which forms a constraint on the charges, under a $U(1)$ gauge field, of matter coupled consistently to quantum gravity. In essence, the weak gravity conjecture requires the existence of a particle in the spectrum of the theory for which the electromagnetic charge is greater than the gravitational charge (the mass) in Planck units. The weak gravity conjecture was initially argued for using examples from string theory and thought experiments involving black holes [30], particularly the requirement that charged black holes must all

be able to quantum mechanically decay into particles, in processes involving Hawking radiation and pair-production of charged states. This requirement of black hole decay, which is motivated by reasons we will discuss, points the way toward a generalization of the weak gravity conjecture. In Chap. 2, we generalize the weak gravity conjecture to models containing multiple Abelian gauge fields (i.e., multiple copies of electromagnetism), which are well motivated in the context of string theory. Surprisingly, we find that the weak gravity bounds become more stringent with the addition of extra forces. Furthermore, we demonstrate that the weak gravity conjecture can be in tension with the principle of *naturalness*, the generic requirement in quantum field theory that cancellations are not happenstance but reliant on the presence of symmetries. Indeed, we discover that it is possible for the weak gravity conjecture to falsify the naturalness principle and we use this fact to build a model that uses the weak gravity conjecture to solve the hierarchy problem in the standard model [32, 33], the mystery of the large ratio between the energy scale of the weak nuclear force and the Planck scale of quantum gravity.

Subsequently, in Chap. 3, we examine the motivation for the weak gravity conjecture from an entirely new perspective. As we noted, the original arguments for the weak gravity conjecture are dependent on assumptions about high-energy physics, namely, string theory and properties of black hole decay. However, swampland bounds can also arise from bottom-up, rather than top-down, reasoning [2–5, 31, 34–37]. In particular, the principles of causality (that is, no time machines), unitarity (consistency of quantum mechanics), and analyticity of scattering amplitudes (related to physical locality) form a powerful toolbox of low-energy physics postulates that will remain valid for any well-behaved ultraviolet completion. Thus, considerations of infrared physics can help constrain the ultraviolet. In Chap. 3, these principles are employed to

explore the possibility of proving the weak gravity conjecture in the infrared. By placing bounds on the coefficients of higher-dimension operators in an effective field theory of an interacting photon and graviton, we find that bounds very similar to the weak gravity conjecture can be derived.

In Chaps. 4 through 6, we move on from the consideration of the weak gravity conjecture to the more general question of placing infrared consistency bounds on gravitational effective field theories. Whatever the final theory of quantum gravity ultimately is found to be, Einstein's equations will receive corrections. At present, physicists can write down the possible form of these corrections, but their coefficients can only be computed after making particular assumptions about the ultraviolet, e.g., assuming a particular version of string theory. However, infrared consistency criteria can be applied to general gravitational effective field theories, allowing us to bound the quantum corrections to Einstein's equations in ways that will remain robust under arbitrary ultraviolet completions.

In Chap. 4, we focus on using the requirement of analyticity of scattering amplitudes to bound higher-curvature operators. While the Einstein-Hilbert action of general relativity is linear in the Riemann curvature tensor, we bound operators at quartic order in the Riemann tensor, by requiring that graviton scattering amplitudes are analytic functions of complex momenta and using physical criteria such as the optical theorem from quantum mechanics. Furthermore, we use this same mathematical technology to prove that the curvature-squared operator is inconsistent unless it is accompanied by new states appearing at the energy scale suggested by effective field theory reasoning, as is the case in string theory and as is required by causality [37]. These bounds help to constrain gravitational effective field theories, providing a useful diagnostic for candidate low-energy limits of quantum gravity.

Next, in Chap. 5, we continue with our exploration of analyticity bounds on modifications of gravity. In general relativity, the graviton is a massless spin-two particle that in four spacetime dimensions describes two degrees of freedom. However, it is a priori possible that the graviton has a small mass [38]; indeed, motivations for considering massive gravity include using a small graviton mass, of order the Hubble scale, to explain the observed small acceleration of the universe in the present epoch. A massive graviton in four dimensions describes five degrees of freedom—two tensors, two vectors, and one scalar—and is famously subject to various theoretical pathologies [39–42]. However, a recently discovered version of the theory, so-called ghost-free massive gravity [43, 44], is free of many of these obstacles and furthermore contains two free parameters in its action. In Chap. 5, we consider tree-level scattering of massive graviton states, demonstrating that there is a finite region of this two-dimensional parameter space of couplings that obeys analyticity bounds. This result gives new hope for future model-building with massive gravity.

Chap. 6 revisits the curvature-squared correction to Einstein gravity first discussed in Chap. 4, but from a very different perspective. Rather than bound the coupling of the curvature-squared operator using analyticity, in Chap. 6 we consider how to build the ultraviolet completion of the operator "from scratch" in an arbitrary unitary tree-level theory. By enumerating such completions of the curvature-squared operator and requiring the principle of quantum mechanical unitarity, we demonstrate that the coupling of this operator must be positive, again providing a robust diagnostic for building models of modifications of general relativity by quantum gravity.

After having considered various applications of perturbative quantum gravity to the swampland program, in Chap. 7 we turn our attention to the representation of perturbation theory for pure Einstein gravity itself. In partic-

ular, we consider the symmetry of the *double copy*, a famous set of relations equating scattering amplitudes for gravitons to the sum of squares of scattering amplitudes for gluons in gauge (Yang-Mills) theory [45, 46]. This amazing relation for graviton scattering amplitudes suggests a hidden symmetry within general relativity: a twofold copy of Lorentz (rotation, translation, and boost) symmetry. In the canonical representation of the graviton perturbations [47–49], the Feynman rules become extremely complicated as one goes to higher-point graviton interaction vertices, containing an exponential number of terms and lacking any simple structure. Moreover, the canonical perturbation expansion of general relativity shows no sign of this twofold Lorentz symmetry. However, in Chap. 7, we discover a reformulation of canonical quantum gravity, a field redefinition and gauge-fixing of the graviton perturbation theory that makes the twofold Lorentz symmetry manifest and furthermore reduces drastically the number of terms in the perturbation expansion at a given order [6]. This result has potential utility in a number of applications for perturbative gravity.

While previous chapters have focused on quantum field theoretic aspects of perturbative quantum gravity, we now turn to nonperturbative questions arising from the combination of general relativity with quantum mechanics. One of the most exciting developments in physics of the past couple decades has been the notion of *holography* [50, 51], which in its modern form is the idea that a full quantum gravity theory in a given region can be represented by the dynamics of a certain non-gravitational quantum field theory on the boundary of the region, i.e., in one dimension lower. Arising from black hole physics and made concrete in the famous *anti-de Sitter/conformal field theory (AdS/CFT) correspondence* [52–54], holography has proven a fruitful field of study in high-energy physics, yielding insights not only into string theory, but having interesting interactions with quantum information theory [7, 8],

condensed matter physics [10], cosmology, and, as we will see, even classical general relativity [9].

In AdS/CFT, quantum gravity on a background spacetime of constant negative curvature (anti-de Sitter space) is represented by a conformal field theory on the boundary. Notably, a wormhole in the bulk geometry, consisting of two black holes connected by a nontraversable Einstein-Rosen (ER) bridge [55], is represented in the conformal field theory as a particular quantum state consisting of a large number of entangled energy eigenstates on either side of the wormhole [56]. Inspired by this example, Ref. [57] proposed a mapping in quantum gravity between entangled qubits (quantum mechanical computational bits)—so-called Einstein-Podolsky-Rosen (EPR) states [58]—and wormholes (ER bridges), resulting in the *ER=EPR correspondence*. In Chaps. 8 through 10, we conduct several nontrivial tests and applications of the ER=EPR correspondence in the classical limit. In particular, if true, the ER=EPR correspondence should result in a mapping between certain theorems in quantum mechanics and results in general relativity. Interestingly, in Chap. 8 we indeed find such a mapping, demonstrating that the no-cloning theorem in quantum mechanics, which prevents perfect quantum duplication of arbitrary states, is equivalent under ER=EPR to the no-go theorem preventing change of the topology of spacetime in general relativity.

In Chap. 9, we push this test of ER=EPR further, finding that the ER=EPR correspondence can be consistent with the linearity property of quantum mechanics. For a given state that consists of two entangled qubits, linearity of quantum mechanics implies that there exists no observation that can decide conclusively whether the state is entangled; one can project onto particular chosen entangled states, but not the entire set of entangled states. Similarly, we prove the ER=EPR analogue of this statement in classical general relativity,

implying that the topology of spacetime is not an observable in quantum gravity. In particular, one cannot detect definitively whether a given black hole interior is a wormhole or not. Thus, the ER=EPR correspondence appears to be well defined in the classical limit.

Another quantum mechanical statement that the ER=EPR correspondence should obey is conservation of entanglement: given two sets of entangled qubits that evolve unitarily but only interact among themselves and not with each other, the entanglement between the two should be conserved. Using this statement, we demonstrate in Chap. 10 that the ER=EPR correspondence predicts a new area theorem in general relativity: for two arbitrary collections of wormholes and black holes, interacting arbitrarily among themselves but not with each other, there should be a surface whose area is constant in time. We discover such a surface and prove the new area theorem purely within classical general relativity, giving a new conserved quantity for black hole and wormhole dynamics. It is striking that this new area theorem, present within the century-old theory of general relativity, was found due to considerations of the ultimate nature of spacetime and the quantum properties of gravity.

A theme that the AdS/CFT correspondence suggests is the importance of emergence and entanglement for understanding quantum gravity. In holography, an additional dimension of space, along with gravitation, emerges from a non-gravitational theory in a lower-dimensional spacetime. As such, the investigation of the phenomenon of emergent space and gravity is a compelling problem, to which Chaps. 11 and 12 of this thesis are devoted. In Chap. 11, we consider a particular proposal, closely related to AdS/CFT, for how an additional dimension of space emerges from the entanglement structure of a quantum state in a conformal field theory. In particular, recent work on *tensor networks*, a computational construction originally designed for efficiently finding

the ground state of certain quantum systems, suggests a potential link between a particular tensor network, the *Multiscale Entanglement Renormalization Ansatz (MERA)*, and anti-de Sitter space. In Chap. 11, we closely investigate this possibility, deriving a set of consistency conditions for a correspondence between the MERA and anti-de Sitter space, which is useful as a guide to future exploration of this intriguing proposition for the emergence of space.

Continuing in our exploration of the connections between gravity and entanglement, we next consider the possible emergence of gravity from entanglement entropy in the context of *entropic gravity* in Chap. 12. In particular, we focus on entropic gravity theories in which gravity emerges from constraints relating entanglement to areas in the geometry [59, 60], inspired by black hole thermodynamics and holography, asking the question of whether there is a consistent microscopic definition of entropy in these models [11]. We find that the classic model of entropic gravity, which we call thermodynamic gravity and in which Einstein's equation emerges as an equation of state from constraints on a dynamical lightsheet in a fixed spacetime background, is inconsistent. However, we find that holographic gravity, in which the constraint leading to Einstein's equation comes from stationarity of entropy in equilibrium for variations of the spacetime and quantum state over a family of nearby configurations, can be consistently defined in terms of the vacuum-subtracted entanglement entropy, the Casini entropy [61].

Having examined various consequences following from the application of quantum mechanics to general relativity, we now turn to the application of properties of general relativity to questions about the beginning of the universe. Cosmic *inflation*, the proposal that the universe underwent a period of exponential expansion in the first fraction of a second, is well known for solving various conundrums in cosmology and is typically implemented by

positing a theory of a scalar field that governs the dynamics of the universe at early times [62, 63]. Most notably, inflation solves the horizon problem, the mystery of why the cosmic microwave background is of uniform temperature across the entire sky, despite the fact that the corresponding regions of space, in the absence of inflation, would not have been in causal contact at the time of the formation of the cosmic microwave background.

Using the fact that general relativity, as a classical field theory, possesses a Hamiltonian constraint, Hawking et al. [64] showed that, for a homogeneous, isotropic universe with dynamics dominated by a scalar field, the natural measure on phase space can be turned into a probability distribution on the set of trajectories. One feature of this measure is that it selects universes that are spatially flat, but because of the divergence of the distribution, cannot make predictions among the subset of flat universes, like our own. In Chap. 13, we solve this problem, finding the unique classically conserved measure for spatially-flat cosmologies with scalar-field dynamics. This result has significant implications for differentiating between competing theories of inflation with similar observable predictions, quantifying the notion of tuning for cosmology [12].

We then demonstrate an example application of the result of Chap. 13 in Chap. 14, by directly comparing the amount of expansion resulting from two different models of inflation, with quadratic and cosine potentials. In particular, the results of Chap. 14 favor large-field inflation, by demonstrating that, under the classical probability distribution, quadratic inflation more naturally produces the amount of expansion required for our own universe than does the small-field inflation one obtains from a cosine potential [13]. As the current golden age of observational cosmology yields new results from terrestrial and space-based observations of the cosmic microwave background [65], such

comparisons will become ever more crucial, to constrain the possibilities for model-building for the dynamics of the very early universe.

The work in this thesis demonstrates the wide array of areas of investigation in current high-energy theoretical physics research into the nature of gravity, from effective field theory and the landscape of possible laws of physics, to holography and the relation between quantum entanglement and spacetime, to the universe itself. The future of work in this field is bright indeed, with many connections forming between different subfields of high-energy theory and new possibilities for future work constantly appearing on the horizon.

Chapter 2
Naturalness and the Weak Gravity Conjecture

The weak gravity conjecture (WGC) is an ultraviolet consistency condition asserting that an Abelian force requires a state of charge q and mass m with $q > m/m_{\rm Pl}$. We generalize the WGC to product gauge groups and study its tension with the naturalness principle for a charged scalar coupled to gravity. Reconciling naturalness with the WGC either requires a Higgs phase or a low cutoff at $\Lambda \sim q m_{\rm Pl}$. If neither applies, one can construct simple models that forbid a natural electroweak scale and whose observation would rule out the naturalness principle.

This chapter is from Ref. [1], C. Cheung and G. N. Remmen, "Naturalness and the Weak Gravity Conjecture," Phys. Rev. Lett. ***113*** *(2014) 051601,* arXiv:1402.2287 [hep-ph].

2.1 Introduction

The naturalness principle asserts that operators not protected by symmetry are unstable to quantum corrections induced at the cutoff. As a tenet of effective field theory, naturalness has provided a key motivation for new physics at the electroweak scale. However, the discovery of the Higgs boson [14, 15] together with null results from direct searches has led many to revisit naturalness as a fundamental principle. Rather than amend naturalness to fit the data, we instead explore its interplay with established concepts in quantum field theory.

Our focus will be the weak gravity conjecture (WGC) [30], which states that a consistent theory of gravity coupled to an Abelian gauge theory must

contain a state of charge q and mass m satisfying[2.1]

$$q > m/m_{\rm Pl}, \tag{2.1.1}$$

i.e., gravity is the weakest force. While Eq. (2.1.1) is certainly true of electromagnetism, Ref. [30] convincingly argued that it is a universal consistency condition of all healthy quantum field theories.

However, in theories with fundamental scalars Eq. (2.1.1) runs afoul of naturalness because it bounds a quadratically divergent mass by a logarithmically divergent charge. For small charge, Eq. (2.1.1) forbids a natural spectrum in which scalars have masses near the cutoff. We illustrate this contradiction with scalar quantum electrodynamics (QED) coupled to general relativity, but this tension is a ubiquitous feature of any model with a hierarchy problem and a small charge. We also generalize Eq. (2.1.1) to the case of multiple forces and particles.

As we will show, reconciling naturalness with Eq. (2.1.1) requires a revision of the original theory: either the gauge symmetry is spontaneously broken or new degrees of freedom enter prematurely at the cutoff

$$\Lambda \sim q m_{\rm Pl}. \tag{2.1.2}$$

Ref. [30] conjectured Eq. (2.1.2) with the stronger interpretation that Λ signals the complete breakdown of four-dimensional quantum field theory. Supporting this claim with compelling string theoretic examples, Ref. [30] fell short of a general argument. However, if one asserts the primacy of naturalness, then our logic provides a reason from quantum field theory for new states at Λ.

To illustrate these ideas we present simple, concrete extensions of the standard model (SM) in which a natural value of the electroweak scale—at the Planck scale—is incompatible with Eq. (2.1.1) due to a new millicharged force.

[2.1] In this chapter, we define the Planck mass, $m_{\rm Pl}$, such that Eq. (2.1.1) is saturated for an extremal black hole.

These models offer the unique opportunity to test naturalness experimentally. Indeed, either naturalness reigns, in which case Eq. (2.1.2) demands a low cutoff, or it fails. Absent additional ultralight states, a discovery of this millicharged force would then invalidate naturalness and mandate an unnatural electroweak scale. In particular, Eq. (2.1.1) would disallow a natural electroweak scale and the hierarchy problem would arise from as-yet-unknown ultraviolet dynamics. More generally, a fifth force discovery of any kind would invalidate the interpretation of Λ advocated in Ref. [30] as the cutoff of four-dimensional quantum field theory. If, as conjectured in Ref. [30], this breakdown is a universal feature of all string compactifications, such an observation would also falsify string theory.

2.2 Evidence for the WGC

Let us summarize the justification for the WGC [30]. Consider a $U(1)$ gauge theory with charged species labeled by i, each representing a particle (anti-particle) of charge q_i ($-q_i$) and mass m_i. We define dimensionless charge-to-mass ratios,

$$z_i = q_i m_{\rm Pl}/m_i, \tag{2.2.1}$$

so Eq. (2.1.1) implies that there exists some particle i with $z_i > 1$. The authors of Ref. [30] offered theoretical evidence in support of Eq. (2.1.1). They presented many examples from field theory and string theory, all satisfying Eq. (2.1.1). Further, they argued that Eq. (2.1.1) reconciles the inherent inconsistency of exact global symmetries with the naively innocuous $q \to 0$ limit of a gauge theory. This limit yields an exact global symmetry; however, such charges are not conserved by quantum gravity [66, 67] because, in accordance with no-hair theorems [68], a stationary black hole is fully characterized by its mass, spin,

and charge.

Of course, examples and consistency with no-hair theorems only provide circumstantial evidence for Eq. (2.1.1). Importantly, Ref. [30] also argues for Eq. (2.1.1) via *reductio ad absurdum*, drawing only on general relativity, conservation of charge and energy, and minimal assumptions about the ultimate theory of quantum gravity. Consider a black hole of charge Q and mass M decaying solely to particles of species i, which can occur via Hawking radiation or Schwinger pair production [69, 70]. By charge conservation, Q/q_i particles are produced. Conservation of energy dictates that the total rest mass of the final state, $m_i Q/q_i$, be less than M. In terms of the black hole charge-to-mass ratio, $Z = Q\, m_{\rm Pl}/M$, this implies $z_i > Z$. An extremal black hole corresponds to $Z = 1$ and is stable unless some state i exists for which $z_i > 1$. If Eq. (2.1.1) fails, the spectrum contains a large number of stable black hole remnants, in tension with holographic bounds [51, 71] and afflicted with various quantum gravitational and thermodynamic pathologies [72, 73].

2.3 The Limits of Naturalness

The WGC is straightforward at tree-level, but radiative corrections introduce subtleties. In fermionic QED, q and m run with renormalization scale, as does their ratio, naively making q/m ambiguous; however, as Ref. [30] notes, the appropriate scale to evaluate q/m is the physical mass of the particle. This is the mass scale that is relevant to the kinematics of extremal black hole decay, which provides the justification for the WGC.

However, the radiative stability question becomes more interesting in scalar QED:

$$\mathcal{L} = -\frac{1}{4}F_{\mu\nu}^2 + |D_\mu \phi|^2 - m^2|\phi|^2 - \frac{\lambda}{4}|\phi|^4, \tag{2.3.1}$$

where $D_\mu = \partial_\mu + iqA_\mu$ is the gauge covariant derivative. As for any effective field theory, we assume an ultraviolet cutoff, Λ, above which new physics enters. Since ϕ is a fundamental scalar, its mass is radiatively unstable and corrected by $m^2 \to m^2 + \delta m^2$ where

$$\delta m^2 = \frac{\Lambda^2}{16\pi^2}(aq^2 + b\lambda). \tag{2.3.2}$$

Here a and b are dimensionless numerical coefficients. We assume that δm^2 is positive so that the theory remains in the Coulomb phase. In a natural theory, the physical mass of ϕ cannot be parametrically smaller than its radiative corrections. Equivalently, the counterterm for the scalar mass should not introduce a delicate cancellation. This is formally equivalent to requiring that the coefficients a and b take on $\mathcal{O}(1)$ values.

Let us set the physical mass squared for ϕ to its natural value, δm^2, which the WGC forbids from exceeding its charge in Planck units. The charge-to-mass ratio of ϕ is

$$z = \frac{4\pi m_{\rm Pl}}{\Lambda}\frac{1}{\sqrt{a + b\lambda/q^2}}, \tag{2.3.3}$$

where the WGC implies that $z > 1$. If $q^2 \gg \lambda$, then

$$\Lambda < \frac{4\pi m_{\rm Pl}}{\sqrt{a}}, \tag{2.3.4}$$

which is the reasonable requirement that the cutoff not exceed the Planck scale.

Turning to the opposite hierarchy, $q^2 \ll \lambda$, which is also radiatively stable, we find that the WGC implies

$$\Lambda < 4\pi m_{\rm Pl}\sqrt{\frac{q^2}{b\lambda}}. \tag{2.3.5}$$

As $q^2/\lambda \to 0$, a sensible cutoff requires $b \to 0$, indicating mandatory fine-tuning in order to satisfy the WGC. We are left with a remarkable conclusion: scalar QED with $q^2 \ll \lambda$ and natural masses fails Eq. (2.1.1) and is thus inconsistent with a quantum theory of gravity.

We have not traded a mass scale hierarchy problem for an equivalent hierarchy problem of couplings. Small charges are radiatively stable and thus technically natural. In principle, $q^2 \ll \lambda$ is no worse than the small electron Yukawa coupling.

To reconcile naturalness with Eq. (2.1.1), one alternative is to argue that the original theory—scalar QED with $q^2 \ll \lambda$—is impossible. For example, this would be true if nature does not permit fundamental scalars or if a hierarchy among couplings is somehow strictly forbidden. However, there are far less drastic options, elaborated below, if one modifies the original scalar QED theory.

i) **Radiative corrections induce the Higgs phase.** It is possible that quantum effects generate a tachyon for ϕ, Higgsing the theory. Charge becomes ill defined; the charge and mass eigenbases need not commute, leaving q/m ambiguous. Further, the WGC is not justified in the Higgs phase. The original argument for the WGC [30] relied on stable extremal black holes. However, no-hair theorems imply that there are no stationary black hole solutions supporting classical hair from a massive photon [74], independent of the size of the black hole relative to the Higgs scale. If a black hole accretes a massive-$U(1)$-charged particle, it briefly supports an associated electric field, but after a time of order the photon Compton wavelength, it balds [75] when the gauge field is radiated away to infinity or through the horizon.

ii) **New physics enters below the Planck scale.** The simplest way to reconcile the WGC with naturalness is for the effective field theory to break down at a cutoff defined by Eq. (2.3.5). There could be new light states regulating quadratic divergences of ϕ, effectively lowering Λ. This option resolves the contradiction tautologically by eliminating the hierarchy problem altogether. However, a more interesting alternative occurs when the new states

do not couple to ϕ. The quadratic divergence of ϕ is robust and m is large. If one of these new states satisfies Eq. (2.1.1), then ϕ is irrelevant: the WGC and naturalness are reconciled. Thus, asserting naturalness offers Eq. (2.3.5) as a more precise version of the low cutoff conjecture of Ref. [30] stated in Eq. (2.1.2).

2.4 More Forces, More Particles

Extending our results to various charged species of different spins, the WGC implies that at least one state in the spectrum must satisfy Eq. (2.1.1) after taking into account radiative corrections. Naturalness is violated in parameter regions with a hierarchy between charges and couplings that generate quadratic divergences (quartic couplings, Yukawa couplings).

The story becomes more interesting for product gauge symmetries. Consider a gauge group $\prod_{a=1}^{N} U(1)_a$ and particles i with charges q_{ia} and masses m_i. We represent the charges, $\vec{q}_i = q_{ia}$, and charge-to-mass ratios, $\vec{z}_i = q_{ia} m_{\rm Pl}/m_i$, as vectors of $SO(N)$, the symmetry transforming the N photons among each other. If present, photon kinetic mixing can be removed by a general linear transformation on the photons, which is equivalent to redefining charge vectors of states in the theory.

To generalize the WGC for multi-charged particles, Eq. (2.1.1) is inadequate and requires upgrading to a constraint on $\vec{q}_i$ and m_i. Ref. [30] briefly alluded to this scenario, but detailed analysis will reveal quantitative differences between the WGC as applied to a single $U(1)$ versus many. By symmetry, the proper generalized WGC must be $SO(N)$ invariant. Naively, the WGC could require at least one species i with $|\vec{z}_i| > 1$. However, this is insufficient—it guarantees the existence of one particle of large total charge, but preserves stability for orthogonally-charged extremal black holes. A stricter alternative is that for

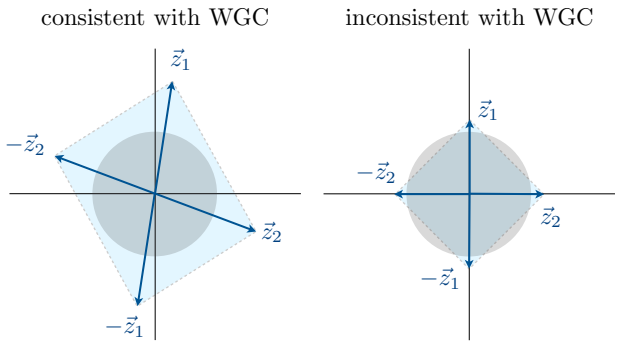

Figure 2.1. Vectors representing charge-to-mass ratios for two species charged under two Abelian gauge symmetries. When the convex hull defined by these vectors contains the unit ball, then extremal black holes can decay to particles and the condition of the WGC is satisfied.

each $U(1)$ there exists a species i charged under that $U(1)$ with $|\vec{z}_i| > 1$. Curiously, this is still actually weaker than the true generalized WGC.

To determine the proper generalized WGC, we revisit black hole decay kinematics. Consider a black hole of charge $\vec{Q}$, mass M, and charge-to-mass ratio $\vec{Z} = \vec{Q}\, m_{\rm Pl}/M$ decaying to a final state comprised of n_i particles of species i. Charge and energy conservation imply $\vec{Q} = \sum_i n_i \vec{q}_i$ and $M > \sum_i n_i m_i$. If $\sigma_i = n_i m_i/M$ is the species i fraction of the total final state mass, then $\vec{Z} = \sum_i \sigma_i \vec{z}_i$ and $1 > \sum_i \sigma_i$; decay requires that $\vec{Z}$ be a subunitary weighted average of $\vec{z}_i$. This criterion has a geometric interpretation in charge space. Draw the vectors $\pm\vec{z}_i$ corresponding to the charge-to-mass ratio of each fundamental particle in the spectrum. A weighted average of $\vec{z}_i$ defines the convex hull spanned by the vectors, delineating the space of $\vec{Z}$ that is unstable to decay. Any state outside the convex hull is stable. Since extremal black holes correspond to $|\vec{Z}| = 1$, the generalized WGC requires that the convex hull spanned by $\pm\vec{z}_i$ contain the unit ball.

Consider a model of two Abelian factors and two charged states. The left and right panels of Fig. 2.1 represent two possible choices for the charge-to-mass ratios of the particles. Black holes of all possible charges are represented by the

unit disc. The left panel of Fig. 2.1 depicts a theory that is consistent with the WGC: the unit disc is contained in the convex hull. Extremal black holes, the boundary of this disc, can decay. However, the right panel of Fig. 2.1 depicts a theory that violates the WGC: there are regions of the unit disc not within the convex hull, corresponding to stable black hole remnants. Remarkably, this theory fails the WGC despite the fact that $|\vec{z}_1| > 1$ and $|\vec{z}_2| > 1$. Simple geometry shows that the WGC imposes the more stringent constraint:

$$(\vec{z}_1^{\,2} - 1)(\vec{z}_2^{\,2} - 1) > (1 + |\vec{z}_1 \cdot \vec{z}_2|)^2. \tag{2.4.1}$$

For example, given orthogonal charges of equal magnitude, $|\vec{z}_1| = |\vec{z}_2| = z$ and $\vec{z}_1 \perp \vec{z}_2$, Eq. (2.4.1) implies $z > \sqrt{2}$, manifestly stronger than the $z > 1$ condition required for theories with a single $U(1)$. Note that the WGC places constraints on $\vec{z}_1$ and $\vec{z}_2$ that are not mathematically independent. Were a particular value of $\vec{z}_1$ experimentally observed, this would fix a bound $\vec{z}_2^{\,2} > (1 - 1/\vec{z}_1^{\,2})^{-1}$.

A similar analysis can be applied for N Abelian factors and N charged states. Suppose each particle is charged under a single $U(1)$, with equal magnitude charge-to-mass ratios, so $z_{ia} = \delta_{ia} z$ for some z. The convex hull defined by $\pm\vec{z}_i$ is an N-dimensional cross-polytope of circumradius z. The largest ball contained in the cross-polytope has radius $z/\sqrt{N}$. Requiring that the radius of this ball be greater than unity then implies $z > \sqrt{N}$, parametrically stronger than the condition required for a single Abelian factor.

The WGC constraint grows at large N for fixed physical Planck scale $m_{\rm Pl}$. However, the presence of N additional species generally renormalizes the strength of gravity [28, 76, 77] as $\delta m_{\rm Pl}^2 \sim N\Lambda^2/16\pi^2$. If corrections enhance $m_{\rm Pl}$ by a factor of $\sqrt{N}$, all factors of N encountered in our earlier analyses cancel. That is, in a theory with fixed Lagrangian parameters and cutoff, the limit from the WGC is N-independent at large N. A similar phenomenon

was discussed in Ref. [30] for N Abelian factors Higgsed to a $U(1)$ subgroup. The large-N limit introduces a $\mathbb{Z}_2^N$ symmetry, which is subject to the large-N species bounds considered in Ref. [76].

The multi-charge generalized WGC has implications for naturalness. Consider a $U(1)^N$ gauge theory with scalars ϕ_i of charges $\vec{q}_i$ and masses m_i,

$$\mathcal{L} = -\frac{1}{4}\sum_a F_{\mu\nu a}^2 + \sum_i |D_\mu \phi_i|^2 - m_i^2|\phi_i|^2 - \frac{\lambda_i}{4}|\phi_i|^4, \tag{2.4.2}$$

where $D_\mu\phi_i = (\partial_\mu + i\sum_a q_{ia}A_{\mu a})\phi_i$. Radiative corrections send $m_i^2 \to m_i^2 + \delta m_i^2$, where

$$\delta m_i^2 = \frac{\Lambda^2}{16\pi^2}(a_i \vec{q}_i^{\,2} + b_i\lambda_i) \tag{2.4.3}$$

and a_i and b_i are $\mathcal{O}(1)$ ultraviolet-sensitive coefficients. The charge-to-mass ratio vector for ϕ_i is

$$\vec{z}_i = \frac{4\pi m_{\rm Pl}}{\Lambda}\frac{\vec{q}_i}{|\vec{q}_i|}\frac{1}{\sqrt{a_i + b_i\lambda_i/\vec{q}_i^{\,2}}}. \tag{2.4.4}$$

A necessary albeit insufficient condition for the WGC is that, for each $U(1)$, there is a state i charged under that Abelian factor such that $|\vec{z}_i| > 1$. This implies

$$\Lambda < 4\pi m_{\rm Pl} \times \begin{cases} \dfrac{1}{\sqrt{a_i}} \ , & \vec{q}_i^{\,2} \gg \lambda_i \\ \\ \sqrt{\dfrac{\vec{q}_i^{\,2}}{b_i\lambda_i}} \ , & \vec{q}_i^{\,2} \ll \lambda_i \end{cases}. \tag{2.4.5}$$

As for the single Abelian case, $\vec{q}_i^{\,2} \gg \lambda_i$ corresponds to the reasonable requirement of a sub-Planckian cutoff, while $\vec{q}_i^{\,2} \ll \lambda_i$ implies tension with naturalness. However, the most stringent requirement of the WGC—that the convex hull spanned by $\pm\vec{z}_i$ contain the unit ball—places a stronger limit than Eq. (2.4.5) by a factor of order $\sqrt{N}$ for fixed $m_{\rm Pl}$.

2.5 The Hierarchy Problem

We have presented explicit models in which naturalness contradicts Eq. (2.1.1). We now construct theories in which natural values of the electroweak scale—at the cutoff—are similarly incompatible. In these models, strict adherence to naturalness implies either a Higgs phase or a parametrically low cutoff given by Eq. (2.1.2).

The obvious path is to relate the electroweak scale to the mass m of a particle that carries a tiny charge q. The SM gauge couplings are $\mathcal{O}(1)$, so we require an additional $U(1)$ gauge symmetry beyond the SM. It is tempting to charge the Higgs, but this will spontaneously break the $U(1)$, invalidating the applicability of the WGC.

However, we can charge the SM fermions under a very weakly gauged unbroken $U(1)_{B-L}$ symmetry. Current limits on $U(1)_{B-L}$ require $q \lesssim 10^{-24}$ [78, 79] and will likely be improved by several orders of magnitude by astrophysical [80], lunar ranging [81], and satellite-based [82–84] tests of apparent equivalence principle violation. To cancel anomalies we introduce a right-handed neutrino $\nu_{\rm R}$ that combines with the left-handed neutrino $\nu_{\rm L}$ to form a $U(1)_{B-L}$ preserving Dirac mass term of the form $m_\nu \bar{\nu}_{\rm L} \nu_{\rm R} + \text{h.c.}$, where $m_\nu \sim y_\nu v$ is controlled by the electroweak symmetry breaking scale. The particle with the largest charge-to-mass ratio is the lightest neutrino. Assuming its mass is of order the neutrino mass scale, $m_\nu \lesssim 0.1$ eV [85, 86], we fix the charge to a technically natural albeit tiny value: $q \sim m_\nu / m_{\rm Pl} \sim 10^{-29}$. For this value of q, Eq. (2.1.1) is just marginally satisfied by the lightest neutrino. While such a charge is permitted in quantum field theory, it may be difficult to engineer in string theory if q arises from a string coupling constant requiring dilaton stabilization at large field values. Similar issues arise in theories of large extra dimensions

and it is a detailed question of string moduli stabilization whether this is possible. In any case, at fixed Yukawa coupling y_ν, were the electroweak scale any higher than its measured value, Eq. (2.1.1) would fail. In this model, regions of parameter space favored by naturalness—and an electroweak scale at the cutoff—are inconsistent with Eq. (2.1.1). Strictly speaking, this logic hinges on the absence of additional $U(1)_{B-L}$ charged states lighter than the neutrino. Depending on the cosmological history, however, such particles may be constrained experimentally by primordial nucleosynthesis.

Our model offers a direct experimental test of naturalness by virtue of a very specific prediction: a new gauge boson very weakly coupled to the SM. As discussed earlier, the assumption of naturalness mandates either a Higgs phase or a low cutoff. The discovery of a fifth force would rule out the former, while current sensitivities would for the latter imply $\Lambda \lesssim$ keV from Eq. (2.1.2). In the absence of such ultralight states, the observation of a $U(1)_{B-L}$ gauge boson at $q \sim 10^{-29}$ would then simultaneously falsify the naturalness principle and suggest an ultraviolet-dependent reason for the why the weak scale takes an unnatural value. Moreover, given present sensitivities, a fifth force discovery of any kind would falsify string theory to the extent to which it predicts the strong interpretation of Λ as the scale at which four-dimensional quantum field theory breaks down [30].

This mechanism can generally be incorporated into any theory where the electroweak scale sources the mass of a $U(1)$ millicharged state, e.g., dark matter [87] charged under an unbroken $U(1)$ dark force. For weak scale dark matter, a charge of $q \sim 10^{-16}$ is sufficient to saturate Eq. (2.1.1).

Chapter 3
Infrared Consistency and the Weak Gravity Conjecture

The weak gravity conjecture (WGC) asserts that an Abelian gauge theory coupled to gravity is inconsistent unless it contains a particle of charge q and mass m such that $q \geq m/m_{\rm Pl}$. This criterion is obeyed by all known ultraviolet completions and is needed to evade pathologies from stable black hole remnants. In this chapter, we explore the WGC from the perspective of low-energy effective field theory. Below the charged particle threshold, the effective action describes a photon and graviton interacting via higher-dimension operators. We derive infrared consistency conditions on the parameters of the effective action using *i*) analyticity of light-by-light scattering, *ii*) unitarity of the dynamics of an arbitrary ultraviolet completion, and *iii*) absence of superluminality and causality violation in certain nontrivial backgrounds. For convenience, we begin our analysis in three spacetime dimensions, where gravity is non-dynamical but has a physical effect on photon-photon interactions. We then consider four dimensions, where propagating gravity substantially complicates all of our arguments, but bounds can still be derived. Operators in the effective action arise from two types of diagrams: those that involve electromagnetic interactions (parameterized by a charge-to-mass ratio q/m) and those that do not (parameterized by a coefficient γ). Infrared consistency implies that q/m is bounded from below for small γ.

This chapter is from Ref. [2], C. Cheung and G. N. Remmen, "Infrared Consistency and the Weak Gravity Conjecture," JHEP ***1412*** *(2014) 087,* arXiv:1407.7865 [hep-th].

3.1 Introduction

The weak gravity conjecture (WGC) [30] asserts a powerful restriction on any Abelian gauge theory coupled consistently to gravity. In particular, it mandates the existence of a state of charge q and mass m satisfying[3.1]

$$q \geq m. \tag{3.1.1}$$

Informally, the WGC states that "gravity is the weakest force" because it bounds the gravitational charge of a state from above by its electric charge. The WGC is a beautiful and sharply defined criterion demarcating the landscape from the swampland.

The authors of Ref. [30] supported their conjecture with numerous examples from field theory and string theory, all satisfying the WGC. Moreover, they offered an elegant argument by contradiction in favor of the WGC. By conservation of charge and energy, the state with the largest charge-to-mass ratio cannot decay, so violation of the WGC implies the absolute stability of extremal black holes, which exactly saturate Eq. (3.1.1). However, stable black hole remnants are thought to be pathological [51, 71–73], so the authors of Ref. [30] argued that the WGC is mandatory in any theory with an Abelian gauge symmetry.

In this chapter, we explore the WGC from the viewpoint of effective field theory. Our central question is simple: does violation of the WGC induce a pathology in the infrared? To seek an answer, we consider energies far below the charged particle threshold, where the dynamics are described by photons and gravitons interacting via higher-dimension operators:

[3.1] Throughout this chapter, we use natural units for mass and charge in which $4\pi G = \epsilon_0 = 1$, with $(+,-,-,\ldots)$ metric signature and curvature tensors $R_{\mu\nu} = R^{\rho}{}_{\mu\rho\nu}$ and $R^{\rho}{}_{\mu\sigma\nu} = \partial_\sigma \Gamma^{\rho}{}_{\mu\nu} - \partial_\nu \Gamma^{\rho}{}_{\mu\sigma} + \Gamma^{\rho}{}_{\alpha\sigma}\Gamma^{\alpha}{}_{\mu\nu} - \Gamma^{\rho}{}_{\alpha\nu}\Gamma^{\alpha}{}_{\mu\sigma}$, all for arbitrary spacetime dimension D.

$$\begin{aligned}\mathcal{L} = &-\frac{1}{4}F_{\mu\nu}F^{\mu\nu} - \frac{1}{4}R \\ &+ a_1(F_{\mu\nu}F^{\mu\nu})^2 + a_2(F_{\mu\nu}\tilde{F}^{\mu\nu})^2 \\ &+ b_1 F_{\mu\nu}F^{\mu\nu}R + b_2 F_{\mu\rho}F_{\nu}{}^{\rho}R^{\mu\nu} + b_3 F_{\mu\nu}F_{\rho\sigma}R^{\mu\nu\rho\sigma} \\ &+ c_1 R^2 + c_2 R_{\mu\nu}R^{\mu\nu} + c_3 R_{\mu\nu\rho\sigma}R^{\mu\nu\rho\sigma},\end{aligned} \tag{3.1.2}$$

where $\tilde{F}_{\mu\nu} = \epsilon_{\mu\nu\rho\sigma}F^{\rho\sigma}/2$. We have dropped terms like $(\nabla_\mu F_{\nu\rho})^2$ and $(\nabla_\mu F^{\mu\nu})^2$, which in the absence of charged sources can be written in terms of the operators already included.

Electromagnetic interactions induce contributions to a_i and b_i that depend on the charges and masses of every state in the spectrum. Each contribution grows with charge and scales inversely with mass, so they are dominated by the state in the spectrum with the largest charge-to-mass ratio, which we will write as $z = q/m$. Crucially, the operator coefficients in the effective Lagrangian (3.1.2) are sensitive to the same quantity as the WGC, which posits that

$$z \geq 1. \tag{3.1.3}$$

Because the photon-graviton effective action is z-dependent, there is hope that an analysis of the infrared dynamics might shed light on the WGC.

From a purely low-energy perspective, it would seem reasonable for the landscape of high-energy completions to span all values of the parameters in the effective action. However, as discussed in Ref. [31], this is a misconception: some effective theories are intrinsically pathological and never emerge from consistent ultraviolet dynamics. This occurs, for example, in the Euler-Heisenberg Lagrangian [88–90], which is Eq. (3.1.2) in the limit that gravity is decoupled. When $a_i < 0$, the theory admits superluminal photon propagation and non-analyticity in the light-by-light scattering amplitude. Unsurprisingly, $a_i \geq 0$ in all known ultraviolet completions. More recently, bounds on graviton interactions were derived in Ref. [37].

The purpose of this chapter is to apply similar methods to determine

infrared consistency conditions on the effective action describing the low-energy interactions of photons and gravitons. In particular we derive constraints on the parameters of Eq. (3.1.2) from three independent criteria:

i) **Analyticity.** We study the analytic properties of the light-by-light scattering amplitude. Forward dispersion relations constrain the effective theory parameters.

ii) **Unitarity.** We construct a spectral representation parameterizing an arbitrary ultraviolet completion. Forbidding ghosts and tachyons constrains the effective theory parameters.

iii) **Causality.** We compute the speed of light in certain nontrivial backgrounds. Absence of superluminality and causality violation constrains the effective theory parameters.

As a warmup, we study the photon-graviton effective theory in three spacetime dimensions (3D), where gravity is purely topological [91]. While the graviton is non-propagating, it still mediates contact interactions for the photon. Remarkably, arguments from analyticity, unitarity, and causality all imply an identical constraint on the parameters of the effective theory:

$$a' \geq 0, \tag{3.1.4}$$

where $a' = a_1 + b_1 - b_3 + c_1 + c_2 + 3c_3$. We can, however, learn more by inputting additional assumptions about the ultraviolet completion. For example, consider the case where the dominant contributions to a_i and b_i originate from diagrams involving electromagnetic interactions of a fermion with charge-to-mass ratio z. As we will see, Eq. (3.1.4) then implies a constraint on a two-dimensional parameter space spanned by z and a coefficient γ parameterizing purely gravitational corrections to the effective action. The theory automatically

satisfies our consistency conditions if γ exceeds a certain critical value. However, below this critical value, the theory is consistent only for certain values of z. In particular, for small γ, infrared consistency implies that $z \geq 1$, a 3D version of the WGC.

Subsequently, we move on to four spacetime dimensions (4D), where dynamical gravity introduces a litany of subtleties, which we discuss at length in the body of this chapter. For now, let us simply summarize our results. As we will see, unitarity arguments imply that

$$a_1' \geq 0 \qquad \text{and} \qquad a_2' \geq 0, \tag{3.1.5}$$

where $a_1' = a_1 - b_2/2 - b_3 + c_2 + 4c_3$ and $a_2' = a_2 - b_2/2 - b_3 + c_2 + 4c_3$. Meanwhile, the absence of superluminal photon propagation in certain nontrivial backgrounds implies that

$$a_1' + a_2' \geq 0, \tag{3.1.6}$$

which is also a consequence of the unitarity bounds in Eq. (3.1.5). Analyticity arguments, on the other hand, are suspect in 4D because they rely crucially on the forward light-by-light scattering amplitude, which is ill defined due to singular t-channel graviton exchange [31]. Nevertheless, if one can assume that dispersion relations apply to contributions to the forward amplitude from higher-dimension operators, then remarkably, Eq. (3.1.6) can also be derived as a consequence of analyticity. In this sense, arguments from analyticity, unitarity, and causality in 4D all point to the set of mutually consistent bounds in Eqs. (3.1.5) and (3.1.6).

These bounds imply 4D constraints on the parameter space defined by z and the coefficients γ that parameterize purely gravitational effects. Our results in 4D are summarized in Fig. 3.3. In all cases, when γ is small, infrared consistency implies a lower bound on z that is numerically stronger than the

WGC. Curiously, in this regime we find that Eq. (3.1.6) results in the exact same bound for fermions and scalars: $z \geq 2$.

The remainder of the chapter is structured as follows. In Sec. 3.2, we derive constraints on the photon-graviton effective action in 3D coming from analyticity, unitarity, and causality. We then present the analogous arguments for the photon-graviton effective action in 4D in Sec. 3.3. Finally, we conclude and discuss future directions in Sec. 3.4.

3.2 Three Dimensions

3.2.1 Setup and Bounds (3D)

To begin, we re-express Eq. (3.1.2) in a form convenient for studying the dynamics of interacting photons. Specifically, we eliminate all dependence on the spacetime curvature in favor of the electromagnetic field strength. We start by rewriting the Riemann tensor in terms of the Ricci scalar, Ricci tensor, and Weyl tensor, which in D dimensions is[3.2]

$$\begin{aligned} C_{\mu\nu\rho\sigma} = R_{\mu\nu\rho\sigma} - \frac{1}{D-2}(g_{\mu[\rho}R_{\sigma]\nu} - g_{\nu[\rho}R_{\sigma]\mu}) \\ + \frac{1}{(D-1)(D-2)}Rg_{\mu[\rho}g_{\sigma]\nu}, \end{aligned} \tag{3.2.1}$$

where in 3D the Weyl tensor identically vanishes and Eq. (3.2.1) implies that

$$C_{\mu\nu\rho\sigma}C^{\mu\nu\rho\sigma} = R^2 - 4R_{\mu\nu}R^{\mu\nu} + R_{\mu\nu\rho\sigma}R^{\mu\nu\rho\sigma}, \tag{3.2.2}$$

so the Gauss-Bonnet term vanishes identically in 3D. Next, we eliminate all dependence on the Ricci tensor and Ricci scalar in the higher-dimension operators by rewriting them via the tree-level Einstein field equations,

$$R_{\mu\nu} - \frac{1}{2}g_{\mu\nu}R = 2T_{\mu\nu}, \tag{3.2.3}$$

[3.2] Throughout this chapter, square brackets denote un-normalized antisymmetrization, viz. $T_{[\mu\nu]} = T_{\mu\nu} - T_{\nu\mu}$.

which at the order of the Lagrangian (3.1.2) is equivalent to a field redefinition of the graviton. Meanwhile, the energy-momentum tensor is

$$T_{\mu\nu} = -F_{\mu\rho}F_{\nu}{}^{\rho} + \frac{1}{4}g_{\mu\nu}F_{\rho\sigma}F^{\rho\sigma}, \tag{3.2.4}$$

so Eq. (3.1.2) can be expressed solely in terms of the electromagnetic field strength. In particular, Eqs. (3.2.3) and (3.2.4) imply that $R^2 = R_{\mu\nu}R^{\mu\nu} = (F_{\mu\nu}F^{\mu\nu})^2$.

At leading order in derivatives, the only invariants constructed from the electromagnetic field strength are $(F_{\mu\nu}F^{\mu\nu})^2$ and $F_{\mu\rho}F_{\nu}{}^{\rho}F^{\mu}{}_{\sigma}F^{\nu\sigma}$. In 3D, these are algebraically related by $F_{\mu\rho}F_{\nu}{}^{\rho}F^{\mu}{}_{\sigma}F^{\nu\sigma} = (F_{\mu\nu}F^{\mu\nu})^2/2$. Thus, the final form of the photon-graviton effective Lagrangian in 3D is remarkably simple:

$$\mathcal{L} = -\frac{1}{4}F_{\mu\nu}F^{\mu\nu} - \frac{1}{4}R + a'(F_{\mu\nu}F^{\mu\nu})^2. \tag{3.2.5}$$

Here we have defined a new higher-dimension operator coefficient,

$$a' = a_1 + b_1 - b_3 + c_1 + c_2 + 3c_3, \tag{3.2.6}$$

written in terms of the original parameters in the Lagrangian (3.1.2) after discarding the operator $(F_{\mu\nu}\tilde{F}^{\mu\nu})^2$, which does not exist in 3D.

Next, we exploit a nice feature of 3D, namely, that a photon is equivalent to a scalar. To simplify our calculations, we dualize the photon according to

$$F_{\mu\nu} = i\epsilon_{\mu\nu\rho}\partial^{\rho}\phi, \tag{3.2.7}$$

where $\epsilon_{\mu\nu\rho}$ is the 3D Levi-Civita tensor and the overall coefficient is fixed so that ϕ is a canonically normalized state with positive norm. After dualization, Eq. (3.2.5) becomes

$$\mathcal{L} = \frac{1}{2}\partial_{\mu}\phi\partial^{\mu}\phi - \frac{1}{4}R + 4a'(\partial_{\mu}\phi\partial^{\mu}\phi)^2, \tag{3.2.8}$$

which is our final form for the photon-graviton effective Lagrangian in 3D. The underlying gauge symmetry of the photon is encoded in the shift symmetry of ϕ.

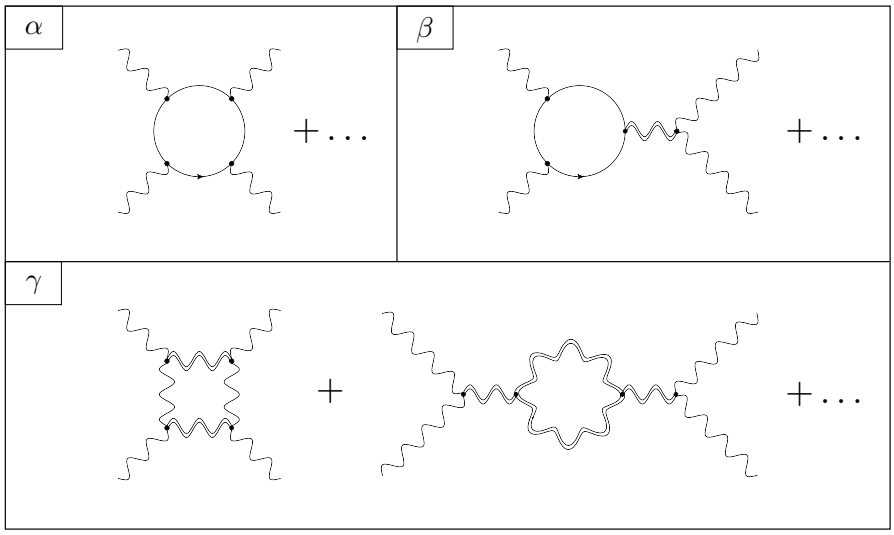

Figure 3.1. Diagrams involving photons (single wavy), gravitons (double wavy), and charged matter (solid) that contribute to light-by-light scattering, organized in terms of scaling with $z = q/m$, as defined in Eq. (3.2.10). Here, γ parameterizes purely gravitational corrections.

As we will derive shortly, the constraints from analyticity, causality, and unitarity in 3D all imply the exact same constraint,

$$a' \geq 0. \tag{3.2.9}$$

How might this bound constrain the spectrum of the ultraviolet completion? As noted earlier, the coefficients a_i and b_i in Eq. (3.1.2) receive calculable contributions from every charged particle in the spectrum, but they are dominated by the state with the largest charge-to-mass ratio, defined to be $z = q/m$. Without loss of generality, we can thus expand a' in powers of z as

$$a' = \alpha z^4 + \beta z^2 + \gamma. \tag{3.2.10}$$

Primordially, α, β, and γ arise from diagrams with four, two, and zero insertions of the electromagnetic coupling, respectively, as shown in Fig. 3.1.

By definition, α and β are contributions coming from diagrams that contain electromagnetic interactions. For example, integrating out a charged fermion in 3D yields calculable threshold corrections to the higher-dimension operator

coefficients [92, 93],

$$\begin{aligned} a_1 &= \frac{q^4}{1920\pi m^5} \\ (b_1, b_2, b_3) &= \left(-\frac{q^2}{1152\pi m^3},\ \frac{13q^2}{2880\pi m^3},\ -\frac{q^2}{2880\pi m^3} \right). \end{aligned} \tag{3.2.11}$$

In 3D, q has mass dimension 1/2. By substituting Eq. (3.2.11) into Eq. (3.2.6) and comparing to Eq. (3.2.10), we straightforwardly obtain α and β. Despite the complicated numerical factors in Eq. (3.2.11), we find that $\alpha/\beta = -1$. Meanwhile, since γ is independent of q, it necessarily parameterizes all contributions arising solely from gravitational interactions. These include the combination of coefficients $c_1 + c_2 + 3c_3$ in Eq. (3.2.6). Because γ is incalculable within the low-energy effective theory, it should be thought of as a high-energy boundary condition encoding the gravitational dynamics of the ultraviolet completion. Finally, rewriting Eq. (3.2.9) in terms of z and γ, we find that

$$z^2(z^2 - 1) \geq -\gamma m \times 1920\pi. \tag{3.2.12}$$

If $\gamma \geq 1/7680\pi m$, then this bound is satisfied for any value of z. This is a sufficient albeit not necessary condition for satisfying bounds from analyticity, unitarity, and causality.

On the other hand, it is interesting to consider the case in which the gravitational corrections are small, so $\gamma \sim 0$. In this case, our bounds imply that

$$z \geq 1, \tag{3.2.13}$$

which is the 3D analogue of the WGC in Eq. (3.1.1). This result is interesting because the argument for Eq. (3.1.1) from Ref. [30] derives from pathologies of stable extremal black holes, which do not exist in asymptotically-flat 3D spacetime. In this sense, infrared consistency conditions have more general applicability than the extremal black hole arguments of Ref. [30].

A priori, the 3D effective theory could arise from the compactification of

a higher-dimensional theory. Of course, even then, the infrared consistency condition in Eq. (3.2.9) would hold. However, if the compactification scale were less than m/q, then interactions generated from integrating out the radion and the Kaluza-Klein modes would dominate over those generated by the charged states. In this case, z would contribute negligibly to the effective action and infrared consistency would simply bound the parameter γ.

3.2.2 Analyticity (3D)

In this section, we exploit the analytic properties of the light-by-light scattering amplitude to constrain the 3D effective Lagrangian in Eq. (3.2.8). Following the procedure of Ref. [31], we consider the scattering amplitude

$$\mathcal{M}(s,t) = 8a'(s^2+t^2+u^2), \tag{3.2.14}$$

where the Mandelstam variables satisfy $s+t+u=0$. The forward scattering amplitude is then $\mathcal{M}(s) = \mathcal{M}(s, t\to 0) = 16a's^2$. Next, to extract the operator coefficient we compute the contour integral of $\mathcal{M}(s)/s^3$ around a contour $\mathcal{C}$ encircling the origin:

$$\begin{aligned} 16a' &= \oint_{\mathcal{C}} \frac{\mathrm{d}s}{2\pi i}\frac{\mathcal{M}(s)}{s^3} = \oint_{\mathcal{C}'} \frac{\mathrm{d}s}{2\pi i}\frac{\mathcal{M}(s)}{s^3} \\ &= \left(\int_{-\infty}^{-s_0} + \int_{s_0}^{\infty}\right)\frac{\mathrm{d}s}{2\pi i}\frac{\mathrm{Disc}[\mathcal{M}(s)]}{s^3} + \text{boundary integral.} \end{aligned} \tag{3.2.15}$$

Following Ref. [31], we have used the Cauchy integral theorem to deform $\mathcal{C}$ into a new contour $\mathcal{C}'$ composed of lines running just above and below the real axis plus a large circular boundary contribution at infinity. The discontinuity function is

$$\mathrm{Disc}[\mathcal{M}(s)] = \mathcal{M}(s+i\epsilon) - \mathcal{M}(s-i\epsilon) = 2i\mathrm{Im}[\mathcal{M}(s)], \tag{3.2.16}$$

where the difference of terms arises from the contour integration above and below the real axis and we used analyticity of $\mathcal{M}(s)$ to apply the Schwarz

reflection principle, $\mathcal{M}(s^*) = \mathcal{M}(s)^*$. Deforming the contour is mathematically permitted, provided $\mathcal{M}(s)$ is analytic in the bulk of the complex s plane and in the neighborhood of $s = 0$. The former is guaranteed by the usual stipulation that all non-analyticities of the S-matrix, e.g., poles and branch cuts, occur near the real axis. The latter is ensured by an additional physical input, which is that the scattering amplitude does not have branch cuts on the real axis extending to $s = 0$. At one-loop order in the effective action, light-by-light scattering will include massless branch cuts from a photon loop and two insertions of the $(F_{\mu\nu}F^{\mu\nu})^2$ operator. However, as discussed in Ref. [31], such cuts can be avoided by a slight deformation of the contour after introducing a regulator mass for the photon. Moreover, there are no branch cuts from gravitons, which are non-dynamical in 3D. For concreteness, we define s_0 to be the mass squared of the lowest-lying degree of freedom produced from light-by-light scattering, so $\mathcal{M}(s)$ is analytic in the region $|s| < s_0$.

The contour integral over $\mathcal{C}'$ includes a contribution from the discontinuity across the real axis as well as a contribution from infinity. In D dimensions, unitarity and polynomial boundedness of amplitudes implies the Froissart bound for large $|s|$, $|\mathcal{M}(s)| \lesssim |s \log^{D-2} s|$ [94, 95], so the boundary term is zero. Evaluating the contour integral along the axis yields

$$\begin{aligned}\left(\int_{-\infty}^{-s_0} + \int_{s_0}^{\infty}\right) \frac{\mathrm{d}s}{2\pi i} \frac{\mathrm{Disc}[\mathcal{M}(s)]}{s^3} &= -\int_{s_0}^{\infty} \frac{\mathrm{d}s}{2\pi i} \frac{\mathrm{Disc}[\mathcal{M}(-s)]}{s^3} \\ &\quad + \int_{s_0}^{\infty} \frac{\mathrm{d}s}{2\pi i} \frac{\mathrm{Disc}[\mathcal{M}(s)]}{s^3} \\ &= 2\int_{s_0}^{\infty} \frac{\mathrm{d}s}{2\pi i} \frac{\mathrm{Disc}[\mathcal{M}(s)]}{s^3}.\end{aligned} \tag{3.2.17}$$

Because the external states are identical, crossing symmetry implies that $\mathcal{M}(s + i\epsilon) = \mathcal{M}(-s - i\epsilon)$, so $\mathrm{Disc}[\mathcal{M}(-s)] = -\mathrm{Disc}[\mathcal{M}(s)]$. Inserting the

optical theorem, $\mathrm{Im}[\mathcal{M}(s)] = s\sigma(s)$, the dispersion relation becomes[3.3]

$$16a' = \frac{2}{\pi}\int_{s_0}^{\infty} \mathrm{d}s \frac{\sigma(s)}{s^2} \geq 0, \tag{3.2.18}$$

where $\sigma(s)$ is the total cross-section. In the last step we have used the fact that the total cross-section is non-negative, implying that $a' \geq 0$.

The above arguments apply provided that high-energy scattering amplitudes comply with the optical theorem, the Froissart bound, and the standard analyticity properties of the S-matrix. The first and third conditions hold under the assumptions of unitarity and locality, respectively, while the second requires both. In Ref. [96], it was noted that locality may break down when quantum gravitational dynamics become important; in particular, black holes may induce non-localities at super-Planckian energies, which violate the Froissart bound and the polynomial boundedness of amplitudes [96], albeit in unphysical regions of complex momentum space [97]. However, these caveats are immaterial because, as previously noted, black holes do not exist in asymptotically-flat 3D spacetime, so our arguments apply. In 4D, the issue is more complex, but we postpone a dedicated discussion to Sec. 3.3.2.

3.2.3 Unitarity (3D)

We now derive effective theory bounds by imposing unitarity on a general parameterization of the ultraviolet completion. Our analysis follows the approach of Ref. [35]. As a consequence of the shift symmetry of ϕ, the leading coupling to high-energy degrees of freedom is uniquely

$$\chi_{\mu\nu}\partial^\mu\phi\partial^\nu\phi, \tag{3.2.19}$$

where $\chi_{\mu\nu}$ is a field representing arbitrary ultraviolet dynamics. Integrating out these states generates the leading four-derivative operator, $(\partial_\mu\phi\partial^\mu\phi)^2$. By neglecting higher-order interactions of ϕ with $\chi_{\mu\nu}$, we are implicitly assuming

[3.3]Note that in 3D, $\mathcal{M}(s)$ and $\sigma(s)$ have mass dimensions $+1$ and -1, respectively.

a perturbative ultraviolet completion. Couplings of the form $\chi_\mu \partial^\mu \phi$ are also allowed in principle but can be eliminated via the transverse condition $\partial_\mu \chi^\mu = 0$. Moreover, couplings of the form $\partial_\mu \chi_\nu \partial^\mu \phi \partial^\nu \phi$ can be neglected because they produce subleading six-derivative operators.

We now decompose $\chi_{\mu\nu}$ into components,

$$\chi_{\mu\nu} = \chi^{(2)}_{\mu\nu} + \eta_{\mu\nu}\chi^{(0)}, \tag{3.2.20}$$

where $\chi^{(2)}_{\mu\nu}$ is by definition traceless. In our conventions, all coupling constants have been absorbed into the overall normalization of the fields, so the leading interactions are of unit strength, $\chi^{(2)}_{\mu\nu}\partial^\mu\phi\partial^\nu\phi + \chi^{(0)}\partial_\mu\phi\partial^\mu\phi$. Without loss of generality, the nonperturbative spectral representation of the $\chi_{\mu\nu}$ propagator in D dimensions is given by

$$\begin{aligned}
\langle \chi^{(0)}(k)\chi^{(0)}(k')\rangle &= i\delta^D(k+k')\int_0^\infty \mathrm{d}\mu^2\, \frac{\rho^{(0)}(\mu^2)}{k^2-\mu^2+i\epsilon} \\
\langle \chi^{(2)}_{\mu\nu}(k)\chi^{(2)}_{\rho\sigma}(k')\rangle &= i\delta^D(k+k')\int_0^\infty \mathrm{d}\mu^2\, \frac{\rho^{(2)}(\mu^2)}{k^2-\mu^2+i\epsilon}\Pi_{\mu\nu\rho\sigma},
\end{aligned} \tag{3.2.21}$$

where $\rho^{(0)}$ and $\rho^{(2)}$ are spectral densities describing an arbitrary collection of single- or multi-particle intermediate states. As usual, these expressions are obtained by inserting a complete set of states into the two-particle correlation function, implying positive definite spectral densities in the absence of tachyon or ghost instabilities. Note also that since we are ultraviolet-completing a local operator, i.e., one that is regular as $k \to 0$, the spectral density must vanish in the neighborhood of $\mu^2 = 0$.

As is well known, the spectral representation of a massive spin-2 state is strongly constrained by unitarity. In D dimensions, the absence of tachyons or ghosts implies that [38]

$$\Pi_{\mu\nu\rho\sigma} = \frac{1}{2}(\Pi_{\mu\sigma}\Pi_{\nu\rho} + \Pi_{\mu\rho}\Pi_{\nu\sigma}) - \frac{1}{D-1}\Pi_{\mu\nu}\Pi_{\rho\sigma}, \tag{3.2.22}$$

where for convenience we have defined the projection operator,

$$\Pi_{\mu\nu} = \eta_{\mu\nu} - k_\mu k_\nu/\mu^2, \tag{3.2.23}$$

such that $k^\mu \Pi_{\mu\nu} = 0$ when $k^2 = \mu^2$ is on-shell. Note that the transverse condition, $k^\mu \Pi_{\mu\nu\rho\sigma} = 0$, applies on-shell so as to eliminate gauge degrees of freedom. Not coincidentally, Eq. (3.2.22) is precisely the propagator numerator for a massive graviton.

At low momentum transfer we integrate out $\chi_{\mu\nu}$, yielding the 3D effective operator,

$$\chi_{\mu\nu}\partial^\mu\phi\partial^\nu\phi \to (\partial_\mu\phi\partial^\mu\phi)^2 \int_0^\infty \mathrm{d}\mu^2 \, \frac{\rho^{(0)}(\mu^2)/2 + \rho^{(2)}(\mu^2)/4}{\mu^2}. \tag{3.2.24}$$

Eq. (3.2.24) shows that the coefficient of $(\partial_\mu\phi\partial^\mu\phi)^2$ is positive for any weakly-coupled ultraviolet completion consistent with a positive spectral density. Thus, unitarity implies that $a' \geq 0$ for the effective Lagrangian defined in Eq. (3.2.8). Conversely, $a' < 0$ signals an instability coming from a tachyon or ghost intermediate state.

The above arguments apply assuming a perturbative ultraviolet completion of the effective theory. This allowed us to ignore operators involving ever higher powers of the field. As discussed in Sec. 3.2.2, while it may be problematic to extrapolate any argument to energies far above the Planck scale, this is not an issue in asymptotically-flat 3D spacetime, since black holes are not permitted.

3.2.4 Causality (3D)

Let us now investigate the causal structure of the 3D photon-graviton effective theory. We expand around nontrivial backgrounds for the photon and graviton,

$$\phi = \overline{\phi} + \varphi \qquad \text{and} \qquad g_{\mu\nu} = \overline{g}_{\mu\nu} + h_{\mu\nu}. \tag{3.2.25}$$

Throughout this chapter, any barred variable represents a field or combination of fields evaluated on its background value. Here φ denotes photon fluctuations,

while in 3D, the graviton is non-dynamical so $h_{\mu\nu} = 0$. To simplify our analysis we introduce vielbein coordinates defined by $\eta_{ab} = \overline{e}^{\mu}_{a}\overline{e}^{\nu}_{b}\overline{g}_{\mu\nu}$, where $\eta_{ab} = \mathrm{diag}(+1, -1, -1)$ is the flat space metric. We use Latin and Greek indices to denote vielbein and metric coordinates, respectively. Importantly, the speed measured in the vielbein frame corresponds to the physical speed measured by an observer in the coordinates of the local Lorentz frame. In terms of these coordinates, the equation of motion for φ in a background is

$$\tilde{\eta}^{ab}\partial_a\varphi\partial_b\varphi = 0, \tag{3.2.26}$$

where $\tilde{\eta}_{ab}$ is defined as the effective metric in the vielbein frame, obtained from Eq. (3.2.8),

$$\tilde{\eta}_{ab} = \eta_{ab} + 16a'\overline{\partial_a\phi\partial_b\phi}. \tag{3.2.27}$$

We study the geometric-optics limit in which φ is a plane wave perturbation of four-momentum $k_a = (k_0, \vec{k})$, with wavelength far shorter than the characteristic length scale of the spacetime curvature. In this case, the dispersion relation for the photon is simply

$$\tilde{\eta}^{ab}k_ak_b = 0. \tag{3.2.28}$$

For now, let us focus on the photon speed in a local neighborhood; we will consider the global effects of gravity shortly.

The local speed of photon fluctuations varies depending on the choice of background. The simplest possibility is a constant electromagnetic field, represented by a constant condensate that breaks Lorentz invariance: $\overline{\partial_a\phi} = w_a = (w_0, \vec{w})$. The effective metric is then $\tilde{\eta}_{ab} = \eta_{ab} + 16a'w_aw_b$. Expanding at leading order in the small parameter a', we obtain the propagation speed of photons,

$$v = \frac{k_0}{|\vec{k}|} = 1 - 8a'(w_0 - \vec{w}\cdot\hat{k})^2, \tag{3.2.29}$$

defining $\hat{k} = \vec{k}/|\vec{k}|$. Superluminal photon propagation occurs when $a' < 0$.

Another interesting background is a thermal gas of photons, which we consider henceforth. For a thermal system, background fields should be evaluated as stochastic expectation values, so in general $\overline{\partial_a\phi\partial_b\phi} \neq \overline{\partial_a\phi}\cdot\overline{\partial_b\phi}$. In particular, for a photon gas the electromagnetic field has zero average background value, $\overline{\partial_a\phi} = 0$, but nonzero variance, $\overline{\partial_a\phi\partial_b\phi} \neq 0$. In 3D, the pressure p and energy density ρ satisfy an equation of state $p = \rho/2$, where $\rho = \zeta(3)T^3/\pi$ for a gas at temperature T [98, 99]. For a scalar field, the energy-momentum tensor is

$$T^{ab} = \partial^a\phi\partial^b\phi - \frac{1}{2}\eta^{ab}\partial_c\phi\partial^c\phi, \tag{3.2.30}$$

the background expectation value of which is $\overline{T}^{ab} = \mathrm{diag}(\rho, p, p)$ in a thermal gas. From this we deduce that $\overline{\partial_c\phi\partial^c\phi} = -2\overline{T}^a{}_a = -2(\rho - 2p) = 0$, so

$$\overline{\partial_a\phi\partial_b\phi} = (3\delta^0_a\delta^0_b - \eta_{ab})\frac{\zeta(3)}{2\pi}T^3. \tag{3.2.31}$$

Putting everything together, we obtain the effective metric for photon propagation,

$$\tilde{\eta}_{ab} = \eta_{ab} + (3\delta^0_a\delta^0_b - \eta_{ab})\frac{8\zeta(3)}{\pi}a'T^3. \tag{3.2.32}$$

The presence of Kronecker delta functions signals the fact that a thermal background breaks Lorentz invariance while preserving isotropy. Expanding at leading order in a', we find that the speed of signal propagation is

$$v = \frac{k_0}{|\vec{k}|} = 1 - \frac{12\zeta(3)}{\pi}a'T^3. \tag{3.2.33}$$

As before, superluminal propagation occurs when $a' < 0$.

Traditionally, superluminal propagation is taken to be a definitive signal of an underlying pathology. However, this diagnosis neglects an important distinction between superluminal propagation in all reference frames versus a preferred frame. The present construction is of the latter type, which as discussed in Ref. [31] introduces oddities in the definition of initial conditions,

but is not, strictly speaking, inconsistent.

To demonstrate a true breakdown of causality, we must construct a closed signal trajectory in spacetime, i.e., a closed causal curve (CCC). To begin, consider a thermal gas of photons localized to a finite bubble in spacetime. The interior of the bubble is described by a zero-curvature, 3D Friedmann-Robertson-Walker (FRW) metric

$$\mathrm{d}s^2 = a(t)^2 \eta_{\mu\nu} \mathrm{d}x^\mu \mathrm{d}x^\nu, \tag{3.2.34}$$

written in a form that is manifestly conformally flat. Inside the bubble, photons deviate from the light-cone by an amount prescribed by the vielbein speed in Eq. (3.2.33). Meanwhile, the vacuum region exterior to the bubble is locally flat because the Weyl tensor vanishes identically in 3D. Consequently, photons are exactly luminal outside the bubble.

What about the boundary of the bubble? Since the interior and exterior spacetimes are conformally flat, regularity of the spacetime across the boundary implies that, in the thin-shell limit, the boundary region itself is parametrically close to conformal flatness. Moreover, one can imagine a boundary formed from "stiff" matter with $\rho = p$, for which the Cotton tensor vanishes [100], thus ensuring conformal flatness exactly.[3.4] In any case, a bubble of thermal photons is well described by a metric that is globally conformally flat,

$$\mathrm{d}s^2 = \Omega^2 \eta_{\mu\nu} \mathrm{d}x^\mu \mathrm{d}x^\nu, \tag{3.2.35}$$

where $\Omega = 1$ in the exterior and $\Omega = a(t)$ in the interior. A feature of conformal flatness is that the speeds of signals as measured in vielbein coordinates and metric coordinates are the same. That is, light signals move at speed $v = \mathrm{d}x/\mathrm{d}t$, where v is given by Eq. (3.2.33).[3.5] In the end, this implies that engineering a

[3.4] The signal itself can be transferred across the boundary either by another particle species that does not interact with the boundary material or by photons through a very small aperture in the circular shell.

[3.5] This is in marked contrast to the gravitational redshift of signals in spacetimes that are

CCC in a conformally-flat spacetime reduces to a special relativistic problem. As is well known, however, a CCC in special relativity requires two frames in relative motion, while the above construction picks out a single preferred frame. To build a CCC we must instead consider two bubbles of thermal photons, both at temperature T and in relative motion. The associated background is described by Eq. (3.2.35), only with a more complicated form for the conformal factor.

Now consider the setup illustrated in Fig. 3.2: two bubbles of equal radii ℓ separated by a distance L and moving in opposite directions at zero impact parameter and constant speed u. Light signals sent between observers at the center of each bubble will have an average speed

$$v_{\text{avg}} = 1 - \epsilon, \tag{3.2.36}$$

as measured in their respective frames. Here, corrections to the speed of light are controlled by a small parameter, $\epsilon \sim a' T^3 \ell / L$. Note that the Friedmann equations imply that the interior of each bubble will evolve on a timescale $\sim \rho^{-1/2}$ in natural units. However, these effects can be neglected by choosing $L^2 T^3 \ll 1$, which is always possible for sufficiently small T. Consequently, we can always treat the temperature as roughly constant over the entire signal trip.

For $a' < 0$, it is then straightforward to construct a CCC. Explicitly, each observer can send a signal that in the reference frame of the other observer propagates at an average superluminal speed defined by Eq. (3.2.36). By transmitting a signal from one bubble to the other and then back, it is possible to form a CCC. This is analogous to the so-called "tachyonic antitelephone" from special relativity [101–103]. Likewise, causality violation will occur here provided the relative speed of the two observers (i.e., the relative speed of the

not conformally flat, such as a signal propagating radially away from a black hole.

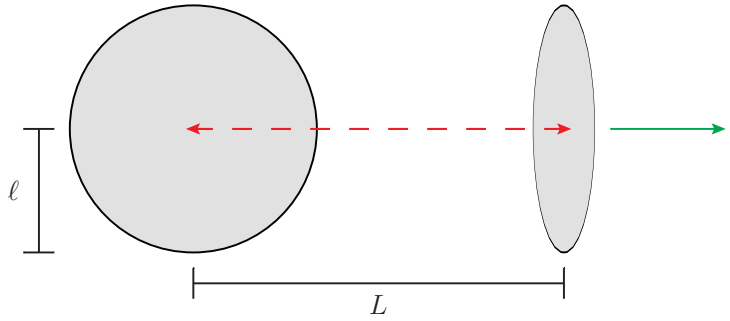

Figure 3.2. Setup for the construction of a CCC in 3D, using superluminal photons in a theory that violates Eq. (3.2.9). The construction, illustrated here in a constant-time slice of the two spatial dimensions, consists of two circular bubbles of thermal radiation, each of radius ℓ and separation $L \gg \ell$, with relative speed $u < 1$ (green arrow). Signals (red dashed arrows) sent back and forth would be superluminal within the bubbles, creating a CCC for large u.

bubbles) satisfies

$$u > \frac{2v_{\rm avg}}{1+v_{\rm avg}^2} \simeq 1 - \frac{1}{2}\epsilon^2. \tag{3.2.37}$$

A diagram of this CCC is nicely depicted in Fig. 2 of Ref. [31], albeit in a slightly different context (Lorentz-violating condensate bubbles passing with finite impact parameter) and without including the effects of gravity. Forbidding the existence of causality violation from a CCC thus requires $a' \geq 0$.

The above arguments apply provided there is no subtlety in constructing this particular background of thermal photons. Naively, one may worry about exceeding the limits of the photon-graviton effective theory due to the relative boost between the bubbles of gas. However, this is not an issue because the bubbles need not overlap and hence do not back-react. While arbitrary configurations of moving masses in 3D sometimes entail topological subtleties [104–106], our CCC construction does not rely on them for causality violation. A detailed study of these issues goes beyond the scope of the current work.

3.3 Four Dimensions

3.3.1 Setup and Bounds (4D)

In this section, we derive bounds on the photon-graviton effective action in 4D. As in Sec. 3.2.1, we rewrite the spacetime curvature in terms of the electromagnetic field strength. To start, we eliminate $R_{\mu\nu\rho\sigma}R^{\mu\nu\rho\sigma}$ from Eq. (3.1.2) in favor of the 4D Gauss-Bonnet term,

$$R^2 - 4R_{\mu\nu}R^{\mu\nu} + R_{\mu\nu\rho\sigma}R^{\mu\nu\rho\sigma}, \tag{3.3.1}$$

which is in turn a total derivative. We also use the definition of the Weyl tensor in Eq. (3.2.1) to rewrite the operator $F_{\mu\nu}F_{\rho\sigma}R^{\mu\nu\rho\sigma}$ in terms of $F_{\mu\nu}F_{\rho\sigma}C^{\mu\nu\rho\sigma}$, $F_{\mu\rho}F_{\nu}{}^{\rho}R^{\mu\nu}$, and $F_{\mu\nu}F^{\mu\nu}R$. Next, we substitute the energy-momentum tensor (3.2.4) in for the Ricci scalar and Ricci tensor in the higher-dimension operators using the tree-level Einstein field equations (3.2.3), which at the present order in couplings is again equivalent to a field redefinition. With the useful identity in 4D,

$$2(F_{\mu\nu}F^{\mu\nu})^2 + (F_{\mu\nu}\tilde{F}^{\mu\nu})^2 = 4F_{\mu\rho}F_{\nu}{}^{\rho}F^{\mu}{}_{\sigma}F^{\nu\sigma}, \tag{3.3.2}$$

we obtain our final form for the effective Lagrangian,

$$\mathcal{L} = -\frac{1}{4}F_{\mu\nu}F^{\mu\nu} - \frac{1}{4}R + a_1'(F_{\mu\nu}F^{\mu\nu})^2 + a_2'(F_{\mu\nu}\tilde{F}^{\mu\nu})^2 + b_3F_{\mu\nu}F_{\rho\sigma}C^{\mu\nu\rho\sigma}, \tag{3.3.3}$$

where we have defined new higher-dimension operator coefficients,

$$a_1' = a_1 - b_2/2 - b_3 + c_2 + 4c_3 \qquad \text{and} \qquad a_2' = a_2 - b_2/2 - b_3 + c_2 + 4c_3. \tag{3.3.4}$$

In Eq. (3.3.3), all explicit curvature dependence has been removed except for the Weyl tensor, which in 4D is nontrivial. In the classical theory, the Weyl tensor represents the component of the gravitational field that propagates freely in the absence of sources and thus decouples from matter at leading order in

Einstein's equations. Later, we will see how this is manifested in the forward scattering amplitudes, which at leading order are explicitly dependent on a_i' but not b_3.

Constraining the parameters in Eq. (3.3.3) using analyticity, unitarity, and causality is substantially more difficult in 4D due to dynamical gravity. We will elaborate on these arguments later on, but let us briefly collect our final results here. We derive bounds coming from unitarity:

$$a_1' \geq 0 \qquad \text{and} \qquad a_2' \geq 0, \tag{3.3.5}$$

while the absence of superluminality in certain backgrounds implies that

$$a_1' + a_2' \geq 0. \tag{3.3.6}$$

Interestingly, if one blithely applies analyticity arguments to the higher-dimension operator contributions, one also obtains Eq. (3.3.6). Just as in Sec. 3.2.1, it is convenient to expand a_i' in terms of their contributions from electromagnetic and gravitational interactions:

$$a_i' = \alpha_i z^4 + \beta_i z^2 + \gamma_i, \tag{3.3.7}$$

where α_i, β_i, and γ_i are generated by diagrams like the ones shown in Fig. 3.1. Contributions coming from integrating out a charged fermion [93] or charged scalar [107–109] are

$$\begin{aligned}
(a_1, a_2) &= \left(\frac{q^4}{1440\pi^2 m^4},\ \frac{7q^4}{5760\pi^2 m^4}\right) && \text{[fermion]} \\
(b_1, b_2, b_3) &= \left(-\frac{q^2}{576\pi^2 m^2},\ \frac{13q^2}{1440\pi^2 m^2},\ -\frac{q^2}{1440\pi^2 m^2}\right) && \text{[fermion]} \\
(a_1, a_2) &= \left(\frac{7q^4}{23040\pi^2 m^4},\ \frac{q^4}{23040\pi^2 m^4}\right) && \text{[scalar]} \\
(b_1, b_2, b_3) &= \left(\frac{q^2}{1152\pi^2 m^2},\ \frac{q^2}{1440\pi^2 m^2},\ \frac{q^2}{2880\pi^2 m^2}\right) && \text{[scalar]},
\end{aligned} \tag{3.3.8}$$

where for the scalar we have assumed minimal coupling to gravity. Given these coefficients, the unitarity bounds in Eq. (3.3.5) imply that

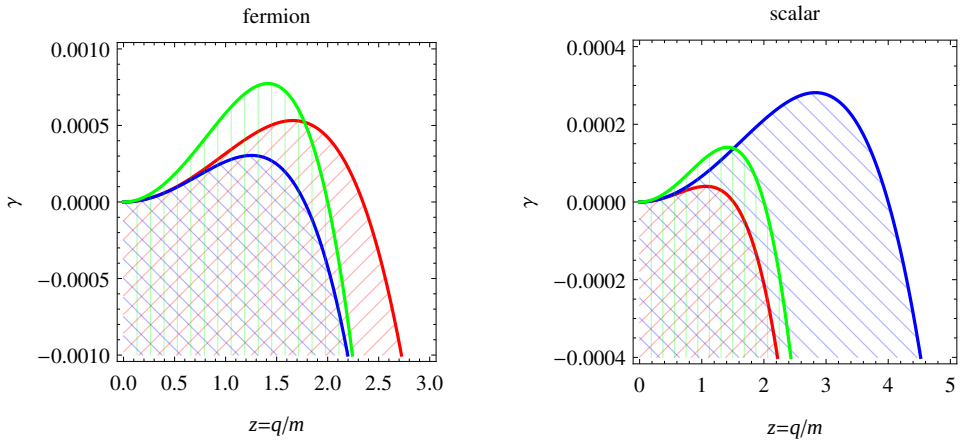

Figure 3.3. Bounds on the 4D photon-graviton effective theory derived from integrating out a fermion (left) or scalar (right) and expressed in terms of the contributions coming from electromagnetism (parameterized by $z = q/m$) and pure gravity (parameterized by γ). The cross-hatched regions are forbidden by arguments from unitarity, which apply to $\gamma = \gamma_1$ (red ◪) and $\gamma = \gamma_2$ (blue ◪), and arguments from analyticity and superluminality, which both apply to $\gamma = \gamma_1 + \gamma_2$ (green ◫). The WGC forbids $z < 1$, which overlaps with much of the region also forbidden by infrared consistency.

$$\begin{aligned} z^2\left(z^2 - 11/2\right) &\geq -\gamma_1 \times 1440\pi^2 && \text{[fermion]} \\ z^2\left(z^2 - 22/7\right) &\geq -\gamma_2 \times 5760\pi^2/7 && \text{[fermion]} \\ z^2\left(z^2 - 16/7\right) &\geq -\gamma_1 \times 23040\pi^2/7 && \text{[scalar]} \\ z^2\left(z^2 - 16\right) &\geq -\gamma_2 \times 23040\pi^2 && \text{[scalar]}, \end{aligned} \tag{3.3.9}$$

while the bounds from analyticity and superluminality in Eq. (3.3.6) are

$$\begin{aligned} z^2\left(z^2 - 4\right) &\geq -(\gamma_1 + \gamma_2) \times 5760\pi^2/11 && \text{[fermion]} \\ z^2\left(z^2 - 4\right) &\geq -(\gamma_1 + \gamma_2) \times 2880\pi^2 && \text{[scalar]}. \end{aligned} \tag{3.3.10}$$

Curiously, for small values of γ_i, both fermions and scalars in 4D are subject to the same bound:

$$z \geq 2. \tag{3.3.11}$$

All of our 4D constraints are summarized in Fig. 3.3. As in Sec. 3.2.1, the coefficients γ_i parameterize all corrections coming from purely gravitational interactions. In 4D, this includes the contribution from $c_2 + 4c_3$ in Eq. (3.3.4),

which runs logarithmically due to graviton loops [110] and is thus controlled by an ultraviolet-sensitive boundary condition. As in 3D, for sufficiently large values of γ_i these bounds are automatically satisfied. Alternatively, we can consider the case where the purely Planck-suppressed corrections are negligible, in which case γ_i is small and our infrared consistency conditions bound z strictly from below.

It is reasonable to assume that a theory that satisfies our consistency conditions, Eqs. (3.3.9) and (3.3.10), at a given energy scale will continue to do so deeper into the infrared. Interestingly, this implies that γ_i should not decrease in the infrared, i.e., the sign of the beta function for γ_i should be negative on general grounds. This is confirmed by explicit computation of the one-loop divergences in the photon-graviton effective theory [110].

3.3.2 Analyticity (4D)

Let us endeavor to apply the analyticity argument of Sec. 3.2.2 to light-by-light scattering in 4D. Using Eq. (3.3.3), we read off the Feynman rules for the photon-graviton theory: there is the usual photon-photon-graviton vertex from the Einstein-Maxwell terms, a higher-order photon-photon-graviton vertex from the b_3 term, and new quartic photon vertices from the a_1' and a_2' terms. Putting these together, we find that the tree-level amplitudes for certain helicity states are

$$\begin{aligned}
\mathcal{M}(1^+2^+3^+4^+) &= \mathcal{M}(1^-2^-3^-4^-) = 8(a_1' - a_2')(s^2+t^2+u^2) \\
\mathcal{M}(1^+2^+3^-4^-) &= \mathcal{M}(1^-2^-3^+4^+) = \frac{2s^4}{stu} + 8(a_1' + a_2')s^2 \\
\mathcal{M}(1^+2^-3^+4^-) &= \mathcal{M}(1^-2^+3^-4^+) = \frac{2t^4}{stu} + 8(a_1' + a_2')t^2 \\
\mathcal{M}(1^+2^-3^-4^+) &= \mathcal{M}(1^-2^+3^+4^-) = \frac{2u^4}{stu} + 8(a_1' + a_2')u^2.
\end{aligned} \tag{3.3.12}$$

Here, we have used a helicity basis defined with all momenta incoming, so

the second and fourth lines of Eq. (3.3.12) correspond to forward amplitudes. Notably, at leading order in the higher-dimension operator coefficients, all forward amplitudes depend explicitly on a_i' but not b_3, which controls the irreducible interactions between the electromagnetic field strength and the Weyl tensor. This is quite reasonable on physical grounds because, in the classical limit, the Weyl tensor does not have a minimal coupling to the energy-momentum tensor. Quantum mechanically, this is manifested as the fact that the Weyl tensor mediates forward light-by-light scattering only at second order in b_3, i.e., coming from two insertions of the higher-dimension operator.

Critically, the scattering amplitudes have terms that are singular in the s-, t-, and u-channels due to leading-order graviton exchange. In the forward limit, the t-channel diagrams scale as $\sim s^2/t$ and formally diverge at forward scattering. In this limit, the partial wave expansion does not converge, the Froissart bound is invalid, and the dispersion relation reasoning from Sec. 3.2.2 does not apply. Hence, dynamical gravity creates a considerable obstacle to any argument from analyticity [31].

There is no immediate justification for simply dropping these singular contributions. Nevertheless, it is interesting to compute the bound that would arise from applying the analyticity argument of Sec. 3.2.2 to the non-singular contributions coming from higher-dimension operators. Notably, a crucial ingredient of the analyticity argument is the requirement that contributions to the scattering amplitude from ultraviolet dynamics be even in s. As a result, contributions to the dispersion relation from negative s can be directly related to the cross-section at positive s. In 3D, this was automatically satisfied because the scattering amplitude was a crossing-symmetric function characterizing indistinguishable scalars. In contrast, the 4D scattering amplitudes describe photons with distinguishable helicity labels. To form an object suitable for

analyticity bounds, we consider the sum of all forward amplitudes, $\mathcal{M}_{\text{sym}}$, which is by construction symmetric under the exchange of $1 \leftrightarrow 3$ and $2 \leftrightarrow 4$:

$$\begin{aligned}\mathcal{M}_{\text{sym}} &= \mathcal{M}(1^+2^+3^-4^-) + \mathcal{M}(1^-2^-3^+4^+) \\ &\qquad + \mathcal{M}(1^+2^-3^-4^+) + \mathcal{M}(1^-2^+3^+4^-) \\ &= \frac{4(s^4+u^4)}{stu} + 16(a_1'+a_2')(s^2+u^2) \\ &\overset{t\to 0}{=} -\frac{8s^2}{t} - 8s + 32(a_1'+a_2')s^2 + \mathcal{O}(t).\end{aligned} \tag{3.3.13}$$

The first two terms of the last line come from single graviton exchange due to the Einstein-Maxwell photon-photon-graviton vertex. If we drop this contribution, then the dispersion relation argument of Sec. 3.2.2 implies that the coefficient of s^2 in $\mathcal{M}_{\text{sym}}$ is non-negative, so

$$a_1' + a_2' \geq 0. \tag{3.3.14}$$

Because the t-channel graviton singularity remains a critical obstruction to this argument, the inequality in Eq. (3.3.14) should not yet be considered a rigorous bound. Nevertheless, it has been noted that singular contributions can be consistently subtracted from a dispersion relation [31, 111], provided the theory has a weak coupling parameter that can discriminate between the contribution from leading-order exchange of massless particles and that from higher-dimension operators. For the photon-graviton effective action, the natural choice for a weak coupling parameter is the gravitational constant, G. However, by sending $G \to 0$, we also eliminate the very higher-dimension, gravitationally-induced interactions that we seek to bound. Thus, we have not identified such a weak coupling parameter here, though it may be possible. More generally, it may be feasible to extract rigorous effective theory bounds from theories with t-channel singularities, but we leave this formidable task for future work.

As discussed in Sec. 3.2.2, the analyticity argument involves additional

subtleties related to taking a contour in the complex s plane to super-Planckian scales, which a priori could involve issues with black hole formation and associated non-localities. However, as $t \to 0$, the impact parameter exceeds the Schwarzschild radius for the scattering particles, implying no black hole production in the forward limit [97, 112]. Pathologies associated with nonperturbative gravitational interactions are thus avoided. In any case, the same assumptions used in our analyticity bounds, which were mentioned in Sec. 3.2.2, have been used previously to constrain string theories from low-energy scattering [31, 113]. In general, this is justified because string amplitudes are analytic and highly convergent at large s [114, 115].

3.3.3 Unitarity (4D)

Next, let us apply the unitarity argument of Sec. 3.2.3 to 4D. In principle, one can define general spectral representations parameterizing the ultraviolet-completing dynamics of $(F_{\mu\nu}F^{\mu\nu})^2$, $(F_{\mu\nu}\tilde{F}^{\mu\nu})^2$, and $F_{\mu\nu}F_{\rho\sigma}C^{\mu\nu\rho\sigma}$. The only substantive difference from the 3D case is the third operator, which depends on the spacetime curvature in a way that cannot simply be eliminated using Einstein's equations. In what follows, we will be interested in bounding the coefficients of the first and second operators.

At leading order, the photon couples to the ultraviolet states according to

$$F^{\mu\nu}F^{\rho\sigma}\chi_{\mu\nu\rho\sigma} \qquad \text{and} \qquad F^{\mu\nu}\tilde{F}^{\rho\sigma}\psi_{\mu\nu\rho\sigma}, \tag{3.3.15}$$

where $\chi_{\mu\nu\rho\sigma}$ and $\psi_{\mu\nu\rho\sigma}$ are parity-even and -odd fields that couple to the photon. Note that these fields have the skew and interchange index symmetries of the Riemann tensor: $\chi_{\mu\nu\rho\sigma} = -\chi_{\nu\mu\rho\sigma} = -\chi_{\mu\nu\sigma\rho}$ and $\chi_{\mu\nu\rho\sigma} = \chi_{\rho\sigma\mu\nu}$ and similarly for $\psi_{\mu\nu\rho\sigma}$. As in Sec. 3.2.3, $\chi_{\mu\nu\rho\sigma}$ and $\psi_{\mu\nu\rho\sigma}$ parameterize an arbitrary set of intermediate single- or multi-particle states, so our unitarity argument remains quite general.

While there can also exist couplings of the form $\chi_{\mu\nu}F^{\mu\nu}$, they can be eliminated by the transverse condition, $\partial_\mu\chi^{\mu\nu}=0$. Likewise, couplings of the form $\partial_\mu\partial_\nu\partial_\rho\chi_\sigma F^{\mu\nu}F^{\rho\sigma}$ and $\partial_\mu\chi_{\nu\rho\sigma}F^{\mu\nu}F^{\rho\sigma}$ need not be considered because they yield operators that are of higher order in the derivative expansion. Because the photon-graviton effective action includes the operator $F_{\mu\nu}F_{\rho\sigma}C^{\mu\nu\rho\sigma}$, it is also possible, in principle, that $\chi_{\mu\nu\rho\sigma}$ couples directly to $C_{\mu\nu\rho\sigma}$. However, as shown in Sec. 3.3.2, interactions mediated through the Weyl tensor do not affect the low-energy forward scattering amplitudes at leading order in the higher-dimension operator coefficients. Hence, at this order, any coupling between $\chi_{\mu\nu\rho\sigma}$ and $C_{\mu\nu\rho\sigma}$ cannot contribute to the coefficients of $(F_{\mu\nu}F^{\mu\nu})^2$ and $(F_{\mu\nu}\tilde{F}^{\mu\nu})^2$ and can be neglected.

As before, we expand $\chi_{\mu\nu\rho\sigma}$ into its components,

$$\chi_{\mu\nu\rho\sigma}=\chi^{(4)}_{\mu\nu\rho\sigma}+\frac{1}{4}(\eta_{\mu[\rho}\chi^{(2)}_{\sigma]\nu}-\eta_{\nu[\rho}\chi^{(2)}_{\sigma]\mu})+\frac{1}{2}\chi^{(0)}\eta_{\mu[\rho}\eta_{\sigma]\nu}, \tag{3.3.16}$$

and similarly for $\psi_{\mu\nu\rho\sigma}$, where $\chi^{(2)}_{\mu\nu}$ and $\chi^{(4)}_{\mu\nu\rho\sigma}$ are by definition traceless. Also as in Sec. 3.2.3, we choose a normalization in which all coupling constants are absorbed into the fields and the photon interacts via $\chi^{(4)}_{\mu\nu\rho\sigma}F^{\mu\nu}F^{\rho\sigma}+\chi^{(2)}_{\mu\nu}F^{\mu}_{\ \rho}F^{\nu\rho}+\chi^{(0)}F_{\mu\nu}F^{\mu\nu}$.

The spectral decompositions for $\chi^{(0)}$ and $\chi^{(2)}_{\mu\nu}$ are the same as in Eq. (3.2.21), while for $\chi^{(4)}_{\mu\nu\rho\sigma}$,

$$\langle\chi^{(4)}_{\mu\nu\rho\sigma}(k)\chi^{(4)}_{\alpha\beta\gamma\delta}(k')\rangle=i\delta^D(k+k')\int_0^\infty \mathrm{d}\mu^2\ \frac{\rho^{(4)}(\mu^2)}{k^2-\mu^2+i\epsilon}\Pi_{\mu\nu\rho\sigma\alpha\beta\gamma\delta}, \tag{3.3.17}$$

where $\rho^{(4)}$ is the spectral function for the four-index state. A priori, the tensor numerator $\Pi_{\mu\nu\rho\sigma\alpha\beta\gamma\delta}$ consists of arbitrary combinations of $\eta_{\mu\nu}$ and k_μ; however, it is actually very constrained. By construction, $\Pi_{\mu\nu\rho\sigma\alpha\beta\gamma\delta}$ is traceless with index (anti-)symmetry properties consistent with those of $\chi^{(4)}_{\mu\nu\rho\sigma}$. In addition, just as for the spin-2 case, there are general arguments that fix the form of

$\Pi_{\mu\nu\rho\sigma\alpha\beta\gamma\delta}$. As discussed in Refs. [116, 117], the tensor numerators of higher-spin propagators are functions of the projection operator $\Pi_{\mu\nu}$ defined in Eq. (3.2.23). This ensures that the transverse condition $k^\mu \Pi_{\mu\nu\rho\sigma\alpha\beta\gamma\delta} = 0$ applies on-shell. This is analogous to the usual transverse conditions required for theories of massive higher-spin fields. We have checked that the only projection operator that satisfies the requisite trace, index symmetry, and transverse conditions can indeed be written in terms of combinations of $\Pi_{\mu\nu}$ and is moreover comprised of two such linearly independent tensor structures, shown in Eqs. (3.A.1) and (3.A.2) of App. 3.A. Last of all, unitarity implies that [20]

$$\Pi_{\mu\nu\rho\sigma\alpha\beta\gamma\delta} = \sum_i \varepsilon_{i\mu\nu\rho\sigma} \varepsilon^*_{i\alpha\beta\gamma\delta}, \tag{3.3.18}$$

so the tensor numerator is equal to the sum over polarization tensors labeled by i, with normalization $\varepsilon_{i\mu\nu\rho\sigma}\varepsilon_j^{*\mu\nu\rho\sigma} = \delta_{ij}$. However, the tensor numerator shown in Eqs. (3.A.1) and (3.A.2) of App. 3.A identically satisfies $\Pi_{\mu\nu\rho\sigma}{}^{\mu\nu\rho\sigma} = 0$, indicating that $\chi^{(4)}_{\mu\nu\rho\sigma}$ carries states of negative norm. Thus, we conclude that $\chi^{(4)}_{\mu\nu\rho\sigma}$ is unphysical and should be eliminated altogether.

Nonetheless, $\chi^{(0)}$ and $\chi^{(2)}_{\mu\nu}$ are still propagating and unitarity dictates that their spectral functions $\rho^{(0)}$ and $\rho^{(2)}$ be positive. At low energies, integrating them out yields

$$\begin{aligned} \chi_{\mu\nu\rho\sigma} F^{\mu\nu} F^{\rho\sigma} \to (F_{\mu\nu}F^{\mu\nu})^2 \int_0^\infty \mathrm{d}\mu^2 \, \frac{\rho^{(0)}/2 + \rho^{(2)}/12}{\mu^2} \\ + (F_{\mu\nu}\tilde{F}^{\mu\nu})^2 \int_0^\infty \mathrm{d}\mu^2 \, \frac{\rho^{(2)}/8}{\mu^2}. \end{aligned} \tag{3.3.19}$$

Thus, the contributions to $(F_{\mu\nu}F^{\mu\nu})^2$ and $(F_{\mu\nu}\tilde{F}^{\mu\nu})^2$ are both positive.

An analogous argument applies to the parity-odd field, $\psi_{\mu\nu\rho\sigma}$. To see this, we define

$$\psi_{\mu\nu\rho\sigma} F^{\mu\nu} \tilde{F}^{\rho\sigma} = \tilde{\chi}_{\mu\nu\rho\sigma} F^{\mu\nu} F^{\rho\sigma}, \tag{3.3.20}$$

where $\tilde{\chi}_{\mu\nu\rho\sigma} = \epsilon^{\alpha\beta}{}_{\rho\sigma}\psi_{\mu\nu\alpha\beta}/2$ is a parity-even field with the exact same symme-

tries as $\chi_{\mu\nu\rho\sigma}$. Running through the same logic as above implies that integrating out $\psi_{\mu\nu\rho\sigma}$ induces positive coefficients for $(F_{\mu\nu}F^{\mu\nu})^2$ and $(F_{\mu\nu}\tilde{F}^{\mu\nu})^2$. Putting it all together, we find that unitarity implies

$$a_1' \geq 0 \qquad \text{and} \qquad a_2' \geq 0 \tag{3.3.21}$$

for a weakly-coupled ultraviolet completion free of ghosts or tachyons.

3.3.4 Causality (4D)

We now turn to the problem of calculating the speed of photon propagation in a nontrivial 4D background. As before, we implement perturbation theory around a background electromagnetic and gravitational field,

$$A_\mu = \overline{A}_\mu + a_\mu, \qquad g_{\mu\nu} = \overline{g}_{\mu\nu} + h_{\mu\nu}, \tag{3.3.22}$$

where the graviton is fully dynamical in 4D. Similarly, the electromagnetic field strength can be expanded as $F_{\mu\nu} = \overline{F}_{\mu\nu} + f_{\mu\nu}$, with $f_{\mu\nu} = \overline{\nabla}_\mu a_\nu - \overline{\nabla}_\nu a_\mu = \partial_\mu a_\nu - \partial_\nu a_\mu$, where the final equality follows from the cancellation of the connection coefficients in the covariant derivatives.

Expanding perturbatively in the photon is straightforward for $(F_{\mu\nu}F^{\mu\nu})^2$ and $(F_{\mu\nu}\tilde{F}^{\mu\nu})^2$, but a slight subtlety arises for $F_{\mu\nu}F_{\rho\sigma}C^{\mu\nu\rho\sigma}$. In particular, this operator carries dependence on graviton fluctuations, which naively can be eliminated in favor of the photon using the linearized Einstein field equations. However, as discussed in Secs. 3.3.2 and 3.3.3, this does not actually happen because the Weyl tensor does not couple minimally to the energy-momentum tensor. Thus, the graviton dependence in $F_{\mu\nu}F_{\rho\sigma}C^{\mu\nu\rho\sigma}$ can be dropped, although this operator still contributes to the photon dispersion relation through the Weyl tensor background value, $\overline{C}_{\mu\nu\rho\sigma}$. This is nicely consistent with the analyticity arguments of Sec. 3.3.2 because of the close relationship between light-by-light scattering and the propagation of photons in a fixed electromagnetic background [31].

Let us consider a photon fluctuation described by a plane wave with circular polarization ε_a and momentum k_a. Throughout this chapter, we work in Lorenz gauge, $k_a\varepsilon^a = 0$. As before, we go to a geometric-optics limit in which the wavelength of the photon is far shorter than the typical scale of spacetime curvature [118]. In this regime, the dispersion relation is

$$\tilde{\eta}^{ab}k_a k_b = 0, \tag{3.3.23}$$

where at leading order in the couplings a_i' and b_3 the effective metric is

$$\tilde{\eta}_{ab} = \eta_{ab} + 32\left(a_1'\overline{F_{ac}F_{bd}} + a_2'\overline{\tilde{F}_{ac}\tilde{F}_{bd}}\right)\varepsilon^{c*}\varepsilon^d + 8b_3\overline{C}_{acbd}\varepsilon^{c*}\varepsilon^d. \tag{3.3.24}$$

Since the speed of propagation depends on the photon polarization, nontrivial electromagnetic fields induce birefringence.

In analogy with Sec. 3.2.4, it is natural to consider a constant electromagnetic background, $\overline{F}_{ab} \neq 0$, defined in vielbein coordinates. However, an additional complication arises due to dynamical gravity: a nontrivial electromagnetic background induces photon-graviton mixing of the form $\overline{F}_a{}^c f_{bc}h^{ab}$. This effect has been neglected in the literature on higher-order corrections to the photon dispersion relation [93, 119], most likely because it is Planck-suppressed. However, these corrections can easily dominate over contributions from higher-dimension operators in the photon-graviton effective action. For example, in the range where the WGC is marginally satisfied, m/q is of order the Planck scale and the effects of photon-graviton mixing will dwarf those of the higher-dimension operators.

To sidestep the issue of photon-graviton mixing, we focus on a background of thermal photons at temperature T. Since the background field values are thermally averaged, $\overline{F_{ab}F_{cd}} \neq \overline{F}_{ab}\cdot\overline{F}_{cd}$. In particular, for a photon gas, the electromagnetic field has zero average value, $\overline{F}_{ab} = 0$, but nonzero variance, $\overline{F_{ab}F_{cd}} \neq 0$. Photon-graviton mixing is identically zero because it scales as a

single power of $\overline{F}_{ab}$. Strictly speaking, this applies to quanta at wavelengths longer than $\sim 1/T$, so the effects of the background photon gas can be coarse-grained on scales relevant to photon-graviton mixing. In practice, this allows us to discard all terms in the action that are odd in the background field strength, $\overline{F}_{ab}$. In this regime, the photon and graviton propagate independently, albeit with a modified dispersion relation induced by the ambient photon gas. To calculate the photon dispersion relation, we then simply extract the part of the effective action (3.3.3) that is quadratic in the photon fluctuation. Note that while the energy of the propagating photon that we consider is, by construction, less than the temperature, the wavelength can still easily be much shorter than the typical scale of spacetime curvature induced by the photon gas. The thermal background sources a conformally-flat FRW metric, which acts effectively as flat space for photon propagation at leading order due to classical conformal invariance of electromagnetism in 4D; in any case, just as in Sec. 3.2.4, a conformally-flat metric in any dimension reduces the question of causality to a special relativistic problem, since coordinate speeds and vielbein speeds coincide.

In 4D, the energy density ρ and pressure p are related by $p = \rho/3$, where $\rho = \pi^2 T^4/15$. Using the fact that $\overline{T}^{ab} = \mathrm{diag}(\rho, p, p, p)$ together with Eq. (3.2.4), we find the simple expression

$$\overline{F_{ab}F_{cd}} = \overline{\tilde{F}_{ab}\tilde{F}_{cd}} = \frac{\pi^2}{45}T^4(\delta_{ac}\delta_{bd} - \delta_{ad}\delta_{bc}), \tag{3.3.25}$$

where δ_{ab} is again the Kronecker delta function. As in Eq. (3.2.31), Eq. (3.3.25) breaks Lorentz invariance due to the existence of the preferred rest frame of the photon gas. Inputting this expression into the effective metric (3.3.24), we find

$$v = \frac{k_0}{|\vec{k}|} = 1 - \frac{32\pi^2}{45}(a_1' + a_2')T^4, \tag{3.3.26}$$

independent of the direction of propagation or polarization, where we have used that $\overline{C}_{abcd} = 0$ because the background FRW metric is conformally flat. In the limit that gravity is decoupled, our expression for the photon velocity agrees with Ref. [120], which considered a thermal photon background in flat space. Note, however, that our formula does not agree with Ref. [93], which computed the photon velocity in a FRW universe but neglected to include the corrections coming from $(F_{\mu\nu}F^{\mu\nu})^2$ and $(F_{\mu\nu}\tilde{F}^{\mu\nu})^2$. In conclusion, we require that

$$a_1' + a_2' \geq 0 \tag{3.3.27}$$

to forbid superluminal propagation within the photon gas.

The relationship between superluminality and causality violation is, however, quite subtle in curved spacetime. A famous example is the seminal work of Ref. [93], which computed the speed of photons near a Schwarzschild black hole, taking into account corrections from the gravitational Euler-Heisenberg Lagrangian obtained by integrating out the electron. Curiously, the authors of Ref. [93] found that orbitally-traversing photons polarized in the radial direction propagate superluminally. However, this superluminal propagation cannot be an authentic signal of causality violation since the theory is literally real-world electrodynamics. While there is no universally-accepted resolution to this puzzle, it is important to note that an explicit CCC was not constructed in Ref. [93].[3.6] Despite the existence of local superluminal propagation, it is therefore clear that spacetime curvature can compensate for these effects in such a way that actual information flow remains causal. This is a prime example of the fallacy of interpreting superluminality as a telltale sign of acausal signal

[3.6] It has been argued (see Ref. [121] and references therein) that the superluminality derived in Ref. [93] is harmless because causality is dictated by high-frequency photon modes that lie outside the regime of the photon-graviton effective theory. However, this interpretation implies non-analyticity of the photon propagator and violation of the Kramers-Kronig dispersion relation.

propagation.

Our ideal goal is then to engineer a CCC in 4D that is analogous to the construction in Sec. 3.2.4, consisting of two bubbles of thermal photon gas in relative motion. However, since 4D gravity is dynamical, a non-vanishing Weyl tensor is induced in the vacuum region exterior to the photon gas. As shown in Eq. (3.2.37), if photons are only slightly superluminal, then a CCC requires a huge relative boost. In turn, the curvature outside the bubbles will be large and thus important for the propagation of photons during their traversal between the bubbles. Indeed, these metric effects will generally dominate over those induced by higher-dimension operators in the effective action. In addition, at such large relative boosts, it is no longer a good approximation to treat the bubbles as independent because they back-react. Of course, none of these effects arise in 3D, where the metric is locally flat in vacuum. Nonetheless, as we shall see, superluminal photon propagation can be linked to sharp pathologies via more elaborate constructions involving black holes.

In particular, consider a Schwarzschild black hole in the Hartle-Hawking vacuum [122]. This describes a black hole in equilibrium with an exterior thermal bath, so the event horizon is static.[3.7] Outside the black hole, the energy-momentum tensor is approximately described by a thermal gas at Hawking temperature T. For a sufficiently massive black hole, T can easily lie below the cutoff of the photon-graviton effective theory. The thermal background outside the black hole causes the speed of light to vary in accordance with our earlier discussion of FRW. However, there is an additional subtlety here in that, unlike the FRW case, the Schwarzschild geometry is not conformally flat, so we must account for the coupling of propagating photons to the background Weyl tensor in Eq. (3.3.24). As shown in Ref. [93], however, this

[3.7] Without this stipulation, Hawking evaporation causes the event horizon to move faster than the tiny corrections to the speed of light that we consider here.

contribution does not affect the speed of radially-propagating photons, so the Weyl component of the Schwarzschild metric can be ignored. Another subtlety is that very close to the horizon, the Hartle-Hawking vacuum actually implies deviations from thermality [123]. Because of these differences, Eq. (3.3.26) does not, strictly speaking, apply; that is, the numerical details of the superluminality bound (3.3.27) may be somewhat different. In any case, these detailed near-horizon corrections affect our results quantitatively but not qualitatively.

Consider the case in which the superluminality bound fails. In this case, photons will traverse slightly outside of the light-cone defined by the spacetime metric, due to the ambient Hawking radiation. Note that this setup differs crucially from that of Ref. [93]. In particular, the authors of Ref. [93] did not consider the effects of Hawking radiation, so modifications to the photon speed arose solely from the non-vanishing Weyl tensor in the vacuum Schwarzschild spacetime. As a result, Ref. [93] found that radially-propagating photons were luminal, so light cannot escape the event horizon. On the other hand, in our construction radial photons are superluminal if the bound fails, because the Hawking radiation modifies the photon speed in all directions. Consequently, a signal sent from inside the horizon can propagate radially to the outside in finite time as measured by an exterior observer. This phenomenon is in tension with black hole complementarity [124], in which the exterior and interior regions are treated as separate but equivalent Hilbert spaces. That is, if one were able to send signals from behind the horizon of a black hole, then the usual challenges to unitarity that come from black hole information theory [125] would no longer be so elegantly solved by complementarity.

Alternatively, one can interpret deviations from luminal photon propagation as a modification of the effective horizon of the black hole. For example, take the case where Eq. (3.3.27) (or its near-horizon analogue) is violated and the photon

is superluminal due to the ambient Hawking radiation. The effective horizon tilts in the space-like direction, shifting to a radius smaller than the usual Schwarzschild radius. Because the effective horizon shrinks, Hawking-radiated photons are emitted at a higher temperature. As the temperature increases, the velocity shift of the photon then increases, thus shrinking the effective horizon even more. In principle, this suggests an instability in the position of the effective black hole horizon. In contrast, if the bound is satisfied, then photon propagation is subluminal, the effective horizon grows, and Hawking-radiated photons exit at a lower temperature. In this case, the ambient photon gas is colder and the photon speed moves closer to unity. Hence, in this scenario the position of the effective horizon is stable.

Last of all, let us consider the maximally-extended Schwarzschild solution [126]. This background supports two asymptotically-flat spacetime regions, I and III, exterior to the two-sided black hole. One interpretation of this spacetime is that it describes a wormhole linking two black hole mouths [127]. When the superluminality bound fails, the concomitant faster-than-light propagation enables observers in regions I and III to communicate by sending signals through region II,[3.8] as shown in Fig. 3.4. Physically, this implies that the Einstein-Rosen bridge is traversable by photons and thus regions I and III are in causal contact. In contrast with usual constructions of traversable wormholes, this setup does not require the existence of exotic matter and associated violations of the averaged null energy condition [128]. As discussed in Ref. [129], if the wormhole mouths are in relative motion, it is possible to construct a CCC, in this case not traversable by matter following timelike or null trajectories, but rather by the superluminal photons that result from

[3.8] As for the one-sided black hole, we require that both wormhole mouths have static event horizons, which can be achieved by putting each in equilibrium with a thermal bath enclosing the mouth.

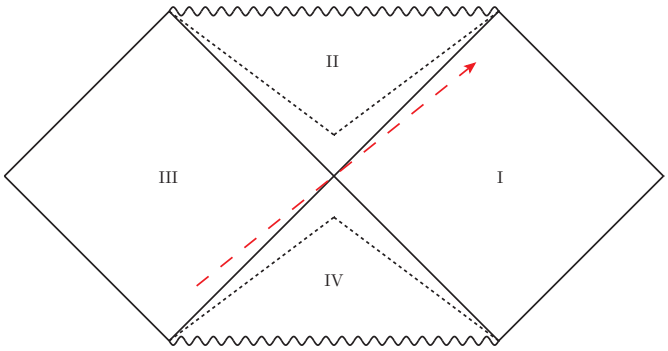

Figure 3.4. Conformal diagram for a maximally-extended Schwarzschild black hole. The effective horizon (dotted black) shrinks in a theory failing our superluminality bound. Superluminal photon propagation (red dashed arrow) allows observers in regions I and III to communicate.

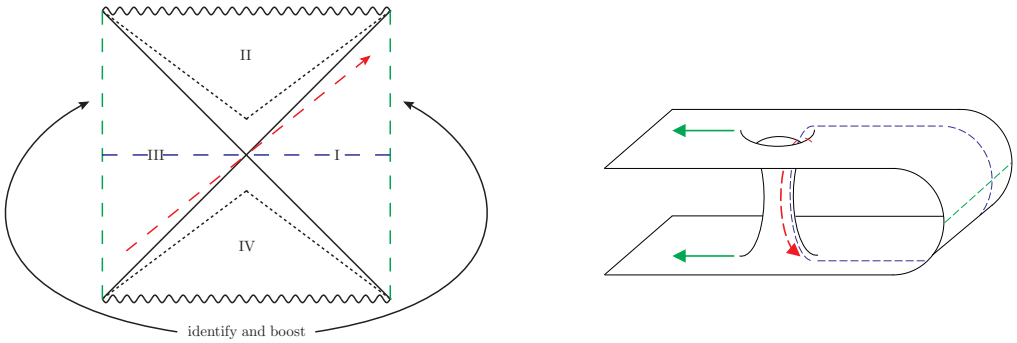

Figure 3.5. Conformal diagram (left) and embedding diagram of a spacelike slice (right) of the maximally-extended Schwarzschild black hole, describing wormhole mouths in relative motion. In a theory with superluminal propagation, the effective horizon (dotted black) shrinks and the wormhole becomes traversable by a signal sent from region III to I (red dashed arrow). The codimension-one surfaces (dashed green) at large spatial distance from the mouths are identified, albeit boosted relative to one another (green arrows). Also shown is a particular tangent codimension-three spacelike surface (dashed blue).

violation of the near-horizon version of Eq. (3.3.27), which is equally destructive to causality. See Fig. 3.5 for an illustration of this setup.

Note, however, that a wormhole can only support a true causal paradox if there is a boost between the wormhole mouths. In essence, the CCC construction is similar to that of Sec. 3.2.4, with the difference being that here we consider signals sent through the wormhole between two observers, one located just outside of each wormhole mouth. In particular, if the mouths are in relative

motion at velocity u, then Eq. (3.2.37) must be satisfied, where $v_{\rm avg}$ is the effective speed at which a light signal appears to propagate between the mouths as seen in the exterior spacetime, i.e., the speed of information propagation as measured by external observers located near each wormhole mouth. For $v_{\rm avg}$ only slightly superluminal, an enormous boost is required, inducing large back-reaction on the metric and invalidating our starting background. However, the wormhole mouths can be taken to be parametrically far apart; since the time the signal takes to go through the wormhole throat is independent of the distance between the mouths, $v_{\rm avg}$ can be made arbitrarily superluminal, overcoming any gravitational redshift effect in the exterior spacetime. With $v_{\rm avg}$ parametrically large, the required boost u can be very small, yielding negligible back-reaction and gravitational radiation while still allowing for the formation of a CCC.

From the perspective of AdS/CFT [52, 54], signal propagation through a traversable wormhole is puzzling and likely pathological [37]. As observed in Ref. [130], traversable wormholes correspond to non-local dynamics in the dual CFT. More concretely, our particular setup can be embedded in the construction of Ref. [131]: a maximally-extended Schwarzschild black hole geometry in asymptotically-AdS spacetime, dual to two entangled non-interacting CFTs on a sphere. In this geometry, the ability to send signals between regions I and III is dual to non-unitary evolution of the CFT, thus disrupting the canonical notions of entanglement entropy between the two CFTs [132, 133]. Moreover, in light of the ER=EPR conjecture [57], communication between mouths of an Einstein-Rosen bridge is dual to pathological information transfer via entanglement. While this scenario is deserving of a more thorough analysis, it lies beyond the scope of the present work.

We have outlined a variety of causal and quantum gravitational pathologies

that suggest that a superluminality bound like Eq. (3.3.27) is a requirement of any consistent low-energy effective theory. Assuming we are permitted to locate regions I and III of the extended Schwarzschild solution within the same asymptotic spacetime, then corrections to photon propagation that violate the superluminality bound transform the Einstein-Rosen bridge into a traversable wormhole and a CCC can be formed.

3.4 Summary and Future Directions

In this chapter, we have derived infrared consistency conditions on the photon-graviton effective action in Eq. (3.1.2) in 3D and 4D. These bounds are deduced from considerations of analyticity of light-by-light scattering, unitarity of the ultraviolet completion, and superluminality of photon fluctuations in nontrivial backgrounds. The 3D setup is a convenient starting point, where gravity is non-dynamical but still has a physical effect on photon-photon interactions. In 4D, many of the arguments are complicated (or, in the case of analyticity, even obstructed) by dynamical gravity. Our bounds on the photon-graviton effective action are summarized in Eqs. (3.1.4), (3.1.5), and (3.1.6) in Sec. 3.1. We then specialize to the case where electromagnetic corrections to the effective action come from a particle of charge-to-mass ratio $z = q/m$. Our infrared consistency conditions are then a constraint on a combination of z and coefficients parameterizing unspecified gravitational corrections, as shown in Eqs. (3.2.12), (3.3.9), and (3.3.10) and in Fig. 3.3.

The present work leaves a number of interesting avenues for future research. For example, as noted in Ref. [1], the WGC is not sharply defined in a theory with a Higgsed Abelian force carrier. In particular, in the Higgs phase, states of different charge can mix, so q and m are non-commuting operators, thus making the WGC ill defined. Furthermore, the original justification of the WGC—that

is, the pathology of exactly stable extremal black holes—is murky in the Higgs phase since a charged black hole can shed charge associated with a massive $U(1)$ and subsequently decay. On the other hand, the photon-graviton effective action is still well defined irrespective of whether the photon is massive or massless. As a result, it is especially interesting to consider infrared consistency conditions in the presence of a nonzero photon mass. For a Proca theory, we can simply add a physical mass. A more interesting case would be to introduce dynamical gauge symmetry breaking with a physical Higgs field.

Another direction for future work relates to the more complicated scenario of multiple Abelian forces. As shown in Ref. [1], it is straightforward to apply the logic of extremal black hole decay to theories with multiple forces and charged particles. The generalization of the WGC then becomes a simple geometric condition on the vectors describing the charge-to-mass ratios of particles in the theory. This generalization demands a more stringent constraint than Eq. (3.1.1) applied to each charge axis. Given this understanding, it would be interesting to see if similar geometric constraints arise from studying the low-energy effective action describing multiple photons interacting with the graviton. In principle, such an action will have many more free parameters than Eq. (3.1.2), but likewise many more constraints coming from analyticity, unitarity, and causality.

Last of all, we have not pursued possible constraints on the photon-graviton action from thermodynamic considerations. As discussed in Ref. [134], variations in the speed of light can allow for violation of the second law of thermodynamics when considering Hawking radiation in a black hole background. Since the speed of photon propagation is modified by higher-dimension operators, it may be possible to derive additional substantive constraints from thermodynamic reasoning.

The boundary between the landscape of healthy ultraviolet-completable theories and the swampland of pathological effective theories offers a promising arena for new physics insights. As we have shown, the particular criterion asserted by the WGC may be studied from purely low-energy reasoning given the nontrivial requirements of infrared consistency. In particular, we have determined regions in the effective theory that are forbidden by violations of analyticity, unitarity, and causality. Rescuing the forbidden regions of parameter space would require loopholes in all three arguments, or alternatively, reasons to countenance all of these pathologies.

3.A Propagator Numerator

In Eq. (3.3.17), we introduced a spectral representation for the field $\chi^{(4)}_{\mu\nu\rho\sigma}$. We now show that the tensor numerator of this spectral representation, $\Pi_{\mu\nu\rho\sigma\alpha\beta\gamma\delta}$, is highly constrained. To begin, note that $\chi^{(4)}_{\mu\nu\rho\sigma}$ does not correspond to a canonical spin-4 state, which is traditionally represented by a four-index, fully symmetric tensor [116, 117, 135, 136]. Like the Riemann tensor, $\chi^{(4)}_{\mu\nu\rho\sigma}$ is instead antisymmetric in its first and second pairs of indices separately and symmetric on the exchange of these pairs. The projection operator $\Pi_{\mu\nu\rho\sigma\alpha\beta\gamma\delta}$ inherits these index symmetry properties and tracelessness, and is furthermore symmetric on the interchange of the entire first and second sets of four indices. To determine $\Pi_{\mu\nu\rho\sigma\alpha\beta\gamma\delta}$, we start with an ansatz tensor that is an arbitrary function of $\eta_{\mu\nu}$ and k_μ. Imposing the transverse condition $k^\mu \Pi_{\mu\nu\rho\sigma\alpha\beta\gamma\delta} = 0$ on-shell, it is straightforward to show that $\Pi_{\mu\nu\rho\sigma\alpha\beta\gamma\delta}$ is necessarily a function of the projection operator $\Pi_{\mu\nu}$ in Eq. (3.2.23). Altogether, these restrictions only allow for two possible tensor structures:

$$
\begin{aligned}
&\Pi_{\mu\rho}\Pi_{\nu\beta}\Pi_{\sigma\delta}\Pi_{\alpha\gamma} + \Pi_{\nu\sigma}\Pi_{\mu\beta}\Pi_{\rho\delta}\Pi_{\alpha\gamma} + \Pi_{\mu\rho}\Pi_{\nu\alpha}\Pi_{\sigma\gamma}\Pi_{\beta\delta} + \Pi_{\nu\sigma}\Pi_{\mu\alpha}\Pi_{\rho\gamma}\Pi_{\beta\delta} \\
&+ \Pi_{\nu\rho}\Pi_{\mu\gamma}\Pi_{\sigma\beta}\Pi_{\alpha\delta} + \Pi_{\mu\sigma}\Pi_{\nu\gamma}\Pi_{\rho\beta}\Pi_{\alpha\delta} + \Pi_{\nu\rho}\Pi_{\mu\delta}\Pi_{\sigma\alpha}\Pi_{\beta\gamma} + \Pi_{\mu\sigma}\Pi_{\nu\delta}\Pi_{\rho\alpha}\Pi_{\beta\gamma} \\
&- \Pi_{\nu\rho}\Pi_{\mu\beta}\Pi_{\sigma\delta}\Pi_{\alpha\gamma} - \Pi_{\mu\sigma}\Pi_{\nu\beta}\Pi_{\rho\delta}\Pi_{\alpha\gamma} - \Pi_{\nu\rho}\Pi_{\mu\alpha}\Pi_{\sigma\gamma}\Pi_{\beta\delta} - \Pi_{\mu\sigma}\Pi_{\nu\alpha}\Pi_{\rho\gamma}\Pi_{\beta\delta} \\
&- \Pi_{\mu\rho}\Pi_{\nu\alpha}\Pi_{\sigma\delta}\Pi_{\beta\gamma} - \Pi_{\nu\sigma}\Pi_{\mu\alpha}\Pi_{\rho\delta}\Pi_{\beta\gamma} - \Pi_{\mu\rho}\Pi_{\nu\beta}\Pi_{\sigma\gamma}\Pi_{\alpha\delta} - \Pi_{\nu\sigma}\Pi_{\mu\beta}\Pi_{\rho\gamma}\Pi_{\alpha\delta} \\
&+ \Pi_{\nu\rho}\Pi_{\mu\alpha}\Pi_{\sigma\delta}\Pi_{\beta\gamma} + \Pi_{\mu\sigma}\Pi_{\nu\alpha}\Pi_{\rho\delta}\Pi_{\beta\gamma} + \Pi_{\nu\rho}\Pi_{\mu\beta}\Pi_{\sigma\gamma}\Pi_{\alpha\delta} + \Pi_{\mu\sigma}\Pi_{\nu\beta}\Pi_{\rho\gamma}\Pi_{\alpha\delta} \\
&+ \Pi_{\mu\rho}\Pi_{\nu\delta}\Pi_{\sigma\beta}\Pi_{\alpha\gamma} + \Pi_{\nu\sigma}\Pi_{\mu\delta}\Pi_{\rho\beta}\Pi_{\alpha\gamma} + \Pi_{\mu\rho}\Pi_{\nu\gamma}\Pi_{\sigma\alpha}\Pi_{\beta\delta} + \Pi_{\nu\sigma}\Pi_{\mu\gamma}\Pi_{\rho\alpha}\Pi_{\beta\delta} \\
&- \Pi_{\nu\rho}\Pi_{\mu\delta}\Pi_{\sigma\beta}\Pi_{\alpha\gamma} - \Pi_{\mu\sigma}\Pi_{\nu\delta}\Pi_{\rho\beta}\Pi_{\alpha\gamma} - \Pi_{\nu\rho}\Pi_{\mu\gamma}\Pi_{\sigma\alpha}\Pi_{\beta\delta} - \Pi_{\mu\sigma}\Pi_{\nu\gamma}\Pi_{\rho\alpha}\Pi_{\beta\delta} \\
&- \Pi_{\mu\rho}\Pi_{\nu\gamma}\Pi_{\sigma\beta}\Pi_{\alpha\delta} - \Pi_{\nu\sigma}\Pi_{\mu\gamma}\Pi_{\rho\beta}\Pi_{\alpha\delta} - \Pi_{\mu\rho}\Pi_{\nu\delta}\Pi_{\sigma\alpha}\Pi_{\beta\gamma} - \Pi_{\nu\sigma}\Pi_{\mu\delta}\Pi_{\rho\alpha}\Pi_{\beta\gamma} \\
&- \Pi_{\mu\alpha}\Pi_{\nu\beta}\Pi_{\rho\gamma}\Pi_{\sigma\delta} + \Pi_{\mu\beta}\Pi_{\nu\alpha}\Pi_{\rho\gamma}\Pi_{\sigma\delta} + \Pi_{\mu\alpha}\Pi_{\nu\beta}\Pi_{\rho\delta}\Pi_{\sigma\gamma} - \Pi_{\mu\beta}\Pi_{\nu\alpha}\Pi_{\rho\delta}\Pi_{\sigma\gamma} \\
&- \Pi_{\mu\gamma}\Pi_{\nu\delta}\Pi_{\rho\alpha}\Pi_{\sigma\beta} + \Pi_{\mu\delta}\Pi_{\nu\gamma}\Pi_{\rho\alpha}\Pi_{\sigma\beta} + \Pi_{\mu\gamma}\Pi_{\nu\delta}\Pi_{\rho\beta}\Pi_{\sigma\alpha} - \Pi_{\mu\delta}\Pi_{\nu\gamma}\Pi_{\rho\beta}\Pi_{\sigma\alpha} \\
&+ 2(\Pi_{\mu\rho}\Pi_{\nu\sigma}\Pi_{\alpha\delta}\Pi_{\beta\gamma} - \Pi_{\mu\rho}\Pi_{\nu\sigma}\Pi_{\alpha\gamma}\Pi_{\beta\delta} - \Pi_{\mu\sigma}\Pi_{\nu\rho}\Pi_{\alpha\delta}\Pi_{\beta\gamma} + \Pi_{\mu\sigma}\Pi_{\nu\rho}\Pi_{\alpha\gamma}\Pi_{\beta\delta})
\end{aligned}
\tag{3.A.1}
$$

and

$$\begin{aligned}
&\Pi_{\mu\alpha}\Pi_{\nu\beta}\Pi_{\rho\gamma}\Pi_{\sigma\delta} - \Pi_{\mu\beta}\Pi_{\nu\alpha}\Pi_{\rho\gamma}\Pi_{\sigma\delta} - \Pi_{\mu\alpha}\Pi_{\nu\beta}\Pi_{\rho\delta}\Pi_{\sigma\gamma} + \Pi_{\mu\beta}\Pi_{\nu\alpha}\Pi_{\rho\delta}\Pi_{\sigma\gamma} \\
&+ \Pi_{\mu\gamma}\Pi_{\nu\delta}\Pi_{\rho\alpha}\Pi_{\sigma\beta} - \Pi_{\mu\delta}\Pi_{\nu\gamma}\Pi_{\rho\alpha}\Pi_{\sigma\beta} - \Pi_{\mu\gamma}\Pi_{\nu\delta}\Pi_{\rho\beta}\Pi_{\sigma\alpha} + \Pi_{\mu\delta}\Pi_{\nu\gamma}\Pi_{\rho\beta}\Pi_{\sigma\alpha} \\
&+ \Pi_{\mu\alpha}\Pi_{\nu\gamma}\Pi_{\rho\delta}\Pi_{\sigma\beta} - \Pi_{\mu\gamma}\Pi_{\nu\alpha}\Pi_{\rho\delta}\Pi_{\sigma\beta} - \Pi_{\mu\alpha}\Pi_{\nu\gamma}\Pi_{\rho\beta}\Pi_{\sigma\delta} + \Pi_{\mu\gamma}\Pi_{\nu\alpha}\Pi_{\rho\beta}\Pi_{\sigma\delta} \\
&+ \Pi_{\mu\alpha}\Pi_{\nu\delta}\Pi_{\rho\beta}\Pi_{\sigma\gamma} - \Pi_{\mu\delta}\Pi_{\nu\alpha}\Pi_{\rho\beta}\Pi_{\sigma\gamma} - \Pi_{\mu\alpha}\Pi_{\nu\delta}\Pi_{\rho\gamma}\Pi_{\sigma\beta} + \Pi_{\mu\delta}\Pi_{\nu\alpha}\Pi_{\rho\gamma}\Pi_{\sigma\beta} \\
&+ \Pi_{\mu\gamma}\Pi_{\nu\beta}\Pi_{\rho\delta}\Pi_{\sigma\alpha} - \Pi_{\mu\beta}\Pi_{\nu\gamma}\Pi_{\rho\delta}\Pi_{\sigma\alpha} - \Pi_{\mu\gamma}\Pi_{\nu\beta}\Pi_{\rho\alpha}\Pi_{\sigma\delta} + \Pi_{\mu\beta}\Pi_{\nu\gamma}\Pi_{\rho\alpha}\Pi_{\sigma\delta} \\
&+ \Pi_{\mu\delta}\Pi_{\nu\beta}\Pi_{\rho\alpha}\Pi_{\sigma\gamma} - \Pi_{\mu\beta}\Pi_{\nu\delta}\Pi_{\rho\alpha}\Pi_{\sigma\gamma} - \Pi_{\mu\delta}\Pi_{\nu\beta}\Pi_{\rho\gamma}\Pi_{\sigma\alpha} + \Pi_{\mu\beta}\Pi_{\nu\delta}\Pi_{\rho\gamma}\Pi_{\sigma\alpha}.
\end{aligned} \tag{3.A.2}$$

Consequently, $\Pi_{\mu\nu\rho\sigma\alpha\beta\gamma\delta}$ must be an arbitrary linear combination of these two tensors. As noted in the body of the text, however, the forms of these tensors imply that $\Pi_{\mu\nu\rho\sigma}{}^{\mu\nu\rho\sigma} = 0$, which cannot be equal to a sum over polarization tensors and is thus in violation of unitarity.

Chapter 4
Quantum Gravity Constraints from Unitarity and Analyticity

We derive rigorous bounds on corrections to Einstein gravity using unitarity and analyticity of graviton scattering amplitudes. In $D \geq 4$ spacetime dimensions, these consistency conditions mandate positive coefficients for certain quartic curvature operators. We systematically enumerate all such positivity bounds in $D = 4$ and $D = 5$ before extending to $D \geq 6$. Afterwards, we derive positivity bounds for supersymmetric operators and verify that all of our constraints are satisfied by weakly-coupled string theories. Among quadratic curvature operators, we find that the Gauss-Bonnet term in $D \geq 5$ is inconsistent unless new degrees of freedom enter at the natural cutoff scale defined by the effective theory. Our bounds apply to perturbative ultraviolet completions of gravity.

This chapter is from Ref. [3], B. Bellazzini, C. Cheung, and G. N. Remmen, "Quantum Gravity Constraints from Unitarity and Analyticity," Phys. Rev. ***D93*** *(2016) 064076,* arXiv:1509.00851 [hep-th].

4.1 Introduction

Low-energy effective theory describes quanta interacting indirectly through kinematically inaccessible states. The dynamics are characterized by an effective action that typically includes all interactions permitted by symmetry with couplings of order unity. However, in certain cases, self-consistency at long distances imposes nontrivial constraints on the coefficients of effective operators. This is famously true in the theory of pions, where the operator coefficients of certain higher-derivative operators are required to be strictly positive to

ensure causal particle propagation together with unitarity and analyticity of scattering amplitudes at complex momenta [31, 111, 113, 137, 138].

As low-energy criteria, causality, unitarity, and analyticity impose constraints that are independent of the detailed ultraviolet dynamics. Consequently, these consistency conditions offer special utility in the context of quantum gravity, where the ultraviolet completion is not known with certainty. For instance, such bounds have been derived for the effective theory of gravitons and photons [2], where consistency necessitates large charge-to-mass ratios precisely of the form of the weak gravity conjecture [30].

Notably, a proper diagnosis of causality violation in curved spacetime is subtle since particle propagation can be locally superluminal even in healthy theories. For example, it has long been known that photons with certain polarizations travel superluminally in the vicinity of a black hole in the effective theory of photons and gravitons describing our actual universe [93]. Instead, a more global measure of causality, e.g., the existence of closed timelike curves, is necessary to establish a true pathology. On the other hand, unitarity and analyticity offer alternative criteria that are mathematically rigorous and applicable in the flat-space limit.

In this chapter, we systematically derive new constraints on curvature corrections in gravity from unitarity and analyticity. The graviton effective theory is described by the action[4.1]

$$S = \int \mathrm{d}^D x \, \sqrt{-g} \, \sum_{n=1}^{\infty} \mathcal{L}_n, \tag{4.1.1}$$

where $\mathcal{L}_n$ are contributions to the action entering at order $2n$ in the derivative expansion and

[4.1] In this chapter, we work in mostly-plus signature and write $\kappa = \sqrt{8\pi G}$, $R_{\mu\nu} = R^{\rho}{}_{\mu\rho\nu}$, and $R^{\mu}{}_{\nu\rho\sigma} = \partial_\rho \Gamma^{\mu}{}_{\nu\sigma} + \cdots$.

$$\mathcal{L}_1 = \frac{R}{2\kappa^2} \qquad \text{and} \qquad \mathcal{L}_2 = \lambda(R_{\mu\nu\rho\sigma}R^{\mu\nu\rho\sigma} - 4R_{\mu\nu}R^{\mu\nu} + R^2) \tag{4.1.2}$$

are the Einstein-Hilbert and Gauss-Bonnet terms. We assume the Gauss-Bonnet form for $\mathcal{L}_2$ throughout this chapter since this is the unique ghost-free quadratic curvature invariant [139, 140] in D dimensions. Moreover, we restrict our analysis to effective theories in which $\mathcal{L}_4$ takes the form

$$\mathcal{L}_4 = \sum_{i=1}^{7} c_i \mathcal{O}_i, \tag{4.1.3}$$

expressed in terms of the minimal basis of quartic Riemann operators in Ref. [141],

$$\begin{aligned}
\mathcal{O}_1 &= R^{\mu\nu\rho\sigma}R_{\mu\nu\rho\sigma}R^{\alpha\beta\gamma\delta}R_{\alpha\beta\gamma\delta} & \mathcal{O}_2 &= R^{\mu\nu\rho\sigma}R_{\mu\nu\rho}{}^{\delta}R^{\alpha\beta\gamma}{}_{\sigma}R_{\alpha\beta\gamma\delta} \\
\mathcal{O}_3 &= R^{\mu\nu\rho\sigma}R_{\mu\nu}{}^{\alpha\beta}R_{\alpha\beta}{}^{\gamma\delta}R_{\rho\sigma\gamma\delta} & \mathcal{O}_4 &= R^{\mu\nu\rho\sigma}R_{\mu\nu}{}^{\alpha\beta}R_{\rho\alpha}{}^{\gamma\delta}R_{\sigma\beta\gamma\delta} \\
\mathcal{O}_5 &= R^{\mu\nu\rho\sigma}R_{\mu\nu}{}^{\alpha\beta}R_{\rho}{}^{\gamma}{}_{\alpha}{}^{\delta}R_{\sigma\gamma\beta\delta} & \mathcal{O}_6 &= R^{\mu\nu\rho\sigma}R_{\mu}{}^{\alpha}{}_{\rho}{}^{\beta}R_{\alpha}{}^{\gamma}{}_{\beta}{}^{\delta}R_{\nu\gamma\sigma\delta} \\
\mathcal{O}_7 &= R^{\mu\nu\rho\sigma}R_{\mu}{}^{\alpha}{}_{\rho}{}^{\beta}R_{\alpha}{}^{\gamma}{}_{\nu}{}^{\delta}R_{\beta\gamma\sigma\delta}.
\end{aligned} \tag{4.1.4}$$

Note that linear dependences arise among operators as the dimension D of spacetime decreases. At quadratic order, $\mathcal{L}_2$ is unphysical in $D \leq 3$, a total derivative in $D = 4$, and dynamical in $D \geq 5$. Meanwhile, at quartic order, the number of algebraically independent[4.2] operators $\mathcal{O}_i$ in $D = 4, 5, 6, 7, 8$ is $2, 4, 6, 6, 7$, respectively [141], with one linear combination—the eight-dimensional Euler density—a total derivative in $D = 8$ and hence dynamical only in $D \geq 9$ [143].

Our analysis hinges on the on-shell four-point graviton scattering amplitude, M, whose forward limit is intimately linked to the total cross-section by well-known analyticity arguments [31, 138]. By marginalizing over the external graviton polarizations, we can then systematically derive a rigorous and inclusive set positivity bounds on the coefficients of operators in the graviton

[4.2] Applying leading-order equations of motion to $\mathcal{L}_n$ is equivalent to a field definition modulo new terms generated in $\mathcal{L}_{n+1}$. Repeating this procedure at each order, we can freely impose $R = R_{\mu\nu} = 0$ in a pure gravity theory [142].

effective action. Throughout this chapter, we assume a perturbative ultraviolet completion of gravity, so there exists a well-defined expansion in $\hbar$.

We begin with an analysis of quartic curvature corrections, proving that in $D = 4$ the coefficients of the $(R^2)^2$ and $(R\tilde{R})^2$ operators are positive. Our results precisely match those of Ref. [144], which derived bounds from the condition of locally subluminal graviton propagation. We then generalize our arguments to $D = 5$ and $D \geq 6$. Subsequently, we obtain positivity constraints on supersymmetric operators in general D, which in the literature are sometimes denoted $t_8 t_8 R^4$ and $t_8 (R^2)^2$. As a consistency check, we verify that all our constraints are satisfied in the bosonic, type II, and heterotic string theories.

Moving on to quadratic curvature corrections, we argue that unitarity and analyticity exclude theories in which $\lambda \gg 1$ with no new degrees of freedom present at or below the mass scale $\Lambda \sim |\lambda\kappa^2|^{-1/2}$, the natural cutoff associated with the Gauss-Bonnet term and the derivative expansion. Our results precisely accord with those of Maldacena et al. [37], who demonstrated that this class of theories is inconsistent with global causality.

The remainder of this chapter is organized as follows. In Sec. 4.2, we review the arguments of Ref. [31] whereby unitarity and analyticity impose rigorous positivity bounds on operator coefficients in effective theories. Afterwards, in Sec. 4.3 we apply these methods to establish the positivity of certain coefficients of quartic curvature operators, starting in $D = 4$ and $D = 5$ and then generalizing to $D \geq 6$. We then apply our bounds to supersymmetric theories and string theories. Finally, we study quadratic curvature operators in Sec. 4.4 and conclude in Sec. 4.5.

4.2 Analyticity Argument

In this section we review how operator coefficients in effective field theories are constrained by the analyticity of scattering amplitudes at complex momenta. Our analysis follows that of Ref. [31], which derived positivity bounds on operator coefficients by relating the low-energy limit of forward amplitudes to strictly positive integrals of cross-sections.

Our object of interest is the on-shell amplitude M describing four-point graviton scattering in D dimensions. Here, the choice of the external polarizations is built into the functional form of M, as is the case for helicity amplitudes in $D = 4$. From this viewpoint, helicity is just a quantum number labeling the external states, no different from baryon or lepton number. Sometimes it will be useful to view M as a function of the external particle labels, $M = M(1, 2, 3, 4)$, and other times as a function of the kinematic invariants, $M = M(s, t, u)$, where

$$s = -(k_1 + k_2)^2, \quad t = -(k_1 + k_3)^2, \quad u = -(k_1 + k_4)^2, \tag{4.2.1}$$

working in the convention where all momenta are incoming, so $k_1+k_2+k_3+k_4 = 0$.

To derive constraints from analyticity, we will be interested in scattering amplitudes that are simultaneously forward and invariant under crossing in the t-channel. Formally, t-channel crossing symmetry implies invariance under swapping particle labels $1 \leftrightarrow 3$ or $2 \leftrightarrow 4$ while leaving the functional form of M—which encodes the polarization choices—fixed, so

$$M(1, 2, 3, 4) = M(3, 2, 1, 4) = M(1, 4, 3, 2) = M(3, 4, 1, 2). \tag{4.2.2}$$

For external gravitons, crossing symmetry is ensured if the exchanged states are indistinguishable. This happens in $D = 4$ if the states have identical helicity

and more generally in D dimensions if the states have the same polarization relative to their momenta. Mathematically, M is crossing symmetric provided the momentum and polarization of particle 1 are related by an improper Lorentz transformation to those of particle 3, and likewise for particles 2 and 4. In terms of kinematic invariants, a crossing symmetric amplitude then satisfies

$$M(s,t) = M(-s-t,t), \tag{4.2.3}$$

where the momenta are swapped but the polarizations relative to momenta—which in $D = 4$ are the helicities—are untouched.

Meanwhile, the forward limit of the amplitude, $M(s, t \to 0)$, corresponds to an identification of particles $1 \leftrightarrow 3$ and $2 \leftrightarrow 4$. This is achieved simultaneously with crossing symmetry if we restrict to the following kinematic regime:

$$\begin{array}{l} \text{forward and} \\ \text{crossing symmetric} \end{array} \quad \Longrightarrow \quad \begin{array}{l} (k_3, \epsilon_3) \leftrightarrow (-k_1, \epsilon_1) \ \text{ and} \\ (k_4, \epsilon_4) \leftrightarrow (-k_2, \epsilon_2), \end{array} \tag{4.2.4}$$

where ϵ_1 and ϵ_2 are real linear polarizations. We choose a basis of linear polarizations because an amplitude with fixed external circular polarizations cannot be simultaneously crossing symmetric and forward.

The forward and crossing symmetric amplitude, $M(s, t \to 0)$, can then be expanded in a power series in s and t. While analytic singularities in s or t arise, their form is severely restricted by the locality of the underlying theory. As noted earlier, we assume throughout this chapter a perturbative ultraviolet completion of gravity, so we are justified in considering only the leading contribution in the $\hbar$ expansion, i.e., tree-level exchange.

At tree level, analytic singularities in kinematic invariants enter at worst as simple poles. Moreover, a t-channel singularity in the forward limit can only arise from non-local terms corresponding to graviton exchange induced by the leading Einstein-Hilbert interactions, so the general form for the forward

amplitude is

$$M(s, t \to 0) = \sum_{n=0}^{\infty} f_n s^n + \mathcal{O}(s^2/t). \tag{4.2.5}$$

The first term is regular, as it is generated by heavy particle exchange, while the second term is singular because it comes from t-channel graviton exchange scaling as $\sim s^2/t$. The form of Eq. (4.2.5) together with the crossing symmetry relation of Eq. (4.2.3) implies that

$$M(s, t \to 0) = M(-s, t \to 0) + \mathcal{O}(s), \tag{4.2.6}$$

where the first term arises because the limit of a regular function is the function evaluated at the limit of its arguments, while the second term is a residual contribution from applying crossing symmetry to the singular $\mathcal{O}(s^2/t)$ contribution.

The parameters f_n depend on the coefficients of operators in the effective action of the low-energy theory. To determine analyticity constraints, we consider the order n residue of $M(s, t \to 0)$ in the complex s plane, yielding

$$f_n = \frac{1}{2\pi i} \oint_{\mathcal{C}} \frac{\mathrm{d}s}{s^{n+1}} \left[M(s, t \to 0) + \mathcal{O}(s^2/t) \right], \tag{4.2.7}$$

where $\mathcal{C}$ denotes a small contour encircling the origin.

As previously noted [2, 31], the $\mathcal{O}(s^2/t)$ singular contribution is formally infinite in the strictly forward limit and therefore a major obstacle to deriving bounds from analyticity. Nevertheless, for $n \neq 2$ this term is eliminated by the residue theorem. While forward singularities of order s^n/t can arise from loop-level graviton exchange diagrams, we are working at leading order in the $\hbar$ expansion so these contributions are formally subdominant. On the other hand, $n = 2$ is more subtle, but we will show that in certain parameter regimes the $\mathcal{O}(s^2/t)$ term can be subdominant to the rest of the amplitude, allowing for a bound to be placed. In any case, we leave a detailed discussion of these issues for later sections and for now simply drop the $\mathcal{O}(s^2/t)$ contribution. Terms

subleading in the forward limit of the Einstein-Hilbert amplitude must by power counting go as $\mathcal{O}(s)$ and will thus be eliminated in the contour integral for all $n \geq 2$.

By Cauchy's theorem, we can blow up $\mathcal{C}$ into a new contour $\mathcal{C}'$ that runs just above and below the real s axis, plus a circular boundary contour at infinity, yielding

$$f_n = \frac{1}{2\pi i}\left(\int_{-\infty}^{-s_0} + \int_{s_0}^{\infty}\right)\frac{\mathrm{d}s}{s^{n+1}}\mathrm{Disc}[M(s, t \to 0)] + \text{boundary integral}, \quad (4.2.8)$$

where s_0 is any scale above zero and below the first massive threshold in the ultraviolet completion. We note that $\mathrm{Disc}[M(s, t \to 0)] = M(s + i\epsilon, t \to 0) - M(s - i\epsilon, t \to 0)$. By the Schwarz reflection principle, $M(s^*, t \to 0) = M(s, t \to 0)^*$, so

$$\mathrm{Disc}[M(s, t \to 0)] = 2i\mathrm{Im}[M(s, t \to 0)]. \quad (4.2.9)$$

The crossing symmetry relation in Eq. (4.2.6) then implies that

$$\mathrm{Disc}[M(-s, t \to 0)] = -\mathrm{Disc}[M(s, t \to 0)], \quad (4.2.10)$$

dropping the $\mathcal{O}(s)$ term that arose from the $\mathcal{O}(s^2/t)$ singularity.

Throughout this chapter we assume that $|M(s)|$ is less divergent than $|s|^n$ at large s so that the boundary term in Eq. (4.2.8) can be dropped.[4.3] This is a physically reasonable assumption applicable to any ultraviolet completion in which the large s behavior of the amplitude at fixed finite physical $t \ll s$ grows more slowly in s than the Einstein-Hilbert contribution, which scales as s^2/t. A theory that fails this criterion would actually have worse ultraviolet behavior than Einstein gravity. Operationally, this translates into the assumption that $|M(s)|$ grows more slowly than $|s|^2$ at large s. For example, this can be verified

[4.3] Strictly speaking, positivity bounds only require that the boundary term be non-negative, which is sometimes true given specific assumptions about the ultraviolet [145]. We do not consider this possibility here.

explicitly in the Regge behavior of string theory amplitudes, which scale as $s^{\alpha(t)}/t$ where $\alpha(t) < 2$ for $t < 0$ [31].

Combining Eq. (4.2.8) with Eqs. (4.2.9) and (4.2.10) yields

$$f_n = \frac{(-1)^n + 1}{\pi} \int_{s_0}^{\infty} \frac{\mathrm{d}s}{s^{n+1}} \mathrm{Im}[M(s, t \to 0)]. \tag{4.2.11}$$

For n odd, this result trivially implies $f_n = 0$ as required by crossing symmetry of M, while for n even, it imposes a positivity condition. In particular, we use the optical theorem to write $\mathrm{Im}[M(s, t \to 0)] = s\sigma(s)$, where $\sigma(s)$ is the total cross-section.[4.4] Crucially, in an interacting theory with new heavy states, $\sigma(s)$ is strictly positive, so

$$f_n = \frac{2}{\pi} \int_{s_0}^{\infty} \mathrm{d}s \frac{\sigma(s)}{s^n} > 0, \tag{4.2.12}$$

thus establishing positivity of f_n for even n.

Here f_n corresponds to the s^n contribution to the low-energy amplitude, which is proportional to the operator coefficients of $\mathcal{L}_n$. By power counting, we know that the low-energy amplitude can be expanded in powers of Mandelstam variables, so

$$M = \sum_{n=1}^{\infty} M_n, \tag{4.2.13}$$

where the leading contribution arises from the Einstein-Hilbert action, which in the forward limit gives an amplitude

$$M_1(s, t \to 0) = -\epsilon_{1\mu\nu}\epsilon_1^{\mu\nu}\epsilon_{2\rho\sigma}\epsilon_2^{\rho\sigma} \frac{\kappa^2 s^2}{t} + \mathcal{O}(s), \tag{4.2.14}$$

where the $\mathcal{O}(s)$ terms are regular in the forward limit. The remaining contributions M_n are generated by $\mathcal{L}_n$. In the subsequent sections, we derive precise analyticity bounds for the quartic and quadratic curvature corrections, $\mathcal{L}_4$ and

[4.4] While the total cross-section diverges in the presence of a t-channel singularity, $\mathrm{Im}\, M(s, t \to 0)$ and by extension $\sigma(s) = \mathrm{Im}\, M(s, t \to 0)/s$ are really proxies for the finite sum over all residues of heavy states in the complex s plane. By factorization, each contribution is positive since $M(hh \to hh) \sim -M(hh \to X)M(X \to hh)/(s - m^2 + i\epsilon)$ on the s-channel resonance of a massive state X of mass m.

$\mathcal{L}_2$.

4.3 Bounds on Quartic Curvature Corrections

In this section we derive bounds on $\mathcal{L}_4$, which encodes quartic curvature corrections to Einstein gravity. The leading contributions from $\mathcal{L}_4$ are quartic graviton vertices, which contribute to graviton scattering amplitudes via contact interactions. Since these corrections are free from kinematic singularities, their forward limit is regular. Thus, to obtain a forward, crossing symmetric amplitude, we simply set $t = 0$, $\epsilon_3 = \epsilon_1$, and $\epsilon_4 = \epsilon_2$, as derived in Eq. (4.2.4).

After a lengthy calculation, we compute the quartic corrections to the graviton scattering amplitude in the forward limit to be

$$\begin{aligned} M_4(s, t \to 0) = \frac{\kappa^4 s^4}{2}\Big[&(2c_6 + c_7)(\epsilon_{1\mu\nu}\epsilon_1{}^{\mu\nu}\epsilon_{2\rho\sigma}\epsilon_2{}^{\rho\sigma}) \\ &+ (32c_1 + 4c_2 + 2c_6)(\epsilon_{1\mu\nu}\epsilon_2{}^{\mu\nu})^2 \\ &+ (4c_2 + 16c_3 + 2c_6)(\epsilon_1{}^{\mu\nu}\epsilon_{2\nu\rho}\epsilon_1{}^{\rho\sigma}\epsilon_{2\sigma\mu}) \\ &+ (4c_2 + 8c_4 + 2c_7)(\epsilon_1{}^{\mu\nu}\epsilon_{1\nu\rho}\epsilon_2{}^{\rho\sigma}\epsilon_{2\sigma\mu})\Big]. \end{aligned} \tag{4.3.1}$$

Eq. (4.2.12) bounds f_4, corresponding to the coefficient of the s^4 contribution to the amplitude, to be positive. To determine the constraint on the coefficients of $\mathcal{L}_4$, we should marginalize over all possible values of the independent polarizations, ϵ_1 and ϵ_2.

To determine the full set of bounds, it will be convenient to map the question of positivity to a linear algebra problem. To do so, we work in the center-of-mass frame, where the polarization tensors, $\epsilon_{1\mu}{}^{\nu}$ and $\epsilon_{2\mu}{}^{\nu}$ are real, symmetric $(D-2)$-by-$(D-2)$ matrices satisfying the usual tracelessness and normalization conditions,

$$\mathrm{Tr}\,(\epsilon_1) = \mathrm{Tr}\,(\epsilon_2) = 0 \quad \text{and} \quad \mathrm{Tr}\,(\epsilon_1 \cdot \epsilon_1) = \mathrm{Tr}\,(\epsilon_2 \cdot \epsilon_2) = 1. \tag{4.3.2}$$

Furthermore, we can define Hermitian matrices $H_+ = \{\epsilon_1, \epsilon_2\}/2$ and $H_- = i[\epsilon_1, \epsilon_2]/2$ encoding the polarization information, which enter the amplitude in terms of the invariants

$$x = \mathrm{Tr}\,(H_+)\mathrm{Tr}\,(H_+), \qquad y = \mathrm{Tr}\,(H_+ \cdot H_+), \qquad z = \mathrm{Tr}\,(H_- \cdot H_-). \quad (4.3.3)$$

We can then express the analyticity bound as

$$\begin{aligned} 0 <& (2c_6 + c_7) \\ &+ (32c_1 + 4c_2 + 2c_6)x \\ &+ (8c_2 + 16c_3 + 8c_4 + 2c_6 + 2c_7)y \\ &+ (-16c_3 + 8c_4 - 2c_6 + 2c_7)z, \end{aligned} \quad (4.3.4)$$

for all (x, y, z) spanned by the graviton polarizations ϵ_1 and ϵ_2. What is the allowed space of (x, y, z)? An obvious set of necessary conditions are

$$0 \leq x, y, z \leq 1 \quad \text{and} \quad y + z \leq 1, \quad (4.3.5)$$

from familiar linear algebra inequalities. In general D, finding the space spanned by the allowed (x, y, z) is a highly nontrivial problem in matrix inequalities.

In the next subsections, we will study various physically well-motivated scenarios, including general theories in $D = 4$ and supersymmetric theories in arbitrary D.

4.3.1 Theories in $D = 4$

The number of linearly independent curvature invariants monotonically increases with the dimension of spacetime. In $D = 4$, there are only two independent quartic curvature invariants. Hence, $\mathcal{L}_4$ in Eq. (4.1.4) collapses to

$$\mathcal{L}_4 = c_1\mathcal{O}_1 + \tilde{c}_1\tilde{\mathcal{O}}_1, \quad (4.3.6)$$

where $\mathcal{O}_1$ is defined as in Eq. (4.1.4) but $\tilde{\mathcal{O}}_1$ is unique to $D = 4$,

$$\mathcal{O}_1 = R^{\mu\nu\rho\sigma}R_{\mu\nu\rho\sigma}R^{\alpha\beta\gamma\delta}R_{\alpha\beta\gamma\delta} \text{ and } \tilde{\mathcal{O}}_1 = R^{\mu\nu\rho\sigma}\tilde{R}_{\mu\nu\rho\sigma}R^{\alpha\beta\gamma\delta}\tilde{R}_{\alpha\beta\gamma\delta}, \quad (4.3.7)$$

where $\tilde{R}_{\mu\nu\rho\sigma} = R_{\mu\nu}{}^{\alpha\beta}\epsilon_{\alpha\beta\rho\sigma}/2$ is the dual Riemann tensor. The operator $\tilde{\mathcal{O}}_1$ can be written as a linear combination of any two the operators in Eq. (4.1.4) modulo contributions proportional to R and $R_{\mu\nu}$, which can be eliminated by the equations of motion. For example,

$$\tilde{\mathcal{O}}_1 = 4\mathcal{O}_2 - 4\mathcal{O}_3 = -4\mathcal{O}_2 + 8\mathcal{O}_4 = \cdots , \tag{4.3.8}$$

corresponding to a choice of operator coefficients, $(c_1, 4\tilde{c}_1, -4\tilde{c}_1, 0, 0, 0, 0)$, $(c_1, -4\tilde{c}_1, 0, 8\tilde{c}_1, 0, 0, 0)$, etc. The ellipses in Eq. (4.3.8) denote equivalent representations in terms of other operators, which are not unique due to the linear dependence in $D = 4$ of all but two of the operators in Eq. (4.1.4).

In $D = 4$, the invariants (x, y, z) are constructed from real, symmetric, traceless 2-by-2 matrices, which we can parameterize by

$$\begin{aligned} \epsilon_1 &= \vec{\epsilon}_1 \cdot \vec{\sigma}/\sqrt{2} \\ \epsilon_2 &= \vec{\epsilon}_2 \cdot \vec{\sigma}/\sqrt{2}, \end{aligned} \tag{4.3.9}$$

where $\vec{\epsilon}_1$ and $\vec{\epsilon}_2$ are real unit polarization vectors and $\vec{\sigma}$ are the Pauli matrices. Since ϵ_1 and ϵ_2 are real and symmetric, they only have components in σ_1 and σ_3, since σ_2 is imaginary and anti-symmetric. From standard matrix identities, we see that $\{\epsilon_1, \epsilon_2\} = \vec{\epsilon}_1 \cdot \vec{\epsilon}_2$ and $[\epsilon_1, \epsilon_2] = i(\vec{\epsilon}_1 \times \vec{\epsilon}_2) \cdot \vec{\sigma}$. Defining θ to be the angle between $\vec{\epsilon}_1$ and $\vec{\epsilon}_2$, we obtain

$$(x, y, z) = \cos^2\theta\ (1, \tfrac{1}{2}, 0) + \sin^2\theta\ (0, 0, \tfrac{1}{2}), \tag{4.3.10}$$

which defines an interval whose endpoints are $(1, \frac{1}{2}, 0)$ and $(0, 0, \frac{1}{2})$. Inserting these (x, y, z) values, along with the coefficient choice given by Eqs. (4.3.6) and (4.3.8), the bound (4.3.4) takes the suggestive form

$$c_1 \cos^2\theta + \tilde{c}_1 \sin^2\theta > 0, \tag{4.3.11}$$

which obviously implies positivity of both coefficients separately,

$$c_1 > 0 \quad \text{and} \quad \tilde{c}_1 > 0, \tag{4.3.12}$$

which correspond to parallel or perpendicular polarization vectors, respectively.

Our results exactly coincide with those derived from requiring subluminal graviton propagation [144].

4.3.2 Theories in $D = 5$

In $D = 5$, there are four linearly independent quartic curvature invariants. For the sake of generality we use the basis of Eq. (4.1.4) with the linear dependences among operators assumed. For this analysis, we ascertain the physically allowed region for the invariants (x, y, z), which in $D = 5$ are constructed from real, symmetric, traceless 3-by-3 matrices. This requirement constrains (x, y, z) to lie in the plane $1 + 2x - 6y - 2z = 0$. Specifically, (x, y, z) are restricted to a planar triangular region,

$$(x, y, z) = \sum_{i=1}^{3} \tau_i v_i, \tag{4.3.13}$$

defined by three vectors

$$v_1 = (0, 0, \tfrac{1}{2}), \qquad v_2 = (1, \tfrac{1}{2}, 0), \qquad \text{and} \qquad v_3 = (0, \tfrac{1}{6}, 0) \tag{4.3.14}$$

for the real parameters $\tau_1, \tau_2, \tau_3 \geq 0$ such that $\tau_1 + \tau_2 + \tau_3 = 1$. The vertices (4.3.14) of this triangle can be reached by choices of physical polarizations. In particular,

$$\begin{aligned}
v_1: \quad & \epsilon_1 = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 0 \end{pmatrix}, && \epsilon_2 = \frac{1}{\sqrt{2}} \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \\
v_2: \quad & \epsilon_1 = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 0 \end{pmatrix}, && \epsilon_2 = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \\
v_3: \quad & \epsilon_1 = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 0 \end{pmatrix}, && \epsilon_2 = \frac{1}{\sqrt{6}} \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & -2 \end{pmatrix}.
\end{aligned} \tag{4.3.15}$$

Plugging in Eqs. (4.3.13) and (4.3.14) back into Eq. (4.3.4), we obtain

$$\begin{bmatrix} \tau_1(-8c_3+4c_4+c_6+2c_7) \\ +\,\tau_2(32c_1+8c_2+8c_3+4c_4+5c_6+2c_7) \\ +\,\frac{1}{3}\tau_3(4c_2+8c_3+4c_4+7c_6+4c_7) \end{bmatrix} > 0, \tag{4.3.16}$$

where we have repackaged the terms independent of (x, y, z) in Eq. (4.3.4) into the coefficients τ_1, τ_2, τ_3 by re-expressing 1 as $\tau_1 + \tau_2 + \tau_3$. Thus, the necessary and sufficient set of bounds on quartic curvature corrections in $D = 5$ are

$$\begin{aligned} -8c_3+4c_4+c_6+2c_7 &> 0 \\ 32c_1+8c_2+8c_3+4c_4+5c_6+2c_7 &> 0 \\ 4c_2+8c_3+4c_4+7c_6+4c_7 &> 0, \end{aligned} \tag{4.3.17}$$

coming from analyticity of the four-point graviton scattering amplitude.

4.3.3 Theories in $D \geq 6$

Consider finally the general case of $D \geq 6$. It is a nontrivial linear algebra problem to determine the parameter space of (x, y, z) corresponding to physical polarization configurations. Each physically allowed point (x, y, z) yields a positivity bound via Eq. (4.3.4). The set of all positive linear combinations of such bounds is given by plugging into Eq. (4.3.4) the set of all points in the convex hull S spanning physically allowed values of (x, y, z). Fully characterizing all such (x, y, z) is beyond the scope of the present work. However, we can derive a general collection of necessary conditions from a subset of extremal vertices on the boundary of S. The details of the calculation are given in App. 4.A, but the vertices are

$$\begin{aligned} v_1 &= \left(0, 0, \tfrac{1}{2}\right) \\ v_2 &= \left(1, 1-\tfrac{3}{D-2}+\tfrac{1}{D-3}, 0\right) \\ v_3 &= \left(0, \tfrac{D-4}{2(D-2)}, 0\right) \\ v_4 &= \left(1, \tfrac{1}{D-2}\left[1+\tfrac{4(D \bmod 2)}{(D-1)(D-3)}\right], 0\right) \\ v_5 &= (0,0,0)\,. \end{aligned} \tag{4.3.18}$$

These vectors can be realized by physical polarization choices. The bounds associated with the (x, y, z) values in Eq. (4.3.18) are necessary for analyticity of four-point scattering amplitudes and moreover are a subset of the minimal basis of sufficient bounds. Numerical evaluation, via the explicit computation of x, y, and z for pseudorandom, traceless, unit-norm matrix pairs of various dimensions, shows that the convex hull defined by the vertices in Eq. (4.3.18) is in fact equal to the full hull S for even D but is slightly smaller than S in odd D. Note that the vectors (4.3.18) are a generalization of those we saw in earlier sections, so v_1, v_2, and v_3 coincide with the vectors from $D = 5$. Moreover, each corner corresponds to a certain extreme configuration of polarizations. For example, v_1 corresponds anticommuting polarizations as in Eq. (4.3.14), while v_2, v_3, v_4, and v_5 correspond to commuting polarizations. For the latter, the polarizations are mutually diagonalizable and can without loss of generality be represented as traceless diagonal matrices. See App. 4.A for details.

Plugging the vectors in Eq. (4.3.18) into the bound in Eq. (4.1.4), we obtain the positivity bounds

$$\begin{aligned}
-8c_3 + 4c_4 + c_6 + 2c_7 &> 0 \\
2\left(1 - \tfrac{3}{D-2} + \tfrac{1}{D-3}\right)(4c_2 + 8c_3 + 4c_4 + c_6 + c_7) & \\
+32c_1 + 4c_2 + 4c_6 + c_7 &> 0 \\
\left(\tfrac{D-4}{D-2}\right)(4c_2 + 8c_3 + 4c_4 + c_6 + c_7) + 2c_6 + c_7 &> 0 \\
\left(\tfrac{2}{D-2}\right)\left[1 + \tfrac{4(D \bmod 2)}{(D-1)(D-3)}\right](4c_2 + 8c_3 + 4c_4 + c_6 + c_7) & \\
+32c_1 + 4c_2 + 4c_6 + c_7 &> 0 \\
2c_6 + c_7 &> 0,
\end{aligned} \tag{4.3.19}$$

which are a stringent set of requirements on quartic curvature corrections to general relativity in $D \geq 6$, necessary to guarantee analyticity of scattering amplitudes.

4.3.4 Supersymmetric Theories

We now consider supersymmetric quartic curvature corrections. Conveniently, Ref. [146] derived a basis for independent off-shell supersymmetric quartic curvature invariants,

$$\mathcal{L}_4 = A\mathcal{O}_A + B\mathcal{O}_B + C\mathcal{O}_C, \tag{4.3.20}$$

where $\mathcal{O}_A$, $\mathcal{O}_B$, and $\mathcal{O}_C$ are proportional to more familiar looking forms denoted in the literature [146, 147] by $t_8t_8R^4$, $t_8(R^2)^2$, and $\epsilon_{10}\epsilon_{10}R^4$, respectively. In terms of the basis defined in Eq. (4.1.4), these supersymmetric operators are

$$\begin{aligned}
\mathcal{O}_A &= \mathcal{O}_1 - 16\mathcal{O}_2 + 2\mathcal{O}_3 - 32\mathcal{O}_5 + 16\mathcal{O}_6 + 32\mathcal{O}_7 \\
\mathcal{O}_B &= -\mathcal{O}_1 + 8\mathcal{O}_2 - 2\mathcal{O}_3 + 4\mathcal{O}_4 \\
\mathcal{O}_C &= \mathcal{O}_1 - 16\mathcal{O}_2 + 2\mathcal{O}_3 + 16\mathcal{O}_4 - 32\mathcal{O}_5 + 16\mathcal{O}_6 - 32\mathcal{O}_7,
\end{aligned} \tag{4.3.21}$$

corresponding to the following choice of operator coefficients:

$$\begin{aligned}
c_1 &= A - B + C & c_2 &= -16A + 8B - 16C \\
c_3 &= 2A - 2B + 2C & c_4 &= 4B + 16C \\
c_5 &= -32A - 32C & c_6 &= 16A + 16C \\
c_7 &= 32A - 32C.
\end{aligned} \tag{4.3.22}$$

Plugging this choice into Eq. (4.3.4), we obtain

$$A + B(y + z) > 0. \tag{4.3.23}$$

Note that C drops out of the calculation completely, since at quartic order in graviton perturbations it is a total derivative in all dimensions [143]. Recall from App. 4.A that while simple matrix identities imply that $y + z \leq 1$, no point in the hull S actually saturates this bound. For example, in $D = 4$, Eq. (4.3.10) implies that $y + z = \frac{1}{2}$, so $A + \frac{1}{2}B > 0$. In $D = 5$, Eq. (4.3.14) implies that $\frac{1}{6} \leq y + z \leq \frac{1}{2}$, so $A + \frac{1}{6}B > 0$ and $A + \frac{1}{2}B > 0$. Finally, in $D \geq 6$, inputting the vertices in Eq. (4.3.18) into Eq. (4.3.23) yields the complete set of positivity bounds for quartic curvature operators in supergravity theories. In summary, we find

$$\begin{aligned} A + \tfrac{1}{6}B > 0 \qquad & (D = 5) \\ A > 0 \qquad & (D \geq 6) \\ A + \left(1 - \tfrac{3}{D-2} + \tfrac{1}{D-3}\right) B > 0, \qquad & (\text{any } D) \end{aligned} \tag{4.3.24}$$

noting that in $D = 4$ and $D = 5$, the final bound reduces to $A + \frac{1}{2}B > 0$.

4.3.5 String Theories

As a consistency check, we now apply our bounds to string theory, which is arguably the leading candidate for the ultraviolet completion of gravity. Conveniently, quartic curvature corrections have been been dutifully computed at tree-level in the existing literature for the bosonic [148, 149], type II [150–152], and heterotic string [151, 152]. The type I string is dual to the heterotic string and has the same low-energy effective action [24, 147], so we need not consider it as a separate case. The resulting effective theory is described by

$$\mathcal{L}_4 = A\mathcal{O}_A + B\mathcal{O}_B + C\mathcal{O}_C + \Delta\mathcal{O}_\Delta, \tag{4.3.25}$$

where $\mathcal{O}_A$, $\mathcal{O}_B$, and $\mathcal{O}_C$ are the supersymmetric operators from the previous section and $\mathcal{O}_\Delta$ is a non-supersymmetric operator defined as

$$\mathcal{O}_\Delta = -\mathcal{O}_1 + 10\mathcal{O}_2 + \mathcal{O}_4. \tag{4.3.26}$$

In various string theories, the operator coefficients are

	A	B	C	Δ
bosonic	$\zeta(3)$	0	$-\zeta(3)$	16
type II	$\zeta(3)$	0	$-\zeta(3)$	0
heterotic	$\zeta(3)$	1	$-\zeta(3)$	0

(4.3.27)

where each entry is normalized by a factor of $\alpha'^3/1024\kappa^2$.

As expected, since the type II and heterotic string theories are supersymmetric, their coefficients in Eq. (4.3.27) satisfy the bound for supersymmetric theories in Eq. (4.3.24). Since the bosonic string is non-supersymmetric, the

bound is more complicated. In particular, plugging the corresponding operator coefficients into Eq. (4.3.4), we obtain

$$\zeta(3) + 2(x + 11y + z) > 0, \tag{4.3.28}$$

which is indeed positive, as $x, y, z \geq 0$. Thus, we have verified that quartic curvature corrections in bosonic, type II, and heterotic string theory are consistent with unitarity and analyticity.

4.4 Bounds on Quadratic Curvature Corrections

Next, we consider analyticity constraints on $\mathcal{L}_2$, which characterizes quadratic curvature corrections in the graviton effective theory. As shown in Ref. [139], the Gauss-Bonnet term

$$\mathcal{L}_2 = \lambda(R_{\mu\nu\rho\sigma}R^{\mu\nu\rho\sigma} - 4R_{\mu\nu}R^{\mu\nu} + R^2) \tag{4.4.1}$$

does not introduce ghost modes in any dimension D, so in this basis the graviton propagator is unmodified. To avoid ghost pathologies, we only consider curvature invariants of this form. For $D = 4$, the Gauss-Bonnet term is furthermore a total derivative and thus does not affect local dynamics. As recently shown [153], however, the Gauss-Bonnet term is critical for computing and interpreting the leading ultraviolet divergences of pure gravity.

Expanding to leading order in the Gauss-Bonnet coefficient λ, we compute the quadratic curvature correction to the graviton scattering amplitude in the forward limit,

$$\begin{aligned} &M_2(s, t \to 0) = \\ &4\lambda\kappa^4 s^2 \Big[\epsilon_1^{\mu\nu}\epsilon_{3\mu\nu}\epsilon_2^{\rho\sigma}\epsilon_{4\rho\sigma} + \epsilon_1^{\mu\nu}\epsilon_{3\nu\rho}\epsilon_2^{\rho\sigma}\epsilon_{4\sigma\mu} + \epsilon_1^{\mu\nu}\epsilon_{3\nu\rho}\epsilon_4^{\rho\sigma}\epsilon_{2\sigma\mu} \\ &\qquad + \frac{2}{t}\left(k_2^\mu k_4^\nu \epsilon_{2\nu}{}^\rho \epsilon_{4\rho\mu}\epsilon_{1\alpha\beta}\epsilon_3^{\alpha\beta} + k_1^\mu k_3^\nu \epsilon_{1\nu}{}^\rho \epsilon_{3\rho\mu}\epsilon_{2\alpha\beta}\epsilon_4^{\alpha\beta}\right)\Big], \end{aligned} \tag{4.4.2}$$

where we have expanded formally in t-dependence arising from propagator denominators, but we have yet to evaluate the numerators in the forward limit.

The first line of Eq. (4.4.2) is manifestly regular in the forward limit $t = 0$, so for these terms we can simply set $\epsilon_3 = \epsilon_1$ and $\epsilon_4 = \epsilon_2$. On the other hand, the second line of Eq. (4.4.2) is naively singular since $1/t$ diverges as $t \to 0$. However, this singularity is canceled by the numerator factor, which vanishes in the forward limit as $\epsilon_3 \to \epsilon_1$ and $\epsilon_4 \to \epsilon_2$. It will be convenient to rewrite this expression in terms of the momentum transfer,

$$q = k_1 + k_3 = -(k_2 + k_4), \tag{4.4.3}$$

where $t = -q^2$. For real kinematics, q is spacelike and vanishes in the forward limit. Note that $q^\mu q^\nu/q^2$ is simply a projection operator in the direction of the spacelike exchanged momentum.

We note that k_3 is simply a real spatial rotation of $-k_1$, and likewise for k_4 and k_2. By symmetry, this then implies that $\epsilon_1^{\mu\nu} k_{3\nu} = \epsilon_3^{\mu\nu} k_{1\mu} = \epsilon_1^{\mu\nu} q_\mu$ and $\epsilon_2^{\mu\nu} k_{4\mu} = \epsilon_4^{\mu\nu} k_{2\mu} = -\epsilon_2^{\mu\nu} q_\mu$ at leading order in q. Rewriting Eq. (4.4.2) in terms of q, we then have

$$\begin{aligned} &M_2(s, t \to 0) = \\ &4\lambda\kappa^4 s^2 \Big[\epsilon_{1\mu\nu}\epsilon_1{}^{\mu\nu}\epsilon_{2\rho\sigma}\epsilon_2{}^{\rho\sigma} + 2\epsilon_1^{\mu\nu}\epsilon_{1\nu\rho}\epsilon_2^{\rho\sigma}\epsilon_{2\sigma\mu} \\ &\qquad - \frac{2q^\mu q^\nu}{q^2}\left(\epsilon_{2\mu}{}^\rho \epsilon_{2\rho\nu}\epsilon_{1\alpha\beta}\epsilon_1{}^{\alpha\beta} + \epsilon_{1\mu}{}^\rho \epsilon_{1\rho\nu}\epsilon_{2\alpha\beta}\epsilon_2{}^{\alpha\beta}\right)\Big], \end{aligned} \tag{4.4.4}$$

which is regular because the projection operator $q^\mu q^\nu/q^2$ is finite in the forward limit. To obtain a bound on λ, we consider all possible choices for the external momenta and polarizations and impose positivity bounds on the forward amplitude in Eq. (4.4.4).

As expected, quadratic curvature corrections to graviton scattering scale as $M_2 \sim \lambda\kappa^4 s^2$, so to extract an analyticity bound we should apply Eq. (4.2.12) for a second-order residue, corresponding to $n = 2$. Unfortunately, this choice

also extracts the t-channel singular contribution from leading-order graviton exchange, $M_1 \sim -\kappa^2 s^2/t$. In the forward limit, this contribution is formally infinite. Of course, in any physical experiment there is an infrared scale μ that regulates these contributions from long distance physics. This would arise, e.g., from a finite detector resolution or beam width [154]. As is common practice for infrared divergences in scattering amplitudes, we introduce a mass regulator, sending $t \to t - \mu^2$ in the denominator. This approach was also used in Ref. [31] to make sense of a theory of interacting massless scalars with trilinear couplings. Note that as in gauge theory, the mass μ^2 is a formal regulator that leaves the number degrees of freedom untouched—so the vDVZ discontinuity [39, 40], which arises for a physical graviton mass included via a Fierz-Pauli Lagrangian term, does not apply here.

While μ^2 tames the formal infrared divergences, for $\lambda \lesssim 1$ the forward amplitude will be dominated by finite but large contributions from Einstein-Hilbert interactions because $|M_1| \gg |M_2|$ in this regime.[4.5] However, by explicit calculation, we can see from Eq. (4.2.14) that $M_1 \sim +\kappa^2 s^2/\mu^2$, which is positive. So while positivity is satisfied, we learn nothing beyond what is already borne out from scattering via the leading Einstein-Hilbert term.

To place a bound on the coefficient λ, we must then restrict to a parameter regime where $|M_1| \lesssim |M_2|$, so the contributions from graviton exchange are subdominant to those from the Gauss-Bonnet term. This implies that $1/\mu^2 \lesssim |\lambda\kappa^2|$. Together with the requirement that $|s| \gg \mu^2$, necessary to treat μ as a regulator, this forces us to consider the regime

$$\sqrt{|s|} \gg \mu \gtrsim \Lambda, \tag{4.4.5}$$

where $\Lambda \sim |\lambda\kappa^2|^{-1/2}$ is the scale of the would-be natural cutoff associated with

[4.5] Note that taking t strictly to zero is not required to derive a positivity bound [155] and positivity holds for any non-negative t below μ^2 [156]. However, we will not need this more general result for our purposes.

the derivative expansion. We assume throughout this chapter that $\Lambda \ll \kappa^{\frac{2}{2-D}}$ so that it is below the Planck scale in D dimensions.

Of course, Eq. (4.4.5) points to a naively pathological region of the effective field theory, given the reasonable expectation of new degrees of freedom of mass m where $m \sim \Lambda$. Moreover, Eq. (4.4.5) indicates that the infrared regulator μ must be larger than some other energy scale Λ. Nevertheless, one can a priori envision an ultraviolet completion in which $m \gg \Lambda$, so new degrees of freedom enter at a parametrically higher scale. In that case, Λ is not the scale of any physical states in the theory and is merely the combination of parameters that appears in the higher-dimension operator. Indeed, at the level of the scattering amplitude, there are no discontinuities that appear around Λ to signal new degrees of freedom.

Thus, μ remains smaller than any physical mass scale in the theory and indeed can be treated consistently as an infrared regulator. In the absence of new states at Λ, the Gauss-Bonnet term acts effectively as a primordial contact operator over a wide range of scales. Precisely such a scenario was considered in Ref. [37], where it was found that such a low-energy effective theory is acausal without new states at or below $m \lesssim \Lambda$. Other authors [157] have likewise argued that a pure Gauss-Bonnet theory is inconsistent with black hole thermodynamics. We will likewise find a pathology in this theory coming from unitarity and analyticity.

To apply constraints from unitarity and analyticity, we must first ensure that the low-energy theory is sensible enough that we can even speak of a long-distance scattering amplitude. Indeed, Eq. (4.4.5) is plainly strange since $|s| \gg \Lambda^2$ violates the derivative expansion. This was required in order for the Gauss-Bonnet interactions to dominate over the Einstein-Hilbert action, as was also assumed in Ref. [37]. Naively, one would expect a gross departure

from perturbative unitarity, e.g., probability amplitudes greater than one as well as a breakdown of the loop expansion. Nevertheless, there is a wide range of scales where neither sickness actually arises. This hinges on the fact that the theory depends on Λ as well as κ, the gravitational coupling constant.

In particular, note that amplitudes can still be perturbatively small in the regime specified by Eq. (4.4.5). For example, $M_2 \sim \kappa^2 s^2/\Lambda^2$ is still sensible provided κ is sufficiently small, corresponding to the weak gravity limit. We can make this more precise by considering the leading effect of the Gauss-Bonnet term, which is a cubic vertex of the schematic form $\lambda\kappa^3\partial^4 h^3$. Inserting this vertex into low-energy amplitudes, we find that the theory remains under perturbative control provided $\lambda\kappa^3$ ($\sim \kappa/\Lambda^2$) times the appropriate powers of energy is sufficiently small. In D dimensions this implies that

$$|s| \ll \left(\frac{\Lambda^2}{\kappa}\right)^{\frac{4}{2+D}} \tag{4.4.6}$$

to safely reside within the regime of perturbativity.[4.6] Moreover, Eq. (4.4.6) also ensures a perturbative loop expansion, since radiative corrections always introduce additional insertions of the Gauss-Bonnet interactions.

For our purposes, we assume a weak gravity limit defined by Eq. (4.4.6), so the low-energy theory is perturbatively unitary. When then apply the method of Sec. 4.2, where the contour around the origin in the complex s plane is widened so as to satisfy Eq. (4.4.5), ensuring that s is large compared to the scale of the infrared cutoff and that the Gauss-Bonnet term dominates the amplitude. Note also that the initial contour encircles a region below the heavy particle threshold, $m \gg \Lambda$.

To see how a pathology arises, it will be convenient to define coordinates transverse to the incoming momenta, $(x_1, x_2, \ldots x_{D-2})$. Without loss of general-

[4.6] A similar statement applies to pions, which have quartic vertices of the form $\partial^4\pi^4/\Lambda^2 v^2$ where v is the breaking scale and Λ controls the derivative expansion. The theory maintains perturbative control provided $s \ll \Lambda v$.

ity, we take the forward limit such that the infinitesimal momentum transfer lies in the x_1 direction, which we henceforth refer to as the "direction of approach." In turn, $q^\mu q^\nu/q^2$ is a projection operator onto this direction. Next, we define a particular subset of polarizations in the transverse plane, defined by rank-two diagonal matrices of the form

$$d^{(i,j)} = \frac{1}{\sqrt{2}}\,\mathrm{diag}(0,\ldots,0,\overbrace{1}^{x_i},0,\ldots,0,\overbrace{-1}^{x_j},0,\ldots), \tag{4.4.7}$$

with zero entries except in the x_i and x_j directions. As the only preferred direction is x_1, labeling the direction from which we approach the forward limit, the relevant physical polarizations are $d^{(1,2)}$, $d^{(2,3)}$, and $d^{(3,1)}$. The forward limit of the quadratic correction to the graviton scattering amplitude in Eq. (4.4.4) for various polarization combinations is

$$M(s,t\to 0) = \quad 2\lambda\kappa^4 s^2 \times \begin{cases} 0, & \epsilon_1=\epsilon_2=d^{(1,2)} \\ 4, & \epsilon_1=\epsilon_2=d^{(2,3)} \\ -1, & \epsilon_1=d^{(1,2)} \text{ and } \epsilon_2=d^{(1,3)}. \end{cases} \tag{4.4.8}$$

In the first case, $\epsilon_1=\epsilon_2=d^{(1,2)}$, corresponding to polarizations that have support in the direction of approach. In the case of $D=4$, this is required because the transverse space only has two dimensions. As expected, the amplitude vanishes in this regime because the Gauss-Bonnet term is a total derivative in $D=4$. Meanwhile, the second case, $\epsilon_1=\epsilon_2=d^{(2,3)}$ occurs when both polarizations are orthogonal to the direction of approach. Of course, this requires dimensions $D\geq 5$. Finally, in the last case, $\epsilon_1=d^{(1,2)}$ and $\epsilon_2=d^{(1,3)}$, the polarizations occupy different planes but share support in the direction of approach, which is only possible in $D\geq 5$.

The upshot of Eq. (4.4.8) is that in $D\geq 5$, different polarization configurations can yield opposite signs for the corrections to the forward scattering amplitude. As a result, this excludes both signs of λ and thus forbids it entirely.

Of course, we made the assumption of Ref. [37] that the Gauss-Bonnet term is an effectively primordial contact operator insofar as new degrees of freedom enter only at a scale far above the naive cutoff. Hence, the positivity violation in Eq. (4.4.8) simply implies that this assumption is false. We conclude that a primordial Gauss-Bonnet term is forbidden and new degrees of freedom are required at or below the cutoff Λ.

4.5 Conclusions

In this chapter, we have derived rigorous bounds on the coefficients of quartic and quadratic curvature corrections in the low-energy effective theory of gravitons. Our results hinge on very general principles: quantum mechanical unitarity and analyticity of scattering amplitudes. Consequently, these constraints apply to any consistent perturbative ultraviolet completion of gravity. For the quartic curvature operators defined in Eqs. (4.1.3), (4.1.4), (4.3.6), and (4.3.7), we derived the positivity bounds in Eq. (4.3.12) in $D = 4$, Eq. (4.3.17) in $D = 5$, and Eq. (4.3.19) in arbitrary $D \geq 6$. We also presented constraints on supergravity theories and checked that all of our results are consistent with known calculations in weakly-coupled string theories. For the quadratic curvature correction in Eq. (4.4.1), we showed that both signs of its coefficient λ are inconsistent unless new degrees of freedom enter at the natural cutoff $\Lambda \sim |\lambda\kappa^2|^{-1/2}$ specified by the effective theory. In short, a primordial Gauss-Bonnet term is forbidden.

Many possibilities remain for future work. While four-point graviton scattering cannot probe curvature operators beyond quartic order, little is known of higher-point amplitudes. Such amplitudes are functions of many more kinematic invariants and should thus enforce commensurately more positivity constraints. Another issue meriting further study is that of cubic curvature

operators. Here, positivity bounds encounter technical challenges due to the vanishing of the associated amplitude in the forward limit [156, 158]. As noted in Ref. [158], this problem is closely related to the a-theorem in $D = 6$.

Distinguishing low-energy effective theories that are consistent with ultraviolet completion from those that are not presents a significant challenge. Systematizing this procedure is important for delineating the space of possible physical laws, but has also become important for model-building in more phenomenological contexts [1] and in inflation [159–163]. In this chapter, the low-energy tools of analyticity and unitarity enabled us to find solutions to this problem in gravitational theories, allowing us to constrain higher-curvature corrections to gravity in our own universe—applying our quartic curvature results to $D = 4$—and further discover bounds applicable in any consistent theory.

4.A Bounding Invariants in General Dimension

We have shown that the graviton scattering amplitude can be expressed in terms of invariant products of graviton polarizations ϵ_1 and ϵ_2, which are real, symmetric, traceless $(D-2)$-by-$(D-2)$ matrices with unit normalization. To recapitulate from Sec. 4.3, given the Hermitian matrices

$$H_+ = \{\epsilon_1, \epsilon_2\}/2 \quad \text{and} \quad H_- = i[\epsilon_1, \epsilon_2]/2, \tag{4.A.1}$$

we can define the invariants

$$x = \mathrm{Tr}\,(H_+)\mathrm{Tr}\,(H_+), \qquad y = \mathrm{Tr}\,(H_+ \cdot H_+), \qquad z = \mathrm{Tr}\,(H_- \cdot H_-). \tag{4.A.2}$$

The space of physical polarizations ϵ_1 and ϵ_2 then maps onto a physical region in (x, y, z), which through Eq. (4.3.4) implies positivity constraints on operator coefficients in the effective theory.

What are the bounds on (x, y, z)? We first note that since H_+ and H_- are Hermitian, their squares are positive semidefinite, so $x, y, z \geq 0$. Moreover, a straightforward application of the Cauchy-Schwarz inequality implies $x = \mathrm{Tr}\,(\epsilon_1 \cdot \epsilon_2)\mathrm{Tr}\,(\epsilon_1 \cdot \epsilon_2) \leq \mathrm{Tr}\,(\epsilon_1 \cdot \epsilon_1)\mathrm{Tr}\,(\epsilon_2 \cdot \epsilon_2) = 1$, with equality if and only if $\epsilon_1 = \pm\epsilon_2$, and similarly $y + z = \mathrm{Tr}\,(\epsilon_1 \cdot \epsilon_1 \cdot \epsilon_2 \cdot \epsilon_2) \leq \mathrm{Tr}\,(\epsilon_1 \cdot \epsilon_1)\mathrm{Tr}\,(\epsilon_2 \cdot \epsilon_2) = 1$. There are, however, many additional constraints on (x, y, z), which we now discuss.

Crucially, a weighted average of any number of positivity bounds yields another valid positivity bound. This implies that a space of necessary conditions can be derived by constructing a convex hull S in (x, y, z) that contains the physically allowed region. Without loss of generality, S is

$$S = (x, y, z) = \left\{ \sum_{i=1}^{n} \tau_i v_i \;\middle|\; \tau_i \geq 0 \text{ and } \sum_{i=1}^{n} \tau_i = 1 \right\}, \tag{4.A.3}$$

where v_i denote extremal points. In this Appendix, we will construct the subset of the v_i that are on the edges of the unit cube in (x, y, z); let the convex hull described by these vertices be $\tilde{S}$. In even dimension, numerical evaluation suggests that $S = \tilde{S}$, while in odd $D > 6$, it is possible for points to lie slightly outside $\tilde{S}$.

Let us first consider the case where ϵ_1 and ϵ_2 are anticommuting, so $x = y = 0$ and we wish to maximize z. Going to a basis in which ϵ_1 is diagonal, we find that antisymmetry of $\epsilon_1 \cdot \epsilon_2$ implies that for each i, j,

$$(\epsilon_{1ii} + \epsilon_{1jj})\epsilon_{2ij} = 0 \tag{4.A.4}$$

and

$$z = -\mathrm{Tr}\,(\epsilon_1 \cdot \epsilon_2 \cdot \epsilon_1 \cdot \epsilon_2) = \sum_{i,j} \epsilon_{1ii}^2 \epsilon_{2ij}^2. \tag{4.A.5}$$

Since Eq. (4.A.4) implies $\epsilon_{1ii}\epsilon_{2ii} = 0$ for each i, the normalization condition $\sum_{i,j} \epsilon_{2ij}^2 = 1$ implies by Eq. (4.A.5) that nonzero diagonal terms in ϵ_2 can only decrease z. We therefore take ϵ_2 to have vanishing diagonal. Similarly, since

$\sum_i \epsilon_{1ii}^2$ is fixed to unity, we should require that, for each i for which $\epsilon_{1ii} \neq 0$, there exists j such that $\epsilon_{2ij} \neq 0$; letting $\epsilon_{1i_0i_0}$ be non-vanishing for some i_0 even if $\epsilon_{2i_0j} = 0$ for all j would decrease z by Eq. (4.A.5). Writing $\sum_j \epsilon_{2ij}^2 = \rho_i$, where $\sum_i \rho_i = 1$ by the normalization constraint, we can then consider z to be a weighted average over the ϵ_{1ii}^2. Thus, z is maximized when we weight the average most in favor of the i for which ϵ_{1ii}^2 is maximal. Suppose there are N such i, which we can without loss of generality take to be 1 through N, for which ϵ_{1ii}^2 takes its maximal value, i.e., $\epsilon_{1i^*i^*}^2 = \max_i \epsilon_{1ii}^2 \equiv \varepsilon^2$ for all $i^* \in \{1, ..., N\}$. Then z is maximized when we have $\rho_i = 1/N$ for $i \in \{1, ..., N\}$ and $\rho_i = 0$ otherwise, for which we obtain $z = \varepsilon^2$. Finally, it remains to determine the maximal possible value of ε^2. Since ϵ_1 is of unit norm, its maximal value is attained when we load as much of the normalization into as few of the ϵ_{1ii}^2 as possible. By tracelessness of ϵ_1, at least two of the ϵ_{1ii} must be nonzero. Thus, ε^2 takes its maximum value of $1/2$ when $\epsilon_1 \propto \sigma_3$ in some 2-by-2 block, up to permutation of coordinate labels. That is, a choice of polarizations that maximizes z for $x = y = 0$ is

$$\epsilon_1 = \tfrac{1}{\sqrt{2}}\sigma_3 \oplus 0_{D-4} \qquad \text{and} \qquad \epsilon_2 = \tfrac{1}{\sqrt{2}}\sigma_1 \oplus 0_{D-4}, \tag{4.A.6}$$

which yields the point

$$v_1 = \left(0, 0, \tfrac{1}{2}\right). \tag{4.A.7}$$

Let us henceforth consider the case where $z = 0$ and explore in x, y. This means that the (real, symmetric) matrices $\epsilon_{1,2}$ commute and so are simultaneously diagonalizable. Taking $x = 1$, we can ask how large y can be, which will give a vertex of S. Since $\epsilon_1 = \pm\epsilon_2$ for $x = 1$, we have $y = \mathrm{Tr}\,(\epsilon_1 \cdot \epsilon_1 \cdot \epsilon_1 \cdot \epsilon_1)$. That is, y has positive first and second derivatives in each of the $|\epsilon_{1ii}|$ values; y is therefore maximized when one of the $|\epsilon_{1ii}|$ is as large as possible and the others are equal and small. (If the smaller numbers in the list

were unequal, we could always make y larger by shifting some weight back to the element in the list with the largest absolute value.) That is, a choice of polarizations maximizing y for $x = 1$ and $z = 0$ is

$$\epsilon_1 = \epsilon_2 = \tfrac{1}{\sqrt{(D-2)(D-3)}}\mathrm{diag}(1, 1, \ldots, -(D-3)), \tag{4.A.8}$$

which corresponds to the vertex

$$v_2 = \left(1, 1 - \tfrac{3}{D-2} + \tfrac{1}{D-3}, 0\right). \tag{4.A.9}$$

Next, still taking $z = 0$, we consider a different extreme, setting $x = 0$ and maximizing y. Again simultaneously diagonalizing ϵ_1 and ϵ_2, we have $y = \sum_i \epsilon_{1ii}^2 \epsilon_{2ii}^2$. Analogously with the case of v_1, we can write $\rho_i = \epsilon_{1ii}^2$ and consider y to be a weighted average over the ϵ_{2ii}^2. Let $\max_i \epsilon_{2ii}^2 \equiv \varepsilon^2$ and, without loss of generality, suppose that $\epsilon_{2ii}^2 = \varepsilon^2$ for $i \in \{1, ..., N\}$ for some N. Then y is maximized if we take $\rho_i = 1/N$ for $i \in \{1, ..., N\}$ and $\rho_i = 0$ otherwise, in which case $y = \varepsilon^2$. Now, by the unit normalization of ϵ_2, ε^2 is maximized when as much of the normalization as possible is loaded into as few terms as possible, i.e., N is minimized. Since ϵ_1 is traceless, at least two of the ρ_i are nonzero, so $N \geq 2$. The maximum value of ε^2 thus occurs when $N = 2$, which fixes $\epsilon_1 \propto \sigma_3 \oplus 0_{D-4}$. We now must maximize the common absolute value of the first two entries in ϵ_{2ii}, subject to the constraints that $\sum_i \epsilon_{2ii} = 0$, $\sum_i \epsilon_{2ii}^2 = 1$, and, since $x = 0$, $\sum_i \epsilon_{1ii}\epsilon_{2ii} = 0$. This last constraint implies that the first two entries in ϵ_{2ii} have the same sign. Thus, y is maximized for $x = z = 0$ for the choice of polarizations

$$\begin{aligned} \epsilon_1 &= \tfrac{1}{\sqrt{2}}\sigma_3 \oplus 0_{D-4} \\ \epsilon_2 &= \sqrt{\tfrac{2}{(D-2)(D-4)}}\mathrm{diag}\left(\tfrac{D-4}{2}, \tfrac{D-4}{2}, -1, \ldots, -1\right), \end{aligned} \tag{4.A.10}$$

for which we find the vertex

$$v_3 = \left(0, \tfrac{D-4}{2(D-2)}, 0\right). \tag{4.A.11}$$

Note that the polarization configuration in Eq. (4.A.10), and hence the vertex

in Eq. (4.A.11), requires $D \geq 5$.

On the other hand, we can minimize y/x for $z = 0$. Using symmetry and reality to diagonalize $H_+ = \operatorname{diag} \vec{h}_+$, we have $x = |\vec{h}_+ \cdot \vec{n}|^2$, where $\vec{n} = (1, 1, \ldots, 1)$, so the Cauchy-Schwarz inequality implies

$$x \leq |\vec{h}_+|^2 |\vec{n}|^2 = (D-2)y. \tag{4.A.12}$$

Eq. (4.A.12) is saturated when $H_+ \propto 1_{D-2}$. If D is even, this choice is possible with

$$\epsilon_1 = \epsilon_2 = \tfrac{1}{\sqrt{D-2}} \operatorname{diag}(1, -1, 1, -1, \ldots), \tag{4.A.13}$$

which yields

$$(x, y, z) = \left(1, \tfrac{1}{D-2}, 0\right). \tag{4.A.14}$$

Let us now consider the odd-dimensional case where $z = 0$ and $x = 1$. Simultaneously diagonalizing ϵ_1 and ϵ_2, we have $y = \sum_i \epsilon_{1ii}^4$. Again, y has positive first and second derivatives in $|\epsilon_{1ii}|$, so it is minimized when the $|\epsilon_{1ii}|$ are all equal. In odd dimension, this is not possible while retaining tracelessness, so the best one can do, making the $|\epsilon_{1ii}|$ as equal as possible, is the choice

$$\epsilon_1 = \epsilon_2 = \sqrt{\tfrac{D-3}{(D-1)(D-2)}} \operatorname{diag}\Big(\underbrace{1, \ldots, 1}_{\frac{D-1}{2}}, -\tfrac{D-1}{D-3}, ..., -\tfrac{D-1}{D-3}\Big), \tag{4.A.15}$$

which results in the vertex

$$(x, y, z) = \left(1, \tfrac{2}{D-1} + \tfrac{2}{D-3} - \tfrac{3}{D-2}, 0\right). \tag{4.A.16}$$

We can combine Eqs. (4.A.14) and (4.A.16) to write the vertex of S as

$$v_4 = \left(1, \tfrac{1}{D-2}\left[1 + \tfrac{4(D \bmod 2)}{(D-1)(D-3)}\right], 0\right). \tag{4.A.17}$$

We note that for both $D = 4$ and $D = 5$, v_2 and v_4 are the same point. Moreover, v_3 generalizes the third vertex applicable in $D = 5$, while the polarization choice for v_3 does not apply in $D = 4$. In $D \geq 6$, there is one remaining linearly independent vertex, which can be obtained by choosing

$\epsilon_1 \cdot \epsilon_2 = 0_{D-2}$, e.g.,

$$\epsilon_1 = \tfrac{1}{\sqrt{2}}\sigma_3 \oplus 0_{D-4} \qquad \text{and} \qquad \epsilon_2 = \tfrac{1}{\sqrt{2}} 0_{D-4} \oplus \sigma_3, \tag{4.A.18}$$

which results in the point

$$v_5 = (0, 0, 0). \tag{4.A.19}$$

Together, Eqs. (4.A.7), (4.A.9), (4.A.11), (4.A.17), and (4.A.19) are the vertices of $\tilde{S}$ given in Eq. (4.3.18). They correspond via Eq. (4.3.4) to a set of linearly independent bounds (4.3.19) that must be satisfied in any gravity theory in $D \geq 6$.

Chapter 5
Positive Signs in Massive Gravity

We derive new constraints on massive gravity from unitarity and analyticity of scattering amplitudes. Our results apply to a general effective theory defined by Einstein gravity plus the leading soft diffeomorphism-breaking corrections. We calculate scattering amplitudes for all combinations of tensor, vector, and scalar polarizations. The high-energy behavior of these amplitudes prescribes a specific choice of couplings that ameliorates the ultraviolet cutoff, in agreement with existing literature. We then derive consistency conditions from analytic dispersion relations, which dictate positivity of certain combinations of parameters appearing in the forward scattering amplitudes. These constraints exclude all but a small island in the parameter space of ghost-free massive gravity. While the theory of the "Galileon" scalar mode alone is known to be inconsistent with positivity constraints, this is remedied in the full massive gravity theory.

This chapter is from Ref. [4], C. Cheung and G. N. Remmen, "Positive Signs in Massive Gravity," JHEP ***04*** *(2016) 002,* arXiv:1601.04068 [hep-th].

5.1 Introduction

Local symmetry breaking is a central concept in quantum field theory with a rich theoretical structure and ubiquitious applications to natural phenomena. While this subject is textbook material in the context of gauge theories, its gravitational analogue remains an active field of study. In particular, theories of massive gravity have spawned an extensive body of literature analyzing its

formal aspects and phenomenology (see Ref. [38] and references therein).

In this chapter, we present new constraints on the parameter space of massive gravity coming from the consistency of scattering amplitudes. For the sake of generality, we assume an effective theory for massive gravity comprised of general relativity plus soft diffeomorphism-breaking corrections proportional to the graviton mass [38]. The theory contains five degrees of freedom: two tensors, two vectors, and one scalar, which is known in the literature as the "Galileon." Importantly, we work in unitarity gauge so that the tensor, vector, and scalar modes are manipulated together as a multiplet rather than as decoupled states in the limit of Goldstone equivalence [42].

To eliminate ghost modes, we restrict to the parameter space of ghost-free massive gravity [43, 44], which is the nonlinear generalization of the Fierz-Pauli tuning for the graviton mass. Notably, ghost-free massive gravity has a parametrically higher cutoff than a generic massive gravity theory [43] and the resulting action has two free coupling constants, (c_3, d_5) [38].

After an intensive computation, we arrive at lengthy expressions for the general tree-level amplitude for the scattering of massive gravitons. As we will show in detail, analyticity and unitarity place positivity constraints on the coefficients that appear in the forward amplitude. Imposing positivity on *all possible* graviton scattering processes, we sculpt an allowed region in (c_3, d_5). For external states that are described by pure tensor, vector, or scalar polarizations—which we dub "definite-helicity" states—we obtain the excluded colored regions shown in Fig. 5.2. Expanding to the scattering of arbitrary superpositions of tensors, vectors, and scalars—which we dub "indefinite-helicity" states—we derive more stringent constraints, leaving a compact allowed region in (c_3, d_5) permitted by unitarity and analyticity shown in Fig. 5.3.

While this result excludes much of the parameter space of massive gravity,

it is actually a boon to the Galileon, which as a stand-alone effective theory actually fails analyticity bounds [31, 156, 164]. However, since this failure is marginal, corrections to the limit of Goldstone equivalence can tip the balance to restore analyticity in the theory. Thus, non-analyticity of the original Galileon may be corrected by embedding it into the full theory of massive gravity.

The structure of this chapter is as follows. In Sec. 5.2, we describe a general effective theory for massive gravity. Next, we compute the massive graviton scattering amplitudes in Sec. 5.3 and verify that they are consistent with existing literature. Finally, in Sec. 5.4 we present our new bounds from analytic dispersion relations, discuss implications in Sec. 5.5, and conclude in Sec. 5.6.

5.2 Effective Theory for Massive Gravity

We consider a general effective theory for massive gravity defined by the Einstein-Hilbert term plus soft diffeomorphism-breaking operators [38]. This starting point is familiar from other contexts, e.g., soft breaking of gauge symmetry or supersymmetry. In such instances, hard symmetry breaking should be avoided since it is radiatively unstable. The action for the massive gravity effective theory is

$$S = \frac{m_{\rm Pl}^2}{2}\int \mathrm{d}^4x\sqrt{-g}\left[R - \frac{m^2}{4}V(g,h)\right]. \tag{5.2.1}$$

The metric is $g_{\mu\nu} = \eta_{\mu\nu} + h_{\mu\nu}$, where $\eta_{\mu\nu}$ is the flat metric, in this chapter using mostly-plus signature, and $h_{\mu\nu}$ corresponds to the graviton. Here m is the soft breaking parameter, to be identified with the graviton mass shortly. Throughout this chapter, $m_{\rm Pl} = 1/\sqrt{8\pi G}$ is the reduced Planck mass.

The graviton potential terms take the general form

$$\begin{aligned} V(g,h) &= V_2(g,h) + V_3(g,h) + V_4(g,h) + \cdots \\ V_2(g,h) &= +b_1\langle h^2\rangle + b_2\langle h\rangle^2 \\ V_3(g,h) &= +c_1\langle h^3\rangle + c_2\langle h\rangle^2\langle h\rangle + c_3\langle h\rangle^3 \\ V_4(g,h) &= +d_1\langle h^4\rangle + d_2\langle h^3\rangle\langle h\rangle + d_3\langle h^2\rangle^2 + d_4\langle h^2\rangle\langle h\rangle^2 + d_5\langle h\rangle^4, \end{aligned} \tag{5.2.2}$$

where angle brackets denote full metric contractions: $\langle h\rangle = g^{\mu\nu}h_{\mu\nu}$, $\langle h^2\rangle = g^{\mu\nu}h_{\nu\rho}g^{\rho\sigma}h_{\sigma\mu}$, etc.

We assume the Fierz-Pauli form for the graviton mass terms,

$$b_1 = -b_2 = 1, \tag{5.2.3}$$

so the linearized theory describes a massive graviton with five polarizations: two tensors, two vectors, and one scalar. Without the Fierz-Pauli tuning in Eq. (5.2.3), the Hamiltonian loses a constraint, activating a scalar ghost degree of freedom [38].

At the nonlinear level, however, numerous pathologies arise. For example, Boulware and Deser [41] observed that a dangerous ghost degree of freedom is reintroduced in nontrivial backgrounds. Moreover, the high-energy behavior of the amplitude signals a parametrically low cutoff Λ_5 [42], where for later convenience we define

$$\Lambda_n = (m^{n-1}m_{\rm Pl})^{1/n}. \tag{5.2.4}$$

More recently, it was observed that the Boulware-Deser ghost can be eliminated with the proper choice of parameters [43, 44, 165]. In particular, working in the high-energy theory of scalars, the couplings at each power in the graviton can be chosen to yield total derivative interactions. For example, in Eq. (5.2.2) this parameter choice corresponds to

$$\begin{aligned} &c_1 = 2c_3 + \frac{1}{2}, \qquad c_2 = -3c_3 - \frac{1}{2}, \qquad d_1 = -6d_5 + \frac{3}{2}c_3 + \frac{5}{16}, \\ &d_2 = 8d_5 - \frac{3}{2}c_3 - \frac{1}{4}, \quad d_3 = 3d_5 - \frac{3}{4}c_3 - \frac{1}{16}, \quad d_4 = -6d_5 + \frac{3}{4}c_3, \end{aligned} \tag{5.2.5}$$

with c_3 and d_5 free parameters. The resulting theory is a nonlinear generalization of the Fierz-Pauli term. Moreover, the theory enjoys a parametrically higher cutoff Λ_3 [42, 43], since the parameter choice eliminates dangerous scalar self-interactions.

5.3 Calculation of Scattering Amplitudes

For our analysis, we have computed the general tree-level amplitude for massive graviton scattering. In what follows, we describe the setup and notation of our amplitudes calculation, followed by a set of consistency checks for our final expressions.

5.3.1 Setup and Notation

A massive graviton has a momentum vector k_μ satisfying $k_\mu k^\mu = -m^2$. To construct a basis of polarization tensors, we decompose the space orthogonal to k_μ in terms of a basis of three polarization vectors ϵ^i_μ satisfying

$$k^\mu \epsilon^i_\mu = 0 \tag{5.3.1}$$

and split according to transverse ($i = 1, 2$) and longitudinal ($i = 3$) polarizations. For example, in a frame in which $k_\mu = (\omega, 0, 0, k)$ and $\omega = \sqrt{k^2 + m^2}$, the polarization vectors satisfy

$$\begin{aligned} \epsilon^1_\mu &= (0, 1, 0, 0) \\ \epsilon^2_\mu &= (0, 0, 1, 0) \\ \epsilon^3_\mu &= \frac{1}{m}(k, 0, 0, \omega), \end{aligned} \tag{5.3.2}$$

with the normalization $\epsilon^i_\mu \epsilon^{j\mu} = \delta^{ij}$. By construction, at high energies $\epsilon^3_\mu \sim k_\mu/m$, which is the Goldstone equivalence limit.

Next, we construct a basis of five polarization tensors $\epsilon^i_{\mu\nu}$, which are symmetric and satisfy the transverse traceless conditions

$$k^\mu \epsilon^i_{\mu\nu} = \epsilon^{i\ \mu}_{\mu} = 0, \tag{5.3.3}$$

normalized to $\epsilon^i_{\mu\nu}\epsilon^{j\mu\nu} = \delta^{ij}$. Here the tensor ($i = 1, 2$), vector ($i = 3, 4$), and scalar ($i = 5$) polarizations are[5.1]

$$\begin{aligned}
\epsilon^1_{\mu\nu} &= \frac{1}{\sqrt{2}}(\epsilon^1_\mu\epsilon^1_\nu - \epsilon^2_\mu\epsilon^2_\nu), & \epsilon^2_{\mu\nu} &= \frac{1}{\sqrt{2}}(\epsilon^1_\mu\epsilon^2_\nu + \epsilon^2_\mu\epsilon^1_\nu), \\
\epsilon^3_{\mu\nu} &= \frac{i}{\sqrt{2}}(\epsilon^1_\mu\epsilon^3_\nu + \epsilon^3_\mu\epsilon^1_\nu), & \epsilon^4_{\mu\nu} &= \frac{i}{\sqrt{2}}(\epsilon^2_\mu\epsilon^3_\nu + \epsilon^3_\mu\epsilon^2_\nu), \\
\epsilon^5_{\mu\nu} &= \sqrt{\frac{3}{2}}\left(\epsilon^3_\mu\epsilon^3_\nu - \frac{1}{3}\Pi_{\mu\nu}\right),
\end{aligned} \tag{5.3.4}$$

where we have defined the projection operator

$$\Pi_{\mu\nu} = \eta_{\mu\nu} + \frac{k_\mu k_\nu}{m^2}. \tag{5.3.5}$$

The polarizations satisfy the completeness relation,

$$\sum_i \epsilon^i_{\mu\nu}\epsilon^{i*}_{\rho\sigma} = \frac{1}{2}\left(\Pi_{\mu\rho}\Pi_{\nu\sigma} + \Pi_{\mu\sigma}\Pi_{\nu\rho}\right) - \frac{1}{3}\Pi_{\mu\nu}\Pi_{\rho\sigma}, \tag{5.3.6}$$

where the right side is the massive graviton propagator numerator. We will often denote the tensor, vector, and scalar polarizations schematically as T, V, and S, respectively. The last is also known in the literature as the Galileon [156, 164, 166, 167].

In terms of the explicit frame used in Eq. (5.3.2), the polarization tensors are

$$\begin{aligned}
\epsilon^1_{\mu\nu} &= \frac{1}{\sqrt{2}}\begin{pmatrix} 0&0&0&0\\ 0&1&0&0\\ 0&0&-1&0\\ 0&0&0&0 \end{pmatrix}, & \epsilon^2_{\mu\nu} &= \frac{1}{\sqrt{2}}\begin{pmatrix} 0&0&0&0\\ 0&0&1&0\\ 0&1&0&0\\ 0&0&0&0 \end{pmatrix}, \\
\epsilon^3_{\mu\nu} &= \frac{i}{\sqrt{2}m}\begin{pmatrix} 0&k&0&0\\ k&0&0&\omega\\ 0&0&0&0\\ 0&\omega&0&0 \end{pmatrix}, & \epsilon^4_{\mu\nu} &= \frac{i}{\sqrt{2}m}\begin{pmatrix} 0&0&k&0\\ 0&0&0&0\\ k&0&0&\omega\\ 0&0&\omega&0 \end{pmatrix}, \\
\epsilon^5_{\mu\nu} &= \sqrt{\frac{2}{3}}\frac{1}{m^2}\begin{pmatrix} k^2&0&0&k\omega\\ 0&-m^2/2&0&0\\ 0&0&-m^2/2&0\\ k\omega&0&0&\omega^2 \end{pmatrix},
\end{aligned} \tag{5.3.7}$$

[5.1] The overall phase of each polarization is unphysical, but we include a factor of i in the vector polarizations to manifest their odd parity under charge conjugation.

which can come in handy for explicit calculations.

The general scattering amplitude of massive gravitons, $M(ABCD)$, depends on the Mandelstam invariants (s,t) together with four external polarization tensors,

$$\begin{aligned} \epsilon^A_{\mu\nu} &= \sum_i \alpha_i \epsilon^i_{\mu\nu}, \qquad & \epsilon^B_{\mu\nu} &= \sum_i \beta_i \epsilon^i_{\mu\nu}, \\ \epsilon^C_{\mu\nu} &= \sum_i \gamma_i \epsilon^i_{\mu\nu}, \qquad & \epsilon^D_{\mu\nu} &= \sum_i \delta_i \epsilon^i_{\mu\nu}, \end{aligned} \tag{5.3.8}$$

where $\alpha, \beta, \gamma, \delta$ are unit vectors.

To determine constraints, we restrict to forward, crossing-symmetric amplitudes. The forward limit implies $t = 0$, which is a regular kinematic regime, as the graviton mass regulates all infrared singularities. Meanwhile, the constraint of crossing symmetry requires that

$$\epsilon^{C*}_{\mu\nu} = \epsilon^A_{\mu\nu} \quad \text{and} \quad \epsilon^{D*}_{\mu\nu} = \epsilon^B_{\mu\nu}. \tag{5.3.9}$$

Thus, the general scattering amplitude is a function of $(s,t,\alpha,\beta,\gamma,\delta)$ while the forward, crossing-symmetric amplitude is a function of (s,α,β). In order to maintain crossing symmetry simultaneously with the forward limit, we must assume linear polarizations for the external states [3], which means that the vectors α and β are real.

We have calculated the massive graviton scattering amplitude at general kinematics using the above definitions of the external polarization tensors, together with the Feynman rules extracted from Eq. (5.2.2) after going to canonical normalization where $h_{\mu\nu}$ is rescaled by $m_{\rm Pl}/2$. As our amplitudes expressions are prohibitively long, we have included them as supplemental material in Ref. [4].

5.3.2 Consistency Checks

To verify consistency we have studied the high-energy behavior for "definite-helicity" gravitons, which are strictly T, V, or S. From power counting, we know that the massive graviton modes enter the action as $T \sim \partial V \sim \partial\partial S$, so the high-energy behavior of amplitudes at fixed angle is

$$\begin{array}{lll} M(TTTT) \sim s, & M(TVTV) \sim s^2, & M(TSTS) \sim s^3, \\ M(VVVV) \sim s^3, & M(VSVS) \sim s^4, & M(SSSS) \sim s^5. \end{array} \tag{5.3.10}$$

Our explicit amplitude expressions agree with this scaling.

In particular, the amplitude for scalar scattering, $M(SSSS)$, is the worst-behaved at high energies and violates unitarity at scales of order Λ_5. We find that

$$M(SSSS) = -\frac{5(1-6c_1-4c_2)^2}{432\Lambda_5^{10}} stu(s^2+t^2+u^2) + \cdots, \tag{5.3.11}$$

in agreement with Ref. [168], which calculated this amplitude including just the Fierz-Pauli term. By choosing $1-6c_1-4c_2=0$, we can raise the cutoff from Λ_5 to Λ_4, so

$$M(SSSS) = \frac{3-16d_1-32d_3}{144\Lambda_4^8}(s^2+t^2+u^2)^2 + \cdots. \tag{5.3.12}$$

By choosing $3-16d_1-32d_3=0$, we can then further raise the cutoff from Λ_4 to Λ_3. Notably, these choices of parameters are consistent with Eq. (5.2.5), which we expected due to the improved cutoff in ghost-free massive gravity. This agreement is a nontrivial check that our calculation of the scattering amplitudes is correct.

Plugging in all the parameters of ghost-free massive gravity from Eq. (5.2.5), we find improved high-energy behavior scaling as

$$\begin{array}{lll} M(TTTT) \sim s, & M(TVTV) \sim s^2, & M(TSTS) \sim s^2, \\ M(VVVV) \sim s^3, & M(VSVS) \sim s^3, & M(SSSS) \sim s^3. \end{array} \tag{5.3.13}$$

From our explicit amplitudes, we find that there is no possible combination of parameters in the action (5.2.2) whereby the high-energy scaling of all amplitudes is s^2; if such a combination existed, it would raise the cutoff further. In particular, $M(VSVS)$ always scales as $\sim s^3$ or worse. This agrees with Ref. [169], which argued that high-energy scaling of $\sim s^2$ is impossible.

After plugging in Eq. (5.2.5), the leading behavior of the all-scalar amplitude is

$$M(SSSS) = -\frac{1 - 4c_3 + 36c_3^2 + 64d_5}{6\Lambda_3^6} stu + \cdots , \tag{5.3.14}$$

which vanishes for $(c_3, d_5) = (1/6, -1/48)$, a parameter choice that indeed results in non-interacting scalars in the decoupling limit of the Λ_3 theory [43]. As a highly nontrivial consistency check, we have verified that the leading high-energy behavior of $M(SSSS)$ in Eq. (5.3.14) is equal to the scattering amplitude for pure Galileons—including signs and numerical factors—as is mandated by the Goldstone equivalence theorem.

For the remainder of this chapter, we assume the parameter choice in Eq. (5.2.5), corresponding to ghost-free massive gravity.

5.4 Derivation of Constraints

In this section, we briefly review the mechanics of analytic dispersion relations for amplitudes and their relation to positivity. We then present our results constraining the parameter space of massive gravity.

5.4.1 Analytic Dispersion Relations

For our analysis, we apply analytic dispersion relations to the amplitude $M(s,t)$, for now dropping the labels for the external polarizations. As noted previously, the forward amplitude $M(s,0)$ is well defined since t-channel singularities are regulated by the graviton mass m. To begin, consider the contour integral

$$f = \frac{1}{2\pi i}\oint_\Gamma \mathrm{d}s\, \frac{M(s,0)}{(s-\mu^2)^3}, \tag{5.4.1}$$

where μ^2 corresponds to an arbitrary mass scale chosen in the interval $0 < \mu^2 < 4m^2$. The reason for this stipulation will become clear shortly.

At tree-level, $M(s,t)$ has singularities from massive graviton exchange at $s,t,u = m^2$, which in the forward limit generate simple poles at $s = m^2$ and $s = 3m^2$. Beyond tree-level, branch cuts arise from multi-particle production, which in the forward limit run from $s = 4m^2$ to $+\infty$ and from $s = 0$ to $-\infty$. The contour Γ in Eq. (5.4.1) is chosen to be a circle of radius at least m^2 and at most $2m^2$, centered on $s = 2m^2$, so that the contour contains the points $s = m^2$, $s = 3m^2$, and $s = \mu^2$, as depicted in Fig. 5.1.

We now use Cauchy's theorem to deform the contour Γ into a new contour Γ' shown in Fig. 5.1, which runs just above and below the real s axis for $s < 0$ and $s > 4m^2$, plus a boundary contour at infinity. Assuming the Froissart unitarity bound [170, 171], the forward amplitude grows sufficiently slowly with s that the boundary contribution at infinity vanishes [3, 31]. Thus,

$$f = \frac{1}{2\pi i}\oint_{\Gamma'} \mathrm{d}s\, \frac{M(s,0)}{(s-\mu^2)^3} = \frac{1}{2\pi i}\left(\int_{-\infty}^{0} + \int_{4m^2}^{\infty}\right)\mathrm{d}s \frac{\mathrm{Disc}\, M(s,0)}{(s-\mu^2)^3}, \tag{5.4.2}$$

where $\mathrm{Disc}\, M(s,0) = M(s+i\epsilon,0) - M(s-i\epsilon,0)$ for real s and infinitesimal positive ϵ. For the integral over the negative real s axis, we switch variables to $u = 4m^2 - s$, yielding

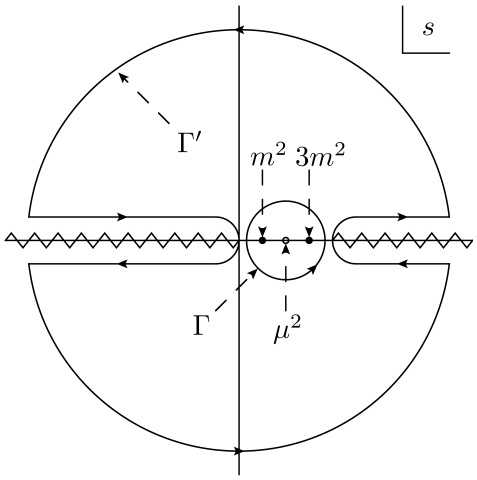

Figure 5.1. Diagram of the analytic structure of the forward amplitude in the complex s plane. The simple poles at $s = m^2$ and $3m^2$ and the branch cuts starting at $s = 4m^2$ and 0 correspond to resonances and multi-particle thresholds in the s- and u-channels, respectively. The scale μ^2 in the dispersion relation is chosen here to be at the symmetric point $\mu^2 = 2m^2$. The contours Γ and Γ' referred to in Eqs. (5.4.1) and (5.4.2) are also depicted.

$$\begin{aligned} f &= \frac{1}{2\pi i}\int_{4m^2}^{\infty} \mathrm{d}u \frac{\mathrm{Disc}\, M(4m^2-u,0)}{(4m^2-u-\mu^2)^3} + \frac{1}{2\pi i}\int_{4m^2}^{\infty} \mathrm{d}s \frac{\mathrm{Disc}\, M(s,0)}{(s-\mu^2)^3} \\ &= \frac{1}{2\pi i}\int_{4m^2}^{\infty} \mathrm{d}s \left[\frac{1}{(s-\mu^2)^3} + \frac{1}{(s+\mu^2-4m^2)^3}\right] \mathrm{Disc}\, M(s,0) \\ &= \frac{1}{\pi}\int_{4m^2}^{\infty} \mathrm{d}s \left[\frac{1}{(s-\mu^2)^3} + \frac{1}{(s+\mu^2-4m^2)^3}\right] \mathrm{Im}\, M(s,0). \end{aligned} \tag{5.4.3}$$

In the second line, we applied the definition $\mathrm{Disc}\, M(4m^2-u,0) = M(4m^2 - u + i\epsilon, 0) - M(4m^2 - u - i\epsilon, 0)$, followed by crossing symmetry, $M(u,0) = M(4m^2-u,0)$, thus yielding $\mathrm{Disc}\, M(4m^2-u) = M(u-i\epsilon) - M(u+i\epsilon) = -\mathrm{Disc}\, M(u)$, and then relabeled u to s as a dummy variable. In the third line, we used the Schwarz reflection principle $M(s^*,0) = [M(s,0)]^*$, so for real s we have $\mathrm{Disc}\, M(s,0) = 2i\,\mathrm{Im}\, M(s,0)$. Finally, by applying the optical theorem, $\mathrm{Im}\, M(s,0) = s\sigma(s)\sqrt{1-4m^2/s}$, we obtain our final expression,

$$f = \frac{1}{\pi}\int_{4m^2}^{\infty} \mathrm{d}s\, \sigma(s) \left[\frac{s}{(s-\mu^2)^3} + \frac{s}{(s+\mu^2-4m^2)^3}\right] \sqrt{1-\frac{4m^2}{s}} > 0, \tag{5.4.4}$$

where for an interacting theory the total cross-section $\sigma(s)$ is strictly positive. Since the integration region is restricted to $s > 4m^2$ and we stipulated earlier that $0 < \mu^2 < 4m^2$, the expressions in brackets and under the radical are strictly positive so f is as well.

We have applied well-known analytic dispersion relations to prove that $f > 0$. Crucially, from Eq. (5.4.1) we can derive f purely from the low-energy effective theory, so

$$0 < f = \left(\operatorname*{Res}_{s=m^2} \left[\frac{M(s,0)}{(s-\mu^2)^3} \right] + \operatorname*{Res}_{s=3m^2} \left[\frac{M(s,0)}{(s-\mu^2)^3} \right] + \operatorname*{Res}_{s=\mu^2} \left[\frac{M(s,0)}{(s-\mu^2)^3} \right] \right)_{\text{EFT}}, \tag{5.4.5}$$

where for emphasis we have included a subscript indicating that all quantities should be computed within the low-energy effective theory, *not* the full theory. There is, however, a shortcut to this calculation: since the poles of the low-energy scattering amplitude are known, we know by Cauchy's theorem that Eq. (5.4.5) can be calculated in a single step by computing the negative of its residue at large s,

$$f = -\left(\operatorname*{Res}_{s=\infty} \left[\frac{M(s,0)}{(s-\mu^2)^3} \right] \right)_{\text{EFT}} > 0, \tag{5.4.6}$$

which is our final expression for f.

Conveniently, we can show that f is μ^2-independent for ghost-free massive gravity. In particular, we saw earlier that fixed-angle scattering in ghost-free massive gravity scales as s^3. The only crossing-symmetric invariant at this order, stu, vanishes in the forward limit, so forward scattering scales as s^2. At large s we can expand $1/(s-\mu^2) = 1/s + \mathcal{O}(\mu^2/s^2)$, in which case only the μ^2-independent piece of Eq. (5.4.6) contributes. We have verified this to be the case in our explicit amplitudes.

Now we can reintroduce the dependence on the external polarization data.

Since the general amplitude is a quartic form in the polarizations $(\alpha, \beta, \gamma, \delta)$, the forward, crossing-symmetric amplitude is a real quartic form in (α, β). As f is a residue of the latter, it takes the form

$$f(\alpha, \beta) = \sum_{ijkl} f(ijkl)\alpha_i\beta_j\alpha_k\beta_l > 0. \tag{5.4.7}$$

Obviously, $f(ijkl)$ is symmetric under $i \leftrightarrow k$ and $j \leftrightarrow l$ due to the structure of the quartic form and also under $ik \leftrightarrow jl$ from exchange of the two incoming particles; that is,

$$f(ijkl) = f(kjil) = f(ilkj) = f(jilk). \tag{5.4.8}$$

In principle, these symmetries leave $f(ijkl)$ with 120 independent components, but as we will see, many of these are zero for the physical amplitude.

In the next subsection, we present $f(ijkl)$ and map the positivity bound from analytic dispersion relations onto the parameter space of massive gravity. We begin by studying "definite-helicity" gravitons described by pure tensor, vector, or scalar polarizations. Afterwards, we consider the "indefinite-helicity" case in which we are scattering superpositions of these states.

5.4.2 Bounds from Definite-Helicity Scattering

To begin, we consider the scattering of definite-helicity gravitons, corresponding to external polarizations that are purely tensor, vector, or scalar. Remarkably, for most combinations of definite-helicity modes, we find that the relative angles between polarizations drop out of our expressions. Writing

$$\begin{aligned} f(1111) = f(1212) = f(2222) &= f(TTTT) \\ f(1313) = f(1414) = f(2323) = f(2424) &= f(TVTV) \\ f(1515) = f(2525) &= f(TSTS) \\ f(3333) = f(4444) &= f(VVVV)_+ \\ f(3434) &= f(VVVV)_- \\ f(3535) = f(4545) &= f(VSVS) \\ f(5555) &= f(SSSS), \end{aligned} \tag{5.4.9}$$

expressed in terms of f for various scattering combinations of T, V, and S, we find, via explicit calculation, that

$$\begin{aligned} f(TTTT) &= \frac{1}{\Lambda_2^4} \\ f(TVTV) &= \frac{5-12c_3}{4\Lambda_2^4} \\ f(TSTS) &= \frac{5-12c_3}{3\Lambda_2^4} \\ f(VVVV)_+ &= \frac{5+72c_3-240c_3^2}{16\Lambda_2^4} \\ f(VVVV)_- &= \frac{23-72c_3+144c_3^2+192d_5}{16\Lambda_2^4} \\ f(VSVS) &= \frac{91-312c_3+432c_3^2+384d_5}{48\Lambda_2^4} \\ f(SSSS) &= \frac{14-12c_3-36c_3^2+96d_5}{9\Lambda_2^4}. \end{aligned} \tag{5.4.10}$$

Note that only in the case of all-vector scattering does f depend on the relative angle between external polarizations. For this reason, we had to define both $f(VVVV)_+$ and $f(VVVV)_-$, corresponding vector polarizations that are parallel and orthogonal, respectively. In contrast, the all-tensor case $f(TTTT)$, for example, is independent of the relative angle between the incoming tensor polarizations.

To obtain new positivity bounds, we simply demand that $f > 0$ for all polarization combinations in Eq. (5.4.10). These constraints can be cast as an excluded region in (c_3, d_5) space, as shown in Fig. 5.2. As one can see, considering the scattering of modes that are pure tensor, vector, or scalar is alone enough to rule out much of the parameter space of massive gravity, except for a strip in d_5 for certain values of c_3. In order to obtain the most stringent possible bounds, we turn to the question of scattering indefinite-helicity states in the next subsection, which will restrict the allowed parameter space to the region inside the black curve in Fig. 5.2.

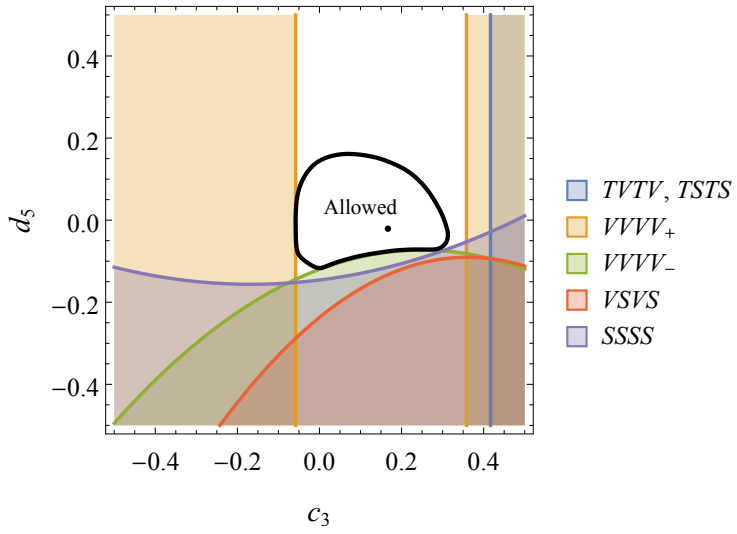

Figure 5.2. Regions in the (c_3, d_5) parameter space of ghost-free massive gravity excluded by analyticity bounds on scattering of definite-helicity gravitons. The tensor, vector, and scalar modes are denoted by T, V, and S, respectively, and the $\pm$ delineation indicates vector polarizations that are parallel or orthogonal, respectively. Ultimately, by considering indefinite-helicity scattering, we will further restrict the allowed region of parameter space to that within the black curve. The dot marks the parameter choice $(c_3, d_5) = (1/6, -1/48)$, which corresponds to a free scalar sector in the decoupling limit.

5.4.3 Bounds from Indefinite-Helicity Scattering

In general, it is possible to scatter arbitrary superpositions of tensor, vector, and scalar modes, corresponding to generic real unit vectors α and β. Our calculation shows that all $f(ijkl)$ vanish except for those in Eq. (5.4.9), together with

$$\begin{aligned} f(1133) = f(1144) &= f(2233) \\ &= f(2244) = -\frac{3(1-4c_3)^2}{8\Lambda_2^4} \\ f(1155) &= f(2255) = \frac{-1+8c_3-24c_3^2-16d_5}{2\Lambda_2^4} \\ f(1335) = -f(1445) &= f(2345) \\ &= f(2435) = \frac{\sqrt{3}(1-12c_3)^2}{96\Lambda_2^4} \end{aligned} \tag{5.4.11}$$

$$f(1353) = -f(1454) = f(2354) = \frac{\sqrt{3}(1 - 8c_3 + 48c_3^2 + 64d_5)}{16\Lambda_2^4}$$
$$f(3344) = \frac{-9 + 72c_3 - 192c_3^2 - 96d_5}{16\Lambda_2^4}$$
$$f(3355) = f(4455) = \frac{-17 + 136c_3 - 336c_3^2}{32\Lambda_2^4},$$

along with the $f(ijkl)$ related to these by the symmetries in Eq. (5.4.8). Varying (α, β) corresponds to different scattering experiments in which the scattered particles are various superpositions of polarizations. Imposing analyticity constraints on the amplitude for all possible scattering processes—that is, marginalizing over all possible choices of (α, β)—implies positivity bounds on the massive graviton parameter space that are much stronger than the bounds derived in the previous subsection.

For example, consider gravitons that are maximal superpositions of scalar and tensor,

$$\alpha_i = \beta_i = \frac{1}{\sqrt{2}}(\cos\phi, \sin\phi, 0, 0, 1). \tag{5.4.12}$$

For any value of ϕ, the corresponding scattering amplitude yields

$$f(\alpha, \beta) = \frac{35 + 60c_3 - 468c_3^2 - 192d_5}{36\Lambda_2^4}. \tag{5.4.13}$$

Requiring positivity of f then excludes arbitrarily large values of d_5, irrespective of c_3. In terms of the (c_3, d_5) parameter space, this example bound already eliminates all but a compact region of the semi-infinite strip of the parameter space permitted by the definite-helicity graviton scattering bounds shown in Fig. 5.2.

To place the most stringent bounds from analytic dispersion relations, we must find all points in (c_3, d_5) for which f is positive for all (α, β). That is, we must marginalize over all choices of external polarizations. Unfortunately, there is no analytic prescription for determining the positivity of quartic forms. While this algebraic problem is strongly NP-hard [172], it can be recast as a

dynamical problem [173] that is numerically tractable. In particular, let us repackage (α, β) into a new ten-dimensional "coordinate,"

$$X_I = (\alpha_1, \alpha_2, \alpha_3, \alpha_4, \alpha_5, \beta_1, \beta_2, \beta_3, \beta_4, \beta_5), \tag{5.4.14}$$

relaxing the normalization constraint $\alpha^2 = \beta^2 = 1$. Next, we assume that X_I evolves in time t according to an equation of motion,

$$\frac{\mathrm{d}X_I}{\mathrm{d}t} = -\frac{\partial f}{\partial X_I}. \tag{5.4.15}$$

This immediately implies that

$$\frac{\mathrm{d}f}{\mathrm{d}t} = -\sum_I \frac{\partial f}{\partial X_I}\frac{\partial f}{\partial X_I} \leq 0, \tag{5.4.16}$$

so f is non-increasing over time. Meanwhile, we know that as long as $X_I \neq 0$, then $\partial f/\partial X_I \neq 0$ since f is quartic in X_I. It thus follows that $\mathrm{d}f/\mathrm{d}t < 0$ strictly everywhere away from $X_I = 0$, i.e., f will decrease monotonically at all X_I except the origin. If there is a direction in which f is unbounded from below, then time evolution will drive it arbitrarily negative. On the other hand, a positive definite f will of course remain positive forever. As a result, f is positive definite if and only if f is stable under the time evolution of X_I.

Concretely, for a given numerical choice of (c_3, d_5), we initialize a random value of $X_I(t_{\text{init}})$, evolve in time to $X_I(t_{\text{final}})$, and then check whether $f(t_{\text{final}})$ is negative. If so, then the polarization choices given by $X_I(t_{\text{final}})$, suitably normalized, contradict the analyticity argument. Thus the parameter point (c_3, d_5) is inconsistent and we discard it. If $f(t_{\text{final}}) \geq 0$, the parameter point remains a possible viable theory. Iterating many times, we are able determine a definitive region in (c_3, d_5) that is excluded by analytic dispersion relations for all possible graviton scattering configurations.

The result of this calculation is that (c_3, d_5) are confined to a small compact region, as shown in Fig. 5.3. Here each colored point corresponds to a point in

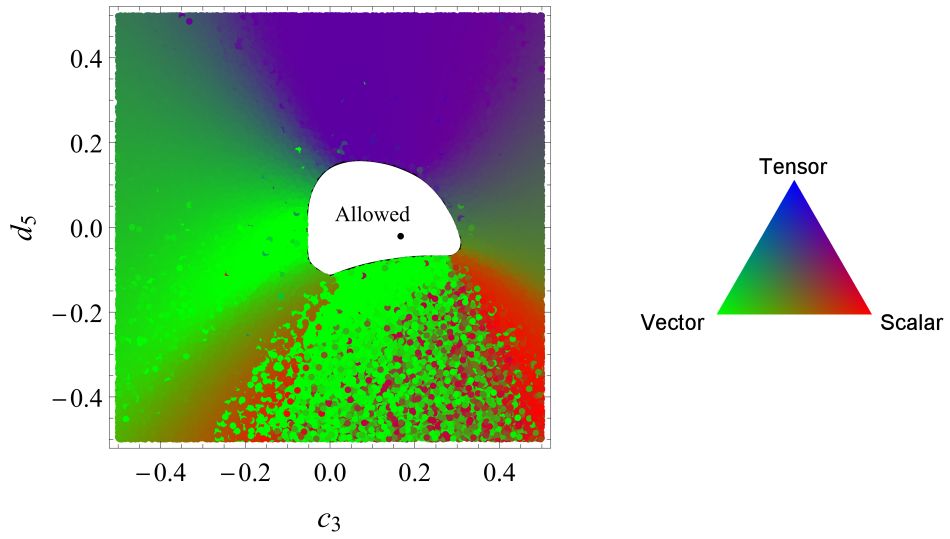

Figure 5.3. Region of (c_3, d_5) parameter space for ghost-free massive gravity excluded by analyticity bounds on scattering of indefinite-helicity gravitons. Each colored point corresponds to a theory excluded by a scattering process that violates analytic dispersion relations. As noted in text, such violations can be diagnosed by evolving a particular dynamical system that tends toward scattering processes of gravitons of similar polarization. The specific color—plotted in blue, green, and red—corresponds to the power of each polarization in tensors ($\alpha_1^2 + \alpha_2^2$ and $\beta_1^2 + \beta_2^2$), vectors ($\alpha_3^2 + \alpha_4^2$ and $\beta_3^2 + \beta_4^2$), and scalars (α_5^2 and β_5^2). The allowed region is shown in white and the black dot marks the choice that corresponds to a free Galileon.

parameter space for which our algorithm has determined a violation of analytic dispersion relations. The color of the point encodes the power distribution in the tensor, vector, and scalar components of the corresponding polarization excluding the point. Interestingly, we find that for many of the points that violate positivity, the numerical algorithm tends to converge to scattering processes in which the two scattered gravitons have the same power distribution.

5.5 Implications for Massive Gravity

Our bounds exclude most of the parameter space for ghost-free massive gravity, subject to the assumptions of analyticity and unitarity of the theory. While this is in part a negative result, the existence of a finite allowed region is actually encouraging, especially given the checkered history of the scalar mode

of massive gravity—the so-called Galileon.

As demonstrated early on, the Galileon is a remarkable effective theory in and of itself [164]. The model is uniquely fixed by an extended shift symmetry that highly constrains allowed interactions, limiting the action to a set of five Galilean-invariant operators in four dimensions. The Galileon is by construction ghost-free, which is natural since it describes the scalar mode of ghost-free massive gravity. Moreover it supports interesting cosmological solutions [174–176] and has scattering amplitudes with unique infrared properties [177].

On the other hand, it has long been known that the Galileon actually violates positivity bounds derived from analytic dispersion relations [31, 156, 164]. The reason is simple: the extended shift symmetry of the Galileon simply forbids interactions of the form $(\partial S)^4$, which induce s^2 contributions to the amplitude. Galileon interactions are instead of the form $(\partial S)^2(\partial\partial S)^2$, which mandates strict s^3 behavior of the fixed-angle amplitude, with no subleading corrections. In turn, the only crossing-symmetric invariant of this type is stu, which is zero in the forward limit. Consequently, $f(SSSS) = 0$, which is not strictly positive, contradicting Eq. (5.4.4).[5.2] Thus, the pure Galileon theory is marginally excluded by analyticity bounds.

These results are consistent with our own because the Galileon only describes the scalar mode of massive gravity in the limit of Goldstone equivalence. In contrast, our results automatically incorporate all contributions coming from the tensor and vector modes as well. More importantly, our calculation implicitly includes subleading corrections to Goldstone equivalence that scale as higher powers in m^2/s relative to the pure Galileon result. Thus, while the leading behavior of Eq. (5.3.14) scales as stu as expected, there are subleading corrections at order s^2 that are nonzero. Since the pure Galileon is only

[5.2] Note that in certain conformal variations of the Galileon [156], the theory is modified, permitting $(\partial S)^4$ corrections that allow accordance with analyticity constraints.

marginally inconsistent with analyticity bounds, the right choice of (c_3, d_5) can tip the scales. In this sense, our calculation shows explicitly that the pathologies of the Galileon are remedied when embedded in a full theory of massive gravity.

5.6 Conclusions

In this chapter, we have used the principles of unitarity and analyticity of scattering amplitudes to bound the general effective theory of a massive graviton. We have shown that the consistency of massive graviton scattering significantly constrains the parameter space of ghost-free massive gravity. Analyticity bounds have been analyzed in other contexts, both in non-gravitational [31, 138] and more recently gravitational [2, 3, 31] theories. Such analyses provide useful criteria for charting the boundary between the landscape and the swampland. As the principles from which these bounds are derived are infrared properties, they apply to any well-behaved ultraviolet completion obeying the canonical axioms of field theory, irrespective of what the ultimate theory of quantum gravity may be.

Chapter 6
Positivity of Curvature-Squared Corrections in Gravity

We study the Gauss-Bonnet (GB) term as the leading higher-curvature correction to pure Einstein gravity. Assuming a tree-level ultraviolet completion free of ghosts or tachyons, we prove that the GB term has a non-negative coefficient in dimensions greater than four. Our result follows from unitarity of the spectral representation for a general ultraviolet completion of the GB term.

This chapter is from Ref. [5], C. Cheung and G. N. Remmen, "Positivity of Curvature-Squared Corrections in Gravity," Phys. Rev. Lett. ***118*** *(2017) 051601,* arXiv:1608.02942 [hep-th].

6.1 Introduction

Effective field theory lore states that, in constructing a Lagrangian, one should include all operators allowed by symmetry and power counting with arbitrary coefficients. Naively, this implies an immense freedom for low-energy model-building. However, not all quantum effective field theories are created equal: some are compatible with ultraviolet completion, while others reside in the so-called swampland [27, 29, 31], impervious to string-theoretic completion or, worse, any completion conforming to the usual axioms of quantum field theory.

An ongoing effort has been undertaken to demarcate the boundaries of healthy effective field theories, with constraints derived from both top-down and bottom-up reasoning. An iconic example of the former is the weak gravity conjecture [30], which was deduced from string-theoretic examples and black hole thought experiments. In the latter approach, one conceives bounds purely

within the logic of low-energy effective theory, e.g., from considerations of causality, unitarity, and locality/analyticity for long-distance observables such as scattering amplitudes and particle trajectories [2–4, 31, 34–37, 111, 113, 138, 145, 156, 159, 178, 179].

In this chapter, we derive a simple bound on curvature-squared corrections to Einstein gravity. Taking a low-energy perspective, we study gravity as an effective field theory described by the Einstein-Hilbert action,[6.1] $S = \int \mathrm{d}^D x \sqrt{-g}\, R/2\kappa^2$, whose higher-curvature corrections a priori include $R_{\mu\nu\rho\sigma}R^{\mu\nu\rho\sigma}$, $R_{\mu\nu}R^{\mu\nu}$, and R^2. However, the usual invariance under field redefinitions implies that leading corrections in the derivative expansion are defined only up to equations of motion, so those operators involving R and $R_{\mu\nu}$ can be discarded. Hence, the only nontrivial leading correction to pure Einstein gravity is effectively $R_{\mu\nu\rho\sigma}R^{\mu\nu\rho\sigma}$, which up to equations of motion is equivalent to the Gauss-Bonnet (GB) term

$$\Delta S = \int \mathrm{d}^D x \sqrt{-g}\, \lambda \Big(R_{\mu\nu\rho\sigma}R^{\mu\nu\rho\sigma} - 4R_{\mu\nu}R^{\mu\nu} + R^2 \Big). \tag{6.1.1}$$

The GB term is a total derivative in $D = 4$, so we take $D > 4$ throughout this chapter. The GB term is ghost-free [139] and is ubiquitous in string-theoretic completions of gravity.

The coupling constant λ is an important low-energy probe of the ultraviolet completion of general relativity. The sign of λ is also of particular interest from holographic considerations, being related to the viscosity-to-entropy ratio of the dual conformal field theory (see Ref. [180] and references therein). More importantly, $\lambda \geq 0$ appears to be a generic prediction of string theory: $\lambda = 0$ in type II superstring theory [150], while $\lambda > 0$ for the bosonic [139], heterotic [151], and type I [147] string.

[6.1] In this chapter, we use mostly-plus signature for the metric, adopt sign conventions $R_{\mu\nu} = R^{\rho}{}_{\mu\rho\nu}$ and $R^{\mu}{}_{\nu\rho\sigma} = \partial_\rho \Gamma^{\mu}{}_{\nu\sigma} + \cdots$ for the curvature tensor, and define $\kappa = \sqrt{8\pi G}$.

Here, we explore theories in which the GB term is generated by weakly-coupled dynamics below the Planck scale, corresponding to large λ in natural units. Furthermore, we assume that "primordial" contributions to the GB term—i.e., contributions present in the ultraviolet but unaccompanied by new states—are subdominant. This assumption is reasonable because a primordial GB term will violate unitarity below the Planck scale. In addition, Ref. [37] demonstrated how a primordial GB term violates causality unless new states are introduced. Moreover, it can incur potential violations of analyticity [3] and the second law of black hole thermodynamics [157]. All of these issues strongly motivate consideration of a GB term generated dominantly by weakly-coupled ultraviolet dynamics.

Within these assumptions, we will prove that $\lambda \geq 0$ for any unitary tree-level ultraviolet completion of the GB term. To do so, we first enumerate interactions that couple gravitons to massive states in order to generate the GB term at tree level. We then introduce a general spectral representation for the two-point function for these massive degrees of freedom. Finally, we show how unitarity of the spectral representation fixes the sign of the curvature-squared operator coefficient in the gravitational effective theory.

6.2 Coupling to Massive States

In this section, we study the structure of weakly-coupled ultraviolet dynamics that generates curvature-squared corrections to gravity at low energies. As noted earlier, we can freely substitute the tree-level equations of motion—i.e., Einstein's equations—into the leading curvature corrections in Eq. (6.1.1). In practice, this means that the GB term is, at leading order in the derivative expansion, equivalent to the Riemann-squared operator and the Weyl-squared operator,

$$C_{\mu\nu\rho\sigma}C^{\mu\nu\rho\sigma} = R_{\mu\nu\rho\sigma}R^{\mu\nu\rho\sigma} - \frac{4}{D-2}R_{\mu\nu}R^{\mu\nu} + \frac{2}{(D-1)(D-2)}R^2, \quad (6.2.1)$$

where the Weyl tensor is

$$\begin{aligned} C_{\mu\nu\rho\sigma} = R_{\mu\nu\rho\sigma} &- \frac{1}{D-2}\left(g_{\mu[\rho}R_{\sigma]\nu} - g_{\nu[\rho}R_{\sigma]\mu}\right) \\ &+ \frac{1}{(D-1)(D-2)}Rg_{\mu[\rho}g_{\sigma]\nu} \end{aligned} \quad (6.2.2)$$

and square brackets on indices denote antisymmetrization without normalization, i.e., $A_{[\mu\nu]} = A_{\mu\nu} - A_{\nu\mu}$. In the presence of massless matter or gauge fields, this equivalence holds modulo additional interactions involving the stress-energy tensor.

This all implies that the low-energy coefficients of the GB term, the Riemann-squared term, and the Weyl-squared term are equal. For technical simplicity, we therefore recast the action as

$$S = \int \mathrm{d}^D x\sqrt{-g}\left(\frac{R}{2\kappa^2} + \lambda C_{\mu\nu\rho\sigma}C^{\mu\nu\rho\sigma}\right) \quad (6.2.3)$$

using the freedom of equations of motion. Let us note that the above action applies to a low-energy theory comprised purely of massless gravitons. If there are additional spectator massless matter fields or gauge fields, there will be additional terms involving the stress-energy tensor that do not affect our arguments.

Throughout our analysis in this chapter, we assume a weakly-coupled ultraviolet completion of gravity. In turn, this assumption implies that high-energy graviton scattering is unitarized by tree-level exchanges of heavy states. The reason for this is as follows. In any theory that is weakly coupled from the ultraviolet to the infrared, there is, by definition, a well-defined $\hbar$ expansion at all scales. Crucially, in general relativity, diffeomorphism symmetry relates the kinetic term for the graviton to its interactions within the Einstein-Hilbert

term. Since the former is manifestly an $\mathcal{O}(1/\hbar)$ tree-level effect, then so, too, is the latter, which means that it can only be unitarized by tree-level exchanges.

A similar line of reasoning applies to the nonlinear sigma model, which is why unitarization of pion scattering at weak coupling can only be achieved via tree-level Higgs exchange. More generally, while the weak coupling assumption could potentially be relaxed through an accounting of loop corrections as in Ref. [113], such an approach would apply to the derivation of positivity bounds via scattering amplitudes and analytic dispersion relations (e.g., Ref. [3]), as opposed to the unitarity-based methods of the present chapter.

In contrast to the leading-order gravity action, operators like the GB term are separately diffeomorphism invariant and are not directly connected to the Einstein-Hilbert term via symmetry. Hence, even at weak coupling, the GB operator can be ultraviolet-completed at tree or loop level. An analogous statement is true for Euler-Heisenberg higher-dimension operators in gauge theory: since they are not connected directly to the gauge kinetic term, they can arise from tree-level exchange or at loop order.

Nevertheless, since high-energy graviton scattering is unitarized at tree level, it is well motivated to focus on tree-level ultraviolet completions of the GB term. Indeed, this is how the GB term arises in the low-energy gravitational effective actions of string theories. Thus, from here on we assume that Eq. (6.1.1) arises from the exchange of heavy states at tree level.

Next, let us systematically enumerate all possible ultraviolet-completing dynamics for the GB term. Denoting a heavy state by χ, we must identify all diffeomorphism-invariant couplings between χ and gravitons. These interactions could involve one, two, or more powers of χ, which we now consider.

For interactions that are linear in χ, any derivatives on χ can always be shuffled onto the gravitons via integration by parts. Since χ is like a matter

field, it by construction transforms as a tensor and thus necessarily couples to some combination of gravitons that also transforms as a tensor.[6.2] If this tensor of gravitons has no derivatives, then in the flat-space limit χ appears as a tadpole in the Lagrangian, so the corresponding term is eliminated once we expand around the proper vacuum. On the other hand, if this tensor has exactly one derivative, then the resulting operator must be a total derivative since the metric is covariantly constant. Finally, if this tensor has two derivatives, then it has mass dimension two and thus just the right power counting to induce a curvature-squared operator. Indeed, any more derivatives will generate operators of higher order than curvature-squared in the derivative expansion.

The only possible tensors of mass dimension two constructed from the metric are the Riemann tensor and its contractions [181]. Hence, any graviton interactions that are linear in χ must take the form

$$y\, C_{\mu\nu\rho\sigma}\chi^{\mu\nu\rho\sigma}, \tag{6.2.4}$$

where $\chi_{\mu\nu\rho\sigma}$ is a field representing all the massive states that generate the GB term and y is a coupling constant. Analogous operators involving $R_{\mu\nu}$ and R can be discarded by equations of motion.

Without loss of generality, we can take $\chi_{\mu\nu\rho\sigma}$ in Eq. (6.2.4) to possess all of the index properties of the Weyl tensor, namely, the requisite (anti-)symmetries, the first Bianchi identity, and on-shell tracelessness. Any components of $\chi_{\mu\nu\rho\sigma}$ that violate these symmetry properties are automatically projected out by the Weyl tensor in Eq. (6.2.4).

Note that Eq. (6.2.4) induces *mixing* between the graviton and the heavy state. However, since this preserves diffeomorphism invariance, the resulting

[6.2] By tensor, we simply mean an object that transforms covariantly under nonlinear coordinate transformations. Since the metric $g_{\mu\nu}$ is a tensor, it is convenient to parameterize all dependence of the graviton through $g_{\mu\nu}$, its associated curvature tensors, and covariant derivatives ∇_μ.

massless eigenstate should still be interpreted as the massless graviton.

On the other hand, interactions that are quadratic in χ will automatically produce new heavy states in pairs. To generate an effective operator involving only gravitons, we can close the loop of heavy states, but this interaction goes beyond tree level and is thus suppressed at weak coupling. An important exception to this occurs if χ mixes with the graviton, in which case we must introduce Eq. (6.2.4) anyway. Similar arguments apply for interactions with higher powers of χ, but the final result is the same: any weakly-coupled ultraviolet completion of the GB term will involve the operator in Eq. (6.2.4).

6.3 Spectrum of Massive States

Next, we construct a general Källén-Lehmann spectral representation [182, 183] for the heavy states χ following the analysis of Refs. [2, 34, 35]. By expanding the metric $g_{\mu\nu}$ around a flat background $\eta_{\mu\nu}$, we can represent the χ two-point function in D dimensions as

$$\langle \chi_{\mu\nu\rho\sigma}(k)\chi_{\alpha\beta\gamma\delta}(k')\rangle = i\delta^D(k+k')\int_0^\infty \mathrm{d}\mu^2 \frac{\rho(\mu^2)}{-k^2-\mu^2+i\epsilon}\Pi_{\mu\nu\rho\sigma\alpha\beta\gamma\delta}, \quad (6.3.1)$$

where k^2 is contracted with the flat metric. Here, $\Pi_{\mu\nu\rho\sigma\alpha\beta\gamma\delta}$ is the propagator numerator for $\chi_{\mu\nu\rho\sigma}$ and $\rho(\mu^2)$ is the spectral density encoding arbitrary ultraviolet dynamics in terms of a distribution of poles corresponding to each massive state. Since we are working at tree level, $\rho(\mu^2)$ is just a sum over delta functions, so the spectral representation is merely a simple way to package a set of resonances.

The absence of tachyons implies that $\mu^2 \geq 0$. As we will soon see, the propagator numerator $\Pi_{\mu\nu\rho\sigma\alpha\beta\gamma\delta}$ is highly constrained by its symmetries and unitarity. In turn, $\rho(\mu^2) \geq 0$ is required if the theory is to be ghost-free

[182, 183]. The fact that the spectrum is gapped implies regularity of the two-point function as $k \to 0$, so the spectral density should vanish as $\mu^2 \to 0$.

Unitarity requires that the on-shell propagator numerator be a sum over the tensor product of the physical polarizations [20]. That is, when the on-shell condition $k^2 = -\mu^2$ is satisfied, the propagator numerator is

$$\Pi_{\mu\nu\rho\sigma\alpha\beta\gamma\delta} = \sum_i \varepsilon_{i\mu\nu\rho\sigma}\varepsilon^*_{i\alpha\beta\gamma\delta}, \tag{6.3.2}$$

where $\varepsilon_{i\mu\nu\rho\sigma}$ are the physical polarization states of $\chi_{\mu\nu\rho\sigma}$, indexed by i and normalized so that $\varepsilon_{i\mu\nu\rho\sigma}\varepsilon_j^{*\mu\nu\rho\sigma} = \delta_{ij}$. By definition, the polarization tensors transform in representations of the $SO(D-1)$ little group for the massive state $\chi_{\mu\nu\rho\sigma}$. Consequently, the polarizations must reside in the subspace transverse to the momentum of $\chi_{\mu\nu\rho\sigma}$. From Eq. (6.3.2), this implies the transversality condition for on-shell k_μ,

$$k^\mu \Pi_{\mu\nu\rho\sigma\alpha\beta\gamma\delta} = 0 \tag{6.3.3}$$

and similarly for all other contractions.

Note that $\chi_{\mu\nu\rho\sigma}$ is not a canonical spin-four state [116, 117, 135, 136] since it is not fully symmetric. Rather, as we noted in the previous section, $\chi_{\mu\nu\rho\sigma}$ can without loss of generality be taken to have the index properties of the Weyl tensor, which are then inherited by the corresponding polarizations as well as the propagator numerator by Eq. (6.3.2). For example, on-shell tracelessness of $\chi_{\mu\nu\rho\sigma}$ implies that, when the on-shell condition is satisfied, $\Pi_{\mu\nu\rho\sigma\alpha\beta\gamma\delta}$ vanishes when any two indices among the first set of four are contracted and similarly for the second set. Because we do not a priori know the form of the propagator numerator, we must construct it purely from its symmetries and the on-shell transversality and tracelessness conditions.

The most general construction begins by considering $\Pi_{\mu\nu\rho\sigma\alpha\beta\gamma\delta}$ to be an arbitrary eight-index tensor built out of $\eta_{\mu\nu}$ and k_μ. Then, in general D, we

impose the requisite symmetries coming from the index properties of the Weyl tensor and symmetry on exchange of the two copies of $\chi_{\mu\nu\rho\sigma}$: antisymmetry on the first and second pairs of indices, symmetry under the exchange of the first and second index pairs, symmetry under the exchange of the first and second sets of four indices, the first Bianchi identity $\Pi_{\mu[\nu\rho\sigma]\alpha\beta\gamma\delta} = \Pi_{\mu\nu\rho\sigma\alpha[\beta\gamma\delta]} = 0$, on-shell tracelessness on each set of four indices (for arbitrary metric contraction of two indices), and on-shell transversality per Eq. (6.3.3). We discover that these conditions are enough to fix the propagator numerator $\Pi_{\mu\nu\rho\sigma\alpha\beta\gamma\delta}$ up to some as-yet-unspecified coefficient β:

$$\begin{aligned}\Pi^{\mu\nu\rho\sigma}{}_{\alpha\beta\gamma\delta} = \beta\Big[&2(D-2)(D-3)\Pi^{\mu}{}_{[\alpha}\Pi^{\nu}{}_{\beta]}\Pi^{\rho}{}_{[\gamma}\Pi^{\sigma}{}_{\delta]} \\ &+ 2(D-2)(D-3)\Pi^{\mu}{}_{[\gamma}\Pi^{\nu}{}_{\delta]}\Pi^{\rho}{}_{[\alpha}\Pi^{\sigma}{}_{\beta]} \\ &+ (D-2)(D-3)\Pi^{[\mu}{}_{\delta}\Pi^{\nu]}{}_{[\alpha}\Pi^{[\rho}{}_{\beta]}\Pi^{\sigma]}{}_{\gamma} \\ &- (D-2)(D-3)\Pi^{[\mu}{}_{\gamma}\Pi^{\nu]}{}_{[\alpha}\Pi^{[\rho}{}_{\beta]}\Pi^{\sigma]}{}_{\delta} \\ &- 3(D-2)\Pi^{[\mu}{}_{[\alpha}\Pi^{\nu][\rho}\Pi_{\beta][\gamma}\Pi^{\sigma]}{}_{\delta]} \\ &- 3(D-2)\Pi^{[\mu}{}_{[\gamma}\Pi^{\nu][\rho}\Pi_{\delta][\alpha}\Pi^{\sigma]}{}_{\beta]} \\ &+ 12\Pi^{\mu[\rho}\Pi^{\sigma]\nu}\Pi_{\alpha[\gamma}\Pi_{\delta]\beta}\Big],\end{aligned} \tag{6.3.4}$$

where we found that the result could be written in terms of the Proca propagator numerator

$$\Pi_{\mu\nu} = \eta_{\mu\nu} + \frac{k_\mu k_\nu}{\mu^2}. \tag{6.3.5}$$

The appearance of this dependence on the projection operator $\Pi_{\mu\nu}$ is not surprising given the transversality condition (6.3.3). However, we emphasize that we did not assume beforehand that $\Pi_{\mu\nu\rho\sigma\alpha\beta\gamma\delta}$ could be expressed as a function of the Proca propagator numerator.

Now, by the completeness relation (6.3.2), the full trace of the propagator numerator counts the number of physical degrees of freedom, so we must have $\Pi_{\mu\nu\rho\sigma}{}^{\mu\nu\rho\sigma} > 0$. Specifically, the number of independent physical degrees of freedom in $\chi_{\mu\nu\rho\sigma}$ is just the number of possible polarizations. This is the

number of tensors $\varepsilon_{i\mu\nu\rho\sigma}$ with the symmetries of the Weyl tensor that respect the transversality condition. Working through the combinatorics is straightforward and one finds that the number of physical degrees of freedom is

$$N = \frac{1}{12}(D+1)D(D-1)(D-4). \tag{6.3.6}$$

On the other hand, from Eq. (6.3.4), we find the beautiful expression

$$\Pi_{\mu\nu\rho\sigma}{}^{\mu\nu\rho\sigma} = 2\beta(D+1)D(D-1)(D-2)(D-3)(D-4), \tag{6.3.7}$$

which for $D > 4$ is positive if and only if $\beta > 0$. Requiring that $\Pi_{\mu\nu\rho\sigma}{}^{\mu\nu\rho\sigma} = N$, we have

$$\beta = \frac{1}{24(D-2)(D-3)}. \tag{6.3.8}$$

Equivalently, we recall that a propagator numerator, when taken on shell, is a projector onto the space orthogonal to k_μ [16] and onto tensors with the requisite index symmetries. Requiring that the propagator numerator be idempotent as a projection operator thus fixes the normalization.

6.4 Integrating Out Massive States

We can now compute the higher-curvature corrections induced by integrating out χ. As noted earlier, interactions between gravitons and two or more powers of χ can contribute to higher-curvature corrections given the mixing term in Eq. (6.2.4). Thus, to study graviton scattering at low energies, it would be necessary to do a proper accounting of all the interactions involving χ beyond even Eq. (6.2.4). As this is rather cumbersome, it is more convenient to compute the off-shell two-point function for the graviton. This low-energy operator receives contributions from Eq. (6.2.4), but crucially is independent of the interactions nonlinear in χ.

Armed with a general parameterization of the couplings and spectrum of the massive states, we can now integrate them out. Using Eqs. (6.3.4) and (6.3.8), one finds

$$C^{\mu\nu\rho\sigma}\Pi_{\mu\nu\rho\sigma\alpha\beta\gamma\delta}C^{\alpha\beta\gamma\delta} \overset{k\to 0}{=} C_{\mu\nu\rho\sigma}C^{\mu\nu\rho\sigma}. \tag{6.4.1}$$

Since we are computing the two-point function for gravitons, we are implicitly expanding $C_{\mu\nu\rho\sigma}$ at linear order in gravitons. Integrating out $\chi_{\mu\nu\rho\sigma}$ at low momentum transfer, we obtain the effective operator

$$\frac{y^2}{2}C_{\mu\nu\rho\sigma}C^{\mu\nu\rho\sigma}\int_0^\infty \frac{\mathrm{d}\mu^2}{\mu^2}\rho(\mu^2). \tag{6.4.2}$$

We then deduce the coefficient of the Weyl-squared operator in Eq. (6.2.3),

$$\lambda = \frac{y^2}{2}\int_0^\infty \frac{\mathrm{d}\mu^2}{\mu^2}\rho(\mu^2) \geq 0. \tag{6.4.3}$$

Thus, since the spectral function is non-negative by unitarity, the sign of the coefficient λ of the GB operator is non-negative in a consistent tree-level ultraviolet completion in $D > 4$.

This bound is consistent with results from string theory [139, 147, 150, 151]. Moreover, our bound constitutes a requisite consistency condition for any candidate tree-level theory of quantum gravity. Proving positivity of the GB coefficient using a different approach—analytic dispersion relations—is the subject of current ongoing research [184], though subtleties exist in applying analyticity bounds to graviton amplitudes [3, 31]. While standard axioms of quantum field theory, e.g., locality, may be violated in quantum gravity, dispersion relations themselves seem to remain robust [97].

Delineating the boundary between the swampland and the landscape can provide insights for model-building and for our broader understanding of gravitational ultraviolet completion of quantum field theories. Open problems include finding ways to apply infrared consistency bounds in nonperturbative contexts, as well as connecting bounds obtained from infrared- and ultraviolet-dependent reasoning.

Chapter 7
Twofold Symmetries of the Pure Gravity Action

We recast the action of pure gravity into a form that is invariant under a twofold Lorentz symmetry. To derive this representation, we construct a general parameterization of all theories equivalent to the Einstein-Hilbert action up to a local field redefinition and gauge fixing. We then exploit this freedom to eliminate all interactions except those exhibiting two sets of independently contracted Lorentz indices. The resulting action is local, remarkably simple, and naturally expressed in a field basis analogous to the exponential parameterization of the nonlinear sigma model. The space of twofold Lorentz invariant field redefinitions then generates an infinite class of equivalent representations. By construction, all off-shell Feynman diagrams are twofold Lorentz invariant while all on-shell tree amplitudes are automatically twofold gauge invariant. We extend our results to curved spacetime and calculate the analogue of the Einstein equations. While these twofold invariances are hidden in the canonical approach of graviton perturbation theory, they are naturally expected given the double copy relations for scattering amplitudes in gauge theory and gravity.

This chapter is from Ref. [6], C. Cheung and G. N. Remmen, "Twofold Symmetries of the Pure Gravity Action," JHEP ***01*** *(2017) 104,* arXiv:1612.03927 [hep-th].

7.1 Introduction

The scattering amplitudes program has revealed extraordinary structures underlying long-studied quantum field theories. One such class of miracles reformulates gravity as the "square" of gauge theory. Discovered by KLT [45] and generalized by BCJ [46], this relationship is encoded in concrete formulae expressing the scattering amplitudes of pure gravity as sums over products of the scattering amplitudes of Yang-Mills theory,

$$A_{\rm GR} \sim \sum A_{\rm YM} \bar{A}_{\rm YM}, \tag{7.1.1}$$

where the barred and unbarred factors need not be the same amplitude. This duality appears in various guises in a variety of contexts, both in field theory and string theory (see Ref. [185] and references therein). Remarkably, the double copy structure also persists in classical field theory, where certain gauge theory backgrounds map directly to solutions of general relativity [186–191].

Pragmatically, these squaring relations simplify certain gravity calculations by connecting them directly to known computations in gauge theory [192]. From a top-down perspective, however, this correspondence suggests a very surprising fact about the underlying symmetries of gravity. In particular, since the right-hand side of Eq. (7.1.1) is a sum over products of Lorentz scalars, it is separately invariant under Lorentz transformations acting individually on each Yang-Mills factor. To see this explicitly, consider graviton polarizations expressed as a bivector,

$$\epsilon_{a\bar{b}} = \epsilon_a \bar{\epsilon}_{\bar{b}}, \tag{7.1.2}$$

where $A_{\rm YM}$ and $\bar{A}_{\rm YM}$ depend only on ϵ and $\bar{\epsilon}$, respectively. Denoting the momenta contracted with unbarred and barred indices by k and $\bar{k}$, respectively, it follows that the right-hand side of Eq. (7.1.1) exhibits a formal twofold invariance under a pair of Lorentz transformations,

$$k_a \to \Lambda_a{}^b k_b \qquad \text{and} \qquad \bar{k}_{\bar{a}} \to \bar{\Lambda}_{\bar{a}}{}^{\bar{b}} \bar{k}_{\bar{b}}, \tag{7.1.3}$$

together with a pair of Ward identity transformations,

$$\epsilon_a \to \epsilon_a + k_a \qquad \text{and} \qquad \bar{\epsilon}_{\bar{a}} \to \bar{\epsilon}_{\bar{a}} + \bar{k}_{\bar{a}}. \tag{7.1.4}$$

The fact that the physical scattering amplitudes of pure gravity can be expressed as products of Yang-Mills amplitudes hints at an underlying "twofold Lorentz symmetry" of pure gravity,

$$SO(D-1,1) \times \overline{SO}(D-1,1). \tag{7.1.5}$$

It should be possible to manifest such a property at the level of the action. Such a formulation would manifest "index factorization," i.e., where all interactions of the graviton field $h_{a\bar{b}}$ involve indices contracted with η_{ab} and $\eta_{\bar{a}\bar{b}}$, thus forbidding contractions between barred and unbarred indices. This condition places stringent restrictions on the allowed interaction terms. For example, something as innocuous as the trace of the graviton, $h_a{}^a = h_{a\bar{b}}\eta^{\bar{b}a}$, is not twofold Lorentz invariant since $\epsilon_a \bar{\epsilon}^a = \epsilon_a \bar{\epsilon}_{\bar{b}} \eta^{\bar{b}a}$ contracts barred and unbarred indices.

The canonical procedure for graviton perturbation theory grossly violates index factorization and, in turn, twofold Lorentz symmetry. In particular, the Einstein-Hilbert (EH) action in D spacetime dimensions is[7.1]

$$S = \int \mathrm{d}^D x \sqrt{-g} \left(\frac{R}{16\pi G} + \mathcal{L}_{\mathrm{GF}} \right), \tag{7.1.6}$$

where $\mathcal{L}_{\mathrm{GF}}$ denotes the Faddeev-Popov gauge-fixing term. To compute graviton scattering amplitudes in perturbation theory, we typically define

$$g_{ab} = \eta_{ab} + h_{ab} \tag{7.1.7}$$

and expand the action in powers of the graviton perturbation h_{ab}. Terms

[7.1] In this chapter, we work in mostly-plus signature and use the conventions $R_{ab} = R^c{}_{acb}$ and $R^a{}_{bcd} = \partial_c \Gamma^a_{bd} - \partial_d \Gamma^a_{bc} + \Gamma^a_{ce}\Gamma^e_{bd} - \Gamma^a_{de}\Gamma^e_{bc}$. We denote the determinant of a metric as the metric's label with no indices, e.g., $g = \det g_{ab}$, etc. For notational reasons, we will adopt Latin indices throughout this chapter.

involving the trace of the graviton, together with other nonfactorized structures, induce propagators and interaction vertices that explicitly violate the twofold Lorentz symmetry.

Nevertheless, in the seminal work of Ref. [193], Bern and Grant showed how the KLT relations can be reverse-engineered to perturbatively construct an action for pure gravity compatible with manifest index factorization. They achieved this feat up to fifth order in graviton perturbations, leaving open the question of an all-orders generalization. Furthermore, to derive this action from the original EH action required the introduction of a dilaton, which when integrated out induced non-local graviton interactions.

In this chapter, we recast the EH action for pure gravity into a form that is local and manifestly twofold Lorentz invariant at all orders in graviton perturbations. To do so, we exploit the fact that the usual EH action of conventional graviton perturbation theory is not particularly meaningful: the freedom of nonlinear field redefinitions and gauge fixing permits one to rewrite the action in an infinite number of different ways, all describing equivalent physics. By exploring this full freedom, we derive a local representation of the EH action that is compatible with index factorization at all nonlinear orders and requires no additional dynamical or auxiliary fields beyond the graviton. The off-shell Feynman propagators and vertices are trivially twofold Lorentz invariant and the resulting tree-level on-shell scattering amplitudes are twofold gauge invariant. The resulting action is derived most naturally in an "exponential basis" for the graviton, reminiscent of the common parameterization of Nambu-Goldstone bosons in the nonlinear sigma model.

By recasting this action in terms of fields on a doubled spacetime of dimension $2D$, we automate the bookkeeping of the barred and unbarred indices at the expense of introducing a two-form field, which decouples from

all tree-level graviton scattering amplitudes. We comment on the link between these representations and those that arise from double field theory [194–200], where Einstein gravity coupled to a dilaton and two-form arises as the low-energy effective field theory of string theory at leading order in the derivative expansion.

An obvious corollary is that our action also generates, via further field redefinitions, an infinite class of equivalent twofold Lorentz invariant actions. Again utilizing this freedom of graviton field basis, we study alternative versions of this action, going from the exponential basis to the analogue of the "Cayley basis" [201] for the nonlinear sigma model. Here, we find that graviton perturbation theory simplifies substantially and manifests some unexpected additional symmetries.

The remainder of this chapter is organized as follows. In Sec. 7.2, we discuss a systematic procedure parameterizing the space of local field redefinitions and gauge-fixing conditions in pure gravity. Afterwards, we show in Sec. 7.3 how this exercise yields a simple action that exhibits index factorization and thus twofold Lorentz invariance. This form is naturally written in terms of a spacetime of doubled dimension. We then discuss the graviton propagator, as well as a more general class of twofold Lorentz invariant theories related by field redefinitions. Next, we generalize this formalism to curved spacetime in Sec. 7.4, establishing index factorization for any Ricci-flat spacetime and deriving the corresponding Einstein equations. We conclude and discuss future directions in Sec. 7.5.

7.2 Building the Action

In this section, we define the space of local actions equivalent to the EH action modulo field redefinitions and gauge fixing. For a particular choice,

the EH action can be recast into a form that manifests index factorization and is thus compatible with twofold Lorentz invariance. Here, we will study graviton perturbation theory as an expansion about flat spacetime in Cartesian coordinates,

$$\eta_{ab} = \mathrm{diag}(-1, 1, \ldots, 1). \tag{7.2.1}$$

In Sec. 7.4, we will generalize our results to arbitrary backgrounds and curvilinear coordinate systems.

7.2.1 Index Factorization

To begin, we identify which terms are compatible and incompatible with index factorization. For later convenience, we define powers of the graviton tensor by

$$\begin{aligned} h^n_{ab} &= h_{ab_1}\eta^{b_1a_1}h_{a_1b_2}\eta^{b_2a_2}\cdots h_{a_{n-2}b_{n-1}}\eta^{b_{n-1}a_{n-1}}h_{a_{n-1}b} \\ &= h_a{}^{a_1}h_{a_1}{}^{a_2}\cdots h_{a_{n-2}}{}^{a_{n-1}}h_{a_{n-1}b}, \end{aligned} \tag{7.2.2}$$

together with a shorthand for the trace,

$$[h^n] = h^n_{ab}\eta^{ba}. \tag{7.2.3}$$

We can now determine when these products of the graviton tensor are compatible with index factorization. Many operators are comprised of gravitons built from objects of the form

$$[h^{2n}] = \text{even cycle} \qquad \text{or} \qquad [h^{2n+1}] = \text{odd cycle}, \tag{7.2.4}$$

where we have suppressed all derivatives and their contractions. The odd cycles necessarily violate index factorization. This is obvious because an odd number of graviton tensors appear with an odd number of barred indices and an odd number of unbarred indices. Thus, contracting all the indices will necessarily involve the contraction of at least one barred and one unbarred index. In contrast, the even cycles are compatible with index factorization, since there exists a consistent assignment of barred and unbarred indices.

As noted before, however, odd cycles appear ubiquitously in the conventional approach to graviton perturbation theory, which is derived by expanding the EH action in the field basis in Eq. (7.1.7). For example, the volume element is given by

$$\sqrt{-g} = \exp\left(\frac{1}{2}\sum_{n=1}^{\infty}\frac{(-1)^{n-1}}{n}[h^n]\right), \tag{7.2.5}$$

which has an infinite number of odd cycles that are incompatible with index factorization. Hence, to construct a representation with manifest index factorization it is necessary to go beyond the standard prescription. To do so, we rewrite the EH action in an arbitrary local graviton field basis and gauge-fixing, which we now discuss.

7.2.2 Field Basis and Gauge Fixing

To construct an arbitrary field basis, we consider all possible local field redefinitions of the graviton defined in Eq. (7.1.7). Due to a theorem of Haag [202] (see also Ref. [203] and references therein), field redefinitions leave all scattering amplitudes invariant, provided the asymptotic states remain unaltered. For example, the local field redefinition of a scalar,

$$\phi \to \alpha_1\phi + \alpha_2\phi^2 + \alpha_3\phi^3 + \cdots, \tag{7.2.6}$$

leaves scattering amplitudes unchanged provided $\alpha_1 = 1$ so that the linearized field is the same. For the graviton, the analogous field redefinition is

$$\begin{aligned} h_{ab} \;\to\;& \alpha_1 h_{ab} + \alpha_2\eta_{ab}[h] \\ &+\alpha_3 h_{ab}^2 + \alpha_4 h_{ab}[h] + \alpha_5\eta_{ab}[h^2] + \alpha_6\eta_{ab}[h]^2 \\ &+\alpha_7 h_{ab}^3 + \alpha_8 h_{ab}^2[h] + \alpha_9 h_{ab}[h^2] + \alpha_{10} h_{ab}[h]^2 \\ &\quad + \alpha_{11}\eta_{ab}[h^3] + \alpha_{12}\eta_{ab}[h^2][h] + \alpha_{13}\eta_{ab}[h]^3 \\ &+\cdots, \end{aligned} \tag{7.2.7}$$

where $\alpha_1 = 1$. Here we will restrict to field redefinitions without any derivatives in order to maintain the familiar two-derivative form of the graviton interactions.

In general, it is straightforward but tedious to enumerate the various tensor structures at higher orders in the graviton. At $\mathcal{O}(h^n)$, there are $\sum_{j=0}^{n} p(j)$ possible terms in the nonlinear field redefinition, where $p(j)$ is the number of partitions of j.

Next, we consider gauge fixing, which also comes with an immense freedom. Using the Faddeev-Popov gauge-fixing procedure, we define

$$\mathcal{L}_{\rm GF} = -\eta^{ab} F_a F_b, \tag{7.2.8}$$

for a local but otherwise arbitrary gauge-fixing vector,

$$\begin{aligned} F_a = \partial^b h^{cd} (&\beta_1\ \eta_{ab}\eta_{cd} + \beta_2\ \eta_{ac}\eta_{bd} + \\ &\beta_3\ h_{ab}\eta_{cd} + \beta_4\ h_{ac}\eta_{bd} + \beta_5\ \eta_{ab}h_{cd} + \beta_6\ \eta_{ac}h_{bd} \\ &+ \beta_7\ \eta_{ab}\eta_{cd}[h] + \beta_8\ \eta_{ac}\eta_{bd}[h] + \cdots), \end{aligned} \tag{7.2.9}$$

which can be thought of as a highly nonlinear generalization of harmonic gauge. At $\mathcal{O}(h^n)$ in the nonlinear gauge-fixing vector, there are $2\sum_{j=1}^{n} j\ \ p(n-j)$ possible terms.

As noted earlier, the α and β parameters that appear in the field basis and gauge-fixing have absolutely no effect on physical scattering amplitudes. However, as a check of our calculation, we have also explicitly computed the three-particle and four-particle scattering amplitudes and verified that they are indeed independent of α and β.

7.3 Factorizing the Action

The α and β parameters of the field basis and gauge-fixing alter the action but have no effect on physical observables. Next, we can examine the action at each order in graviton perturbations, fixing the α and β parameters so as to precisely eliminate all appearances of odd cycles, as defined in Eq. (7.2.4). This is a necessary condition for manifest index factorization. By explicit computation,

we have verified that this criterion can be satisfied at least up to fifth order in the graviton. Perhaps surprisingly, we have also found that a special choice of the α and β parameters follows a simple pattern that straightforwardly generalizes to all orders in perturbation theory, taking a simple analytic form. One can then prove that this choice of nonlinear field redefinition and gauge-fixing allows for index factorization of the action at *all orders* in the graviton. It is to this special class of field redefinitions and gauge-fixing that we now turn.

7.3.1 Definition of the Action

We focus on a special field basis for the graviton defined by

$$g_{ab} = \eta_{ab} + \pi_{ab} + \frac{1}{2!}\pi^2_{ab} + \frac{1}{3!}\pi^3_{ab} + \cdots, \text{ where } \pi_{ab} = h_{ab} - \frac{1}{D-2}\eta_{ab}[h]. \quad (7.3.1)$$

It will often be convenient to invoke the shorthand notation

$$g_{ab} = (e^{\pi})_{ab} \qquad \text{and} \qquad g^{ab} = (e^{-\pi})^{ab}, \quad (7.3.2)$$

where by construction $g_{ab}g^{bc} = \delta^c_a$. We emphasize here that g_{ab} and g^{ab} are matrix inverses, not related by raising and lowering with respect to η_{ab}. The utility of an exponential basis for gravity, in that it treats the metric and its inverse symmetrically in the perturbation expansion, was understood previously in Ref. [204]. Our Faddeev-Popov gauge-fixing term is

$$\mathcal{L}_{\rm GF} = -\frac{1}{64\pi G(D-2)} e^{[h]/(D-2)} (e^{-h})^{ab} \partial_a[h]\partial_b[h]. \quad (7.3.3)$$

Using Eqs. (7.3.1) and (7.3.2), we see that we can write the gauge-fixing term in the compact form

$$\mathcal{L}_{\rm GF} = -\frac{D-2}{64\pi G} g^{ab}\omega_a\omega_b, \quad (7.3.4)$$

where we have defined the vector

$$\omega_a = \partial_a \log\sqrt{-g} = -\frac{1}{D-2}\partial_a[h]. \quad (7.3.5)$$

We will postpone further discussion of the physical meaning of this gauge condition to Sec. 7.4. For now, let us simply view $\mathcal{L}_{\rm GF}$ in Eq. (7.3.4) as a particular choice of the coefficients in the general gauge-fixing vector in Eq. (7.2.9). However, note that the above gauge-fixing term does not eliminate the full gauge freedom: the propagator is not yet invertible, although we will see in Sec. 7.3.3 how this is remedied by an additional gauge fixing.

Putting everything together, we find that EH action in Eq. (7.1.6) is drastically simplified, in part because derivatives act nicely on the exponential form of Eq. (7.3.1). The resulting action is independent of the spacetime dimension D and can be written compactly as

$$S = \frac{1}{16\pi G}\int \mathrm{d}^D x\, \partial_a \sigma_{ce} \partial_b \sigma^{de} \left(\frac{1}{4}\sigma^{ab}\delta^c_d - \frac{1}{2}\sigma^{cb}\delta^a_d\right), \tag{7.3.6}$$

expressed in terms of a new exponential field,

$$\sigma_{ab} = \eta_{ab} + h_{ab} + \frac{1}{2!}h^2_{ab} + \frac{1}{3!}h^3_{ab} + \cdots, \tag{7.3.7}$$

which we will often express in the shorthand

$$\sigma_{ab} = (e^h)_{ab} \qquad \text{and} \qquad \sigma^{ab} = (e^{-h})^{ab}, \tag{7.3.8}$$

where $\sigma_{ab}\sigma^{bc} = \delta^c_a$. Note that to obtain Eq. (7.3.6) we applied the useful identity $\sigma^{ab}\partial_c \sigma_{ab} = \partial_c[h]$, valid in Cartesian coordinates so the metric has unit determinant.

Eq. (7.3.6) is a primary result of this chapter, so let us pause to discuss some salient points. First, since we derived this action directly from the EH action, it is a completely equivalent description of pure gravity expanded around flat spacetime. Consequently, the scattering amplitudes computed with this action are exactly equal to those obtained in conventional graviton perturbation theory.

Second, Eq. (7.3.6) is constructed so that every interaction is compatible with index factorization. Consequently, it is always possible to assign distinct

sets of barred and unbarred indices that are separately contracted. For example, our field basis is chosen to precisely eliminate the $\sqrt{-g} = e^{-[h]/(D-2)}$ factor, which was a persistent source of odd cycles in the action. This factor is precisely canceled by the factors of $[h]$ in the definition of π_{ab} in Eq. (7.3.1). Formally, two sets of independently contracted indices exhibit an enhanced twofold Lorentz symmetry. However, these are not, at least in this particular form, symmetries in the literal sense because they act as rigid transformations on the barred and unbarred indices, as for, e.g., an internal symmetry. In terms of the scattering amplitudes relations in Eq. (7.1.1), this restriction of the enhanced symmetry comes from the fact that the two Yang-Mills amplitudes are separately Lorentz invariant, but crucially must have the same external momenta. As we will soon see, by introducing auxiliary extra dimensions one can promote this property of index factorization into a bona fide symmetry of the action.

Third, it is remarkable how the exponential field defined in Eq. (7.3.7) arises naturally from our prescription for eliminating odd cycles. This object is curiously reminiscent of the exponential parameterization of the nonlinear sigma model. It is tempting to imagine that this form of the EH action implies some form of underlying spontaneous symmetry breaking within gravity. However, as we will see later, there are many alternative field bases that are not exponential.

Fourth, Eq. (7.3.6) is extremely simple compared to the standard action for graviton perturbations, which is derived by inserting the field basis of Eq. (7.1.7) into Eq. (7.1.6). Expanding Eq. (7.3.6) in perturbations, we find that

$$S = \frac{1}{16\pi G} \int \mathrm{d}^D x \sum_n \mathcal{O}_n, \tag{7.3.9}$$

where the first few orders of the operators $\mathcal{O}_n$ are

$$
\begin{aligned}
\mathcal{O}_2 =& + \frac{1}{2}\partial_c h_{ab}\partial^b h^{ac} - \frac{1}{4}\partial_c h_{ab}\partial^c h^{ab} \\
\mathcal{O}_3 =& + \frac{1}{4}h^{ab}\partial_a h_{cd}\partial_b h^{cd} - \frac{1}{2}h^{ab}\partial_c h_{ad}\partial_b h^{cd} \\
\mathcal{O}_4 =& + \frac{1}{8}h_{ab}h^{cd}\partial^b h_{ce}\partial_d h^{ae} - \frac{1}{8}h^{ab}h_{ac}\partial_b h_{de}\partial^c h^{de} \\
& - \frac{1}{12}h^{ab}h^{cd}\partial_c h_{be}\partial^e h_{ad} + \frac{1}{24}h^{ab}h^{cd}\partial^e h_{cb}\partial_e h_{ad} \\
& + \frac{1}{6}h^{ab}h_{ac}\partial^c h^{de}\partial_e h_{db} + \frac{1}{24}h^{ab}h_{ac}\partial^d h^{ec}\partial_e h_{db} \\
& - \frac{1}{24}h^{ab}h_{ac}\partial^e h^{dc}\partial_e h_{db} \\
\mathcal{O}_5 =& - \frac{1}{12}h^{ab}h_{ac}h_{de}\partial^c h^{fe}\partial^d h_{fb} + \frac{1}{24}h^{ab}h_{ac}h_{db}\partial^c h^{ef}\partial^d h_{ef} \\
& + \frac{1}{24}h^{ab}h^{cd}h^{ef}\partial_c h_{eb}\partial_f h_{ad} + \frac{1}{24}h^{ab}h_{ac}h^{de}\partial_d h_{fb}\partial_e h^{fc} \\
& - \frac{1}{24}h^{ab}h^{cd}h^{ef}\partial_e h_{ad}\partial_f h_{cb} + \frac{1}{24}h^{ab}h_{ac}h^{de}\partial^c h_{fe}\partial^f h_{db} \\
& - \frac{1}{24}h^{ab}h_{ac}h^{de}\partial_e h^{fc}\partial_f h_{db} - \frac{1}{24}h^{ab}h_{ac}h_{db}\partial^d h_{ef}\partial^f h^{ec}.
\end{aligned}
\tag{7.3.10}
$$

It is straightforward to check that in all of these interactions it is always possible to assign independent sets of barred and unbarred indices that never contract with one another.

While Eq. (7.3.6) is compatible with index factorization, it is certainly not ideal that checking this requires running through each interaction term one at a time and intelligently assigning barred and unbarred indices. Indeed, the situation would be substantively improved with a formalism that does not require a case-by-case analysis of each term, instead treating indices as barred and unbarred from the very beginning. We construct just such a representation in the next subsection.

7.3.2 Adding Auxiliary Dimensions

To automate the proper contraction of barred and unbarred indices, we introduce an additional set of auxiliary dimensions. In particular, let us extend the

D dimensions of spacetime into $2D$ dimensions, where

$$x^A = (x^a, \bar{x}^{\bar{a}}) \qquad \text{and} \qquad \partial_A = (\partial_a, \partial_{\bar{a}}) \tag{7.3.11}$$

and the original D-dimensional spacetime corresponds to the restriction to the "diagonal" spacetime

$$x^a = \bar{x}^{\bar{a}}. \tag{7.3.12}$$

Here, indices in $2D$-dimensional spacetime are contracted with the metric tensors

$$\eta_{AB} = \begin{bmatrix} \eta_{ab} & 0 \\ 0 & \eta_{\bar{a}\bar{b}} \end{bmatrix} \quad \text{and} \quad \eta^{AB} = \begin{bmatrix} \eta^{ab} & 0 \\ 0 & \eta^{\bar{a}\bar{b}} \end{bmatrix}, \tag{7.3.13}$$

so all terms are automatically twofold Lorentz invariant with respect to barred and unbarred indices.

Next, we repackage the graviton into a tensor in $2D$-dimensional spacetime,

$$H_{AB} = \begin{bmatrix} 0 & h_{a\bar{b}} \\ h_{\bar{a}b} & 0 \end{bmatrix}, \tag{7.3.14}$$

where the two off-diagonal blocks are transposes of each other. The structure of this representation explicitly breaks the underlying $SO(2D-2,2)$ symmetry of the doubled $2D$-dimensional spacetime down to the symmetry in Eq. (7.1.5). Since barred and unbarred indices are distinct, $h_{a\bar{b}}$ is automatically lifted to a general D-dimensional matrix. The usual physical graviton modes correspond to the symmetric components of this tensor. As we will see shortly, the antisymmetric component can be neglected at tree level for graviton scattering amplitudes. In terms of this new field, we define the exponential field

$$\Sigma_{AB} = (e^H)_{AB} \quad \text{and} \quad \Sigma^{AB} = (e^{-H})^{AB}. \tag{7.3.15}$$

A simple computation shows that

$$\Sigma_{AB} = \begin{bmatrix} (\cosh h)_{ab} & (\sinh h)_{a\bar{b}} \\ (\sinh h)_{\bar{a}b} & (\cosh h)_{\bar{a}\bar{b}} \end{bmatrix}, \tag{7.3.16}$$

where, in our shorthand,

$$
\begin{aligned}
(\cosh h)_{ab} &= \eta_{ab} + \frac{1}{2!}h^2_{ab} + \frac{1}{4!}h^4_{ab} + \cdots\,, \\
(\sinh h)_{a\bar{b}} &= h_{a\bar{b}} + \frac{1}{3!}h^3_{a\bar{b}} + \frac{1}{5!}h^5_{a\bar{b}} + \cdots
\end{aligned}
\tag{7.3.17}
$$

are even and odd functions in the graviton, respectively. Because these terms have distinct parity, we can, in analogy with Eq. (7.2.2), define

$$
\begin{aligned}
h^{2n}_{ab} &= h_{a\bar{b}_1}\eta^{\bar{b}_1\bar{a}_1}h_{\bar{a}_1 b_2}\eta^{b_2 a_2}\cdots h_{a_{2n-2}\bar{b}_{2n-1}}\eta^{\bar{b}_{2n-1}\bar{a}_{2n-1}}h_{\bar{a}_{2n-1}b} \\
&= h_a{}^{\bar{a}_1}h_{\bar{a}_1}{}^{a_2}\cdots h_{a_{2n-2}}{}^{\bar{a}_{2n-1}}h_{\bar{a}_{2n-1}b}
\end{aligned}
\tag{7.3.18}
$$

for even powers of the graviton, while for odd powers of the graviton,

$$
\begin{aligned}
h^{2n+1}_{a\bar{b}} &= h_{a\bar{b}_1}\eta^{\bar{b}_1\bar{a}_1}h_{\bar{a}_1 b_2}\eta^{b_2 a_2}\cdots h_{\bar{a}_{2n-1}b_{2n}}\eta^{b_{2n}a_{2n}}h_{a_{2n}\bar{b}} \\
&= h_a{}^{\bar{a}_1}h_{\bar{a}_1}{}^{a_2}\cdots h_{\bar{a}_{2n-1}}{}^{a_{2n}}h_{a_{2n}\bar{b}}
\end{aligned}
\tag{7.3.19}
$$

and similarly for the other tensors. By construction, we see that the barred and unbarred indices are never contracted with each other.

In terms of these new variables, the action takes the form

$$
\begin{aligned}
S = \frac{1}{16\pi G}\int \mathrm{d}^D x\,\mathrm{d}^D\bar{x}\,\delta^D(x-\bar{x})\times \\
\times\,\partial_A\Sigma_{CE}\partial_B\Sigma^{DE}\left(\frac{1}{16}\Sigma^{AB}\delta^C_D - \frac{1}{4}\Sigma^{CB}\delta^A_D\right),
\end{aligned}
\tag{7.3.20}
$$

where the numerical factors are slightly different from those in Eq. (7.3.6) due to additional factors of two coming from the trace over the $2D$-dimensional spacetime. Notably, Eq. (7.3.20) has several properties not manifest in the usual representation of the EH action, which we now discuss.

First and foremost, the action is manifestly invariant under a twofold Lorentz symmetry that acts separately on x and $\bar{x}$. Due to the δ function in Eq. (7.3.20), i.e., the fact that the action is only integrated over the diagonal combination $x = \bar{x}$, the two corresponding conserved currents are one and the same. In particular, they produce the usual single conservation of energy, momentum, and angular momentum in D-dimensional spacetime. To see the index factorization explicitly, we can again expand the action in perturbations

to obtain

$$
\begin{aligned}
\mathcal{O}_2 =& + \frac{1}{4}\partial_c h_{a\bar{b}}\partial^a h^{c\bar{b}} + \frac{1}{4}\partial_{\bar{c}} h_{a\bar{b}}\partial^{\bar{b}} h^{a\bar{c}} - \frac{1}{8}\partial_c h_{a\bar{b}}\partial^c h^{a\bar{b}} - \frac{1}{8}\partial_{\bar{c}} h_{a\bar{b}}\partial^{\bar{c}} h^{a\bar{b}} \\
\mathcal{O}_3 =& + \frac{1}{4}h^{a\bar{b}}\partial_a h_{c\bar{d}}\partial_{\bar{b}} h^{c\bar{d}} - \frac{1}{4}h^{a\bar{b}}\partial_d h_{a\bar{c}}\partial_{\bar{b}} h^{d\bar{c}} - \frac{1}{4}h^{a\bar{b}}\partial_{\bar{d}} h_{c\bar{b}}\partial_a h^{c\bar{d}} \\
\mathcal{O}_4 =& - \frac{1}{16}h^{a\bar{b}}h_{c\bar{b}}\partial_a h^{d\bar{e}}\partial^c h_{d\bar{e}} - \frac{1}{16}h^{a\bar{b}}h_{a\bar{c}}\partial_{\bar{b}} h^{d\bar{e}}\partial^{\bar{c}} h_{d\bar{e}} \\
&+ \frac{1}{16}h_{a\bar{b}}h^{c\bar{d}}\partial^a h_{e\bar{d}}\partial_c h^{e\bar{b}} + \frac{1}{16}h_{a\bar{b}}h^{c\bar{d}}\partial^{\bar{b}} h_{c\bar{e}}\partial_{\bar{d}} h^{a\bar{e}} \\
&- \frac{1}{24}h_{a\bar{b}}h_{c\bar{d}}\partial^c h^{e\bar{b}}\partial_e h^{a\bar{d}} - \frac{1}{24}h_{a\bar{b}}h_{c\bar{d}}\partial^{\bar{b}} h^{c\bar{e}}\partial_{\bar{e}} h^{a\bar{d}} \\
&+ \frac{1}{48}h^{a\bar{b}}h^{c\bar{d}}\partial_e h_{c\bar{b}}\partial^e h_{a\bar{d}} + \frac{1}{48}h^{a\bar{b}}h^{c\bar{d}}\partial_{\bar{e}} h_{c\bar{b}}\partial^{\bar{e}} h_{a\bar{d}} \\
&+ \frac{1}{12}h^{a\bar{b}}h_{c\bar{b}}\partial^c h^{e\bar{d}}\partial_e h_{a\bar{d}} + \frac{1}{12}h^{a\bar{b}}h_{a\bar{c}}\partial^{\bar{c}} h^{d\bar{e}}\partial_{\bar{e}} h_{d\bar{b}} \\
&- \frac{1}{96}h^{a\bar{b}}h_{c\bar{b}}\partial^e h^{c\bar{d}}\partial_e h_{a\bar{d}} - \frac{1}{96}h^{a\bar{b}}h_{c\bar{b}}\partial^{\bar{e}} h^{c\bar{d}}\partial_{\bar{e}} h_{a\bar{d}} \\
&+ \frac{1}{48}h^{a\bar{b}}h_{a\bar{c}}\partial^d h^{e\bar{c}}\partial_e h_{d\bar{b}} + \frac{1}{48}h^{a\bar{b}}h_{c\bar{b}}\partial^{\bar{d}} h^{c\bar{e}}\partial_{\bar{e}} h_{a\bar{d}} \\
&- \frac{1}{96}h^{a\bar{b}}h_{a\bar{c}}\partial^e h^{d\bar{c}}\partial_e h_{d\bar{b}} - \frac{1}{96}h^{a\bar{b}}h_{a\bar{c}}\partial^{\bar{e}} h^{d\bar{c}}\partial_{\bar{e}} h_{d\bar{b}} \\
\mathcal{O}_5 =& + \frac{1}{24}h^{a\bar{b}}h_{a\bar{c}}h_{d\bar{b}}\partial^{\bar{c}} h^{e\bar{f}}\partial^d h_{e\bar{f}} - \frac{1}{24}h^{a\bar{b}}h_{a\bar{c}}h_{d\bar{e}}\partial^{\bar{c}} h^{f\bar{e}}\partial^d h_{f\bar{b}} \\
&+ \frac{1}{24}h^{a\bar{b}}h^{c\bar{d}}h^{e\bar{f}}\partial_{\bar{b}} h_{c\bar{f}}\partial_e h_{a\bar{d}} - \frac{1}{24}h^{a\bar{b}}h_{c\bar{b}}h_{d\bar{e}}\partial^c h^{d\bar{f}}\partial^{\bar{e}} h_{a\bar{f}} \\
&+ \frac{1}{48}h^{a\bar{b}}h_{c\bar{b}}h^{d\bar{e}}\partial_d h_{a\bar{f}}\partial_{\bar{e}} h^{c\bar{f}} + \frac{1}{48}h^{a\bar{b}}h_{a\bar{c}}h^{d\bar{e}}\partial_d h_{f\bar{b}}\partial_{\bar{e}} h^{f\bar{c}} \\
&- \frac{1}{24}h^{a\bar{b}}h^{c\bar{d}}h^{e\bar{f}}\partial_e h_{a\bar{d}}\partial_{\bar{f}} h_{c\bar{b}} + \frac{1}{48}h^{a\bar{b}}h_{c\bar{b}}h^{d\bar{e}}\partial^c h_{d\bar{f}}\partial^{\bar{f}} h_{a\bar{e}} \\
&- \frac{1}{48}h^{a\bar{b}}h_{c\bar{b}}h^{d\bar{e}}\partial_d h^{c\bar{f}}\partial_{\bar{f}} h_{a\bar{e}} + \frac{1}{48}h^{a\bar{b}}h_{a\bar{c}}h^{d\bar{e}}\partial^{\bar{c}} h_{f\bar{e}}\partial^f h_{d\bar{b}} \\
&- \frac{1}{48}h^{a\bar{b}}h_{a\bar{c}}h^{d\bar{e}}\partial_{\bar{e}} h^{f\bar{c}}\partial_f h_{d\bar{b}} - \frac{1}{48}h^{a\bar{b}}h_{a\bar{c}}h_{d\bar{b}}\partial^{\bar{c}} h_{f\bar{e}}\partial^f h^{d\bar{e}} \\
&- \frac{1}{48}h^{a\bar{b}}h_{a\bar{c}}h_{d\bar{b}}\partial^d h_{e\bar{f}}\partial^{\bar{f}} h^{e\bar{c}}.
\end{aligned}
\tag{7.3.21}
$$

As expected, the barred and unbarred indices are all contracted consistently.

Second, the action (7.3.20) is manifestly invariant under a $\mathbb{Z}_2$ parity that swaps the two D-dimensional spacetimes,

$$
\begin{aligned}
x^a &\leftrightarrow \bar{x}^{\bar{a}} \\
h_{a\bar{b}} &\leftrightarrow h_{\bar{a}b}.
\end{aligned}
\tag{7.3.22}
$$

In terms of the full $2D$-dimensional objects, this $\mathbb{Z}_2$ parity acts as

$$H \;\leftrightarrow\; \tau H \tau \qquad \text{and} \qquad \eta \;\leftrightarrow\; \tau \eta \tau = \eta, \tag{7.3.23}$$

where we have defined the swap operator

$$\tau_{AB} = \begin{bmatrix} 0 & \mathbb{1} \\ \mathbb{1} & 0 \end{bmatrix}. \tag{7.3.24}$$

The symmetric and antisymmetric components of $h_{a\bar{b}}$ are manifestly even and odd under this parity, respectively. The former corresponds to the usual physical graviton modes, while the latter is an additional two-form field. However, since the antisymmetric component is odd under the $\mathbb{Z}_2$ parity, it enters the action in pairs and thus does not contribute to tree-level graviton scattering amplitudes. Thus, since Eq. (7.3.20) is expressed in terms of a general graviton tensor $h_{a\bar{b}}$, it is, strictly speaking, only equivalent to pure gravity at tree level.

The above construction is very much reminiscent of one discovered previously in the context of double field theory and there is a close link between our approaches. Double field theory [194–200] is derived from the massless modes of closed string field theory on a doubled torus exhibiting a manifest $O(D, D)$ T-duality group. The resulting low-energy effective theory is comprised of the graviton plus additional massless degrees of freedom: a dilaton and Kalb-Ramond two-form field necessary to maintain diffeomorphism invariance of the full space. Similarly motivated by Ref. [193], Hohm [197] constructed a form of the double field theory action that maintains index factorization as a low-energy remnant of the underlying T-duality. The resulting action is quite similar to our Eq. (7.3.20), except that is has both a massless dilaton and two-form. In this sense, our result is a derivation of a consistent truncation of this action in which the dilaton is not present. Conversely, the fact that our results are applicable in standard general relativity, i.e., without a dilaton, mean that they are directly relevant for calculations pertinent to our own

universe, e.g., scattering amplitudes in Einstein gravity and gravitational wave computations.

7.3.3 Scattering Amplitudes

The action in Eq. (7.3.20) is a rewriting of the EH action that manifests index factorization and twofold Lorentz symmetry. We now study how these properties are encoded in scattering amplitudes. All interaction vertices will be twofold Lorentz invariant even off-shell. To determine the symmetries of the propagator, we study the kinetic term in momentum space. Sending $\partial \to ip$, we obtain

$$\mathcal{O}_2 = \frac{1}{4} h_{a\bar{b}} h_{c\bar{d}} K^{a\bar{b}c\bar{d}}, \qquad \text{where} \qquad K^{a\bar{b}c\bar{d}} = -p^2 \eta^{ac} \eta^{\bar{b}\bar{d}} + \eta^{ac} p^{\bar{b}} p^{\bar{d}} + p^a p^c \eta^{\bar{b}\bar{d}}. \tag{7.3.25}$$

We can systematically determine the zero eigenvectors of the kinetic term by solving

$$0 = K^{a\bar{b}c\bar{d}} h_{c\bar{d}} = -p^2 h^{a\bar{b}} + h^{a\bar{d}} p_{\bar{d}} p^{\bar{b}} + p^a p_c h^{c\bar{b}}, \tag{7.3.26}$$

where indices are raised and lowered with η_{ab} and $\eta_{\bar{a}\bar{b}}$. Dotting this equation into p_a, we obtain

$$0 = p_a h^{a\bar{d}} p_{\bar{d}} p^{\bar{b}}, \tag{7.3.27}$$

which is trivially satisfied for the antisymmetric component of $h_{a\bar{b}}$. This equation also vanishes for the symmetric component of $h_{a\bar{b}}$ when it takes the form of a transverse diffeomorphism, $h_{ab} = \partial_a \xi_b + \partial_b \xi_a$ with $\partial_b \partial_a \xi^a = 0$. The existence of zero eigenmodes of the kinetic term implies that $h_{a\bar{b}}$ does not yet have an invertible kinetic term.

To remedy this, recall that the antisymmetric component of $h_{a\bar{b}}$ enters the action in pairs on account of the underlying $\mathbb{Z}_2$ parity, so it decouples from tree-level graviton scattering. We must also, however, modify the gauge-fixing of

the symmetric piece in order to produce an invertible kinetic term. In principle, there are many prescriptions for doing so. Here we consider a gauge-fixing that is manifestly twofold Lorentz symmetric at the expense of the $\mathbb{Z}_2$ parity, so

$$K_\xi^{a\bar{b}c\bar{d}} = -p^2\eta^{ac}\eta^{\bar{b}\bar{d}} + \left(1+\frac{1}{\xi}\right)\eta^{ac}p^{\bar{b}}p^{\bar{d}} + \left(1-\frac{1}{\xi}\right)p^a p^c \eta^{\bar{b}\bar{d}}, \tag{7.3.28}$$

where we take $\xi \to 0$ in the analogue of Landau gauge for gauge theory. The corresponding propagator, $\Delta_{a\bar{b}c\bar{d}}$, satisfies

$$K_\xi^{a\bar{b}c\bar{d}}\Delta_{c\bar{d}e\bar{f}} = i\delta^a_e\delta^{\bar{b}}_{\bar{f}}, \tag{7.3.29}$$

from which we obtain

$$\Delta_{a\bar{b}c\bar{d}} = -\frac{i}{p^2}\left(\eta_{ac}\eta_{\bar{b}\bar{d}} - (1+\xi)\eta_{ac}\frac{p_{\bar{b}}p_{\bar{d}}}{p^2} - (1-\xi)\frac{p_a p_c}{p^2}\eta_{\bar{b}\bar{d}}\right). \tag{7.3.30}$$

At zeroth order in ξ, the $\mathbb{Z}_2$ parity of the propagator is restored, yielding a simple and convenient propagator for explicit computations. Contributions first order in ξ also vanish because the underlying $\mathbb{Z}_2$ parity of the interactions eliminates all odd powers of ξ dependence from the tree-level graviton scattering amplitude. Note that to obtain consistent answers, it is crucial to use the fully gauge-fixed propagator in Eq. (7.3.30) with the the factorized action in Eq. (7.3.20) and its perturbative expansion. That is, dropping the delineation between barred and unbarred indices will yield inconsistent results. In this sense, the two-form is critical for the gauge-fixing introduced in Eq. (7.3.28), even though it does not appear as an external state in graviton scattering amplitudes. We have checked explicitly that the Feynman diagrams constructed from the propagator in Eq. (7.3.30) and the interaction vertices of Eq. (7.3.20) produce the correct three-, four-, and five-point amplitudes, even for finite ξ.

More generally, in this gauge, all off-shell Feynman diagrams are invariant under twofold Lorentz symmetry as well as $\mathbb{Z}_2$ exchange. Furthermore, the

resulting tree-level scattering amplitudes are invariant under the twofold Ward identities defined in Eq. (7.1.4). The reason for this is simple: the symmetric combination of gauge transformations is an invariance of graviton scattering, while the antisymmetric combination decouples because this mode only enters in pairs and thus does not contribute to pure graviton scattering at tree level.

Finally, let us emphasize that the action presented here is distinct from the action constructed perturbatively up to fifth order in Ref. [193]. This is evident from our propagator, which is different from the propagator assumed in Ref. [193].

7.3.4 Alternative Representations

We have presented a simple representation of the EH action that manifests index factorization and in turn twofold Lorentz symmetry. Now, by again exploiting the freedom to choose a field basis, we can generate an infinite class of physically equivalent actions that manifest the same symmetries. In particular, we can consider field redefinitions of the form

$$h_{a\bar{b}} \to \alpha_1 h_{a\bar{b}} + \alpha_3 h^3_{a\bar{b}} + \alpha_5 h^5_{a\bar{b}} + \cdots , \tag{7.3.31}$$

where again we have $\alpha_1 = 1$ to maintain the form of the asymptotic states. Here, the field redefinition involves only odd powers of the graviton defined by Eq. (7.3.19), so that barred and unbarred indices are properly contracted. More generally, one can consider an arbitrary sum over $h^n_{a\bar{b}}$ for odd n, with each term multiplied by $[h^m]$ for some even m, which preserves the ability to consistently factorize indices.

In general, this additional set of field redefinitions can further simplify various parts of the action. For example, to eliminate the appearance of hyperbolic functions in Eq. (7.3.16), we could send

$$h_{a\bar{b}} \to (\sinh^{-1} h)_{a\bar{b}} = h_{a\bar{b}} - \frac{1}{6} h^3_{a\bar{b}} + \frac{3}{40} h^5_{a\bar{b}} + \cdots , \tag{7.3.32}$$

so that the EH action is just as in Eq. (7.3.20), except with a new field defined as

$$\Sigma_{AB} \to \begin{bmatrix} (\sqrt{1+h^2})_{ab} & h_{a\bar{b}} \\ h_{\bar{a}b} & (\sqrt{1+h^2})_{\bar{a}\bar{b}} \end{bmatrix}. \tag{7.3.33}$$

In what follows, we discuss an alternative field basis for the action that results in even simpler expressions for graviton perturbation theory.

In particular, inspired by the so-called Cayley basis for the nonlinear sigma model action [201], it is natural to consider the field redefinition

$$h_{a\bar{b}} \to \log\left(\frac{1+\frac{1}{2}h}{1-\frac{1}{2}h}\right)_{a\bar{b}} = h_{a\bar{b}} + \frac{1}{12}h^3_{a\bar{b}} + \frac{1}{80}h^5_{a\bar{b}} + \cdots, \tag{7.3.34}$$

for which the field in the doubled spacetime becomes

$$\Sigma_{AB} = \begin{bmatrix} \left(\frac{1+h^2/4}{1-h^2/4}\right)_{ab} & \left(\frac{h}{1-h^2/4}\right)_{a\bar{b}} \\ \left(\frac{h}{1-h^2/4}\right)_{\bar{a}b} & \left(\frac{1+h^2/4}{1-h^2/4}\right)_{\bar{a}\bar{b}} \end{bmatrix}. \tag{7.3.35}$$

The first few terms in the perturbation expansion are

$$\begin{aligned}
\mathcal{O}_2 =& + \frac{1}{2}\partial_c h_{ab}\partial^b h^{ac} - \frac{1}{4}\partial_c h_{ab}\partial^c h^{ab} \\
\mathcal{O}_3 =& + \frac{1}{4}h^{ab}\partial_a h_{cd}\partial_b h^{cd} - \frac{1}{2}h^{ab}\partial_c h_{ad}\partial_b h^{cd} \\
\mathcal{O}_4 =& + \frac{1}{8}h_{ab}h^{cd}\partial^b h_{ce}\partial_d h^{ae} - \frac{1}{8}h^{ab}h_{ac}\partial_b h_{de}\partial^c h^{de} \\
&+ \frac{1}{4}h^{ab}h_{ac}\partial^c h^{de}\partial_e h_{db} + \frac{1}{8}h^{ab}h_{ac}\partial^d h^{ec}\partial_e h_{db} \\
&- \frac{1}{8}h^{ab}h_{ac}\partial^e h^{dc}\partial_e h_{db} \\
\mathcal{O}_5 =& - \frac{1}{8}h^{ab}h_{ac}h_{de}\partial^c h^{fe}\partial^d h_{fb} + \frac{1}{16}h^{ab}h_{ac}h_{db}\partial^c h^{ef}\partial^d h_{ef} \\
&+ \frac{1}{8}h^{ab}h_{ac}h^{de}\partial_d h_{fb}\partial_e h^{fc} - \frac{1}{8}h^{ab}h_{ac}h^{de}\partial_e h^{fc}\partial_f h_{db} \\
&- \frac{1}{8}h^{ab}h_{ac}h_{db}\partial^d h_{ef}\partial^f h^{ec},
\end{aligned} \tag{7.3.36}$$

after dropping the distinction between barred and unbarred indices.

We immediately note that the Cayley-like basis yields fewer terms than our action in Eq. (7.3.6)—for which the first few orders are given in Eq. (7.3.10)—and far fewer terms than occur in the canonical graviton perturbation theory of

the EH action. In particular, at $\mathcal{O}(h^n)$ for $n = 2, 3, 4, 5$, the canonical graviton perturbation yields 4, 13, 35, 76 terms in the action, respectively, counted such that no single graviton is acted upon with two derivatives.

A unique aspect of the Cayley-like basis (7.3.34) is that it makes the action invariant up to a sign-flip under the duality of small and large graviton perturbations. Specifically, consider a metric perturbation h_{ab} that has a nonsingular matrix inverse h^{-1}_{ab}. Then, in the Cayley-like basis, the transformation

$$\frac{h_{ab}}{2} \to \left(\frac{h_{ab}}{2}\right)^{-1} \tag{7.3.37}$$

merely induces a sign in the field

$$\sigma_{ab} \to -\sigma_{ab} \tag{7.3.38}$$

and thus sends the action to minus itself, which simply flips the sign of $\hbar$ and is thus an invariance of the interactions. This invariance, which is unique to the Cayley-like basis, is reminiscent of T-duality, but is more general in the sense that it applies to arbitrary invertible metric perturbations, while more specific in that it applies to the pure gravity theory considered in this chapter.

7.4 Generalizing to Curved Spacetime

In Secs. 7.2 and 7.3, we presented a factorized form of the pure gravity action expanded around a flat background. We will now generalize this construction to curved spacetime, first in terms of the full metric and then for perturbations around a nontrivial background. Afterwards, we derive the corresponding factorized Einstein equations.

7.4.1 Lifting to Curved Spacetime

Although the action in Eq. (7.3.6) was derived by expanding about flat spacetime, it remains valid to all orders in the graviton perturbation. This implies

that this action encodes the physics of large graviton field variations away from flat spacetime, i.e., a curved background. In particular, by combining Eq. (7.3.1) with Eq. (7.3.7), we see that the nonlinear field defined earlier is simply

$$\sigma^{ab} = \sqrt{-g}\, g^{ab}. \tag{7.4.1}$$

Remarkably, this combination of fields arises naturally from the EH action in curved spacetime. After some rearrangement, one can show that

$$\begin{aligned}\sqrt{-g}\, R &= \sqrt{-g}\left[\partial_a g_{ce}\partial_b g^{de}\left(\frac{1}{4}g^{ab}\delta^c_d - \frac{1}{2}g^{cb}\delta^a_d\right) - g^{ab}\partial_a\partial_b(\log\sqrt{-g})\right] \\ &\qquad + \cdots \\ &= \sqrt{-g}\left[\partial_a\left(\frac{g_{ce}}{\sqrt{-g}}\right)\partial_b\left(\sqrt{-g}\, g^{de}\right)\left(\frac{1}{4}g^{ab}\delta^c_d - \frac{1}{2}g^{cb}\delta^a_d\right)\right. \\ &\qquad \left. + \frac{D-2}{4}g^{ab}\partial_a(\log\sqrt{-g})\partial_b(\log\sqrt{-g})\right] + \cdots,\end{aligned} \tag{7.4.2}$$

which is naturally a function of σ_{ab} and σ^{ab} and where ellipses denote total derivative contributions. Something similar arises when we expand in graviton perturbations around a background spacetime $\tilde{g}_{ab}$. To see this, we lift the nonlinear field into curved spacetime, defining

$$\sqrt{-\tilde{g}}\, \sigma^{ab} = \sqrt{-g}\, g^{ab}. \tag{7.4.3}$$

Furthermore, we define ω_a as before and $\tilde{\omega}_a$ analogously,

$$\omega_a = \partial_a \log\sqrt{-g} \qquad \text{and} \qquad \tilde{\omega}_a = \partial_a \log\sqrt{-\tilde{g}}, \tag{7.4.4}$$

as well as their difference,

$$\Omega_a = \omega_a - \tilde{\omega}_a, \tag{7.4.5}$$

which enters the curved-background generalization of Eq. (7.3.4),

$$\mathcal{L}_{\text{GF}} = -\frac{D-2}{64\pi G} g^{ab}\Omega_a\Omega_b. \tag{7.4.6}$$

Let us comment on the physical interpretation of this gauge-fixing. At the level of the gravity action, the gauge condition is a constraint on the full metric

g_{ab} or, equivalently, on the metric perturbation h_{ab} in a given field basis. The gauge-fixing Lagrangian $\mathcal{L}_{\rm GF}$ itself can be viewed as being added simply to cancel expressions in the non-gauge-fixed equations of motion that vanish when the gauge condition is satisfied. In our case, the gauge condition associated with $\mathcal{L}_{\rm GF}$ is

$$\Omega_a = \partial_a \log \sqrt{\frac{-g}{-\tilde{g}}} = 0, \tag{7.4.7}$$

which is different from the commonly used harmonic gauge condition, $\partial_b(g^{ab}\sqrt{-g}) = 0$. A gauge condition on the metric can be recast as a condition on the choice of coordinate system x^a, regarded as a set of D scalar functions on spacetime. In harmonic gauge, this corresponds to $\nabla_b \nabla^b x^a = 0$. The coordinate condition corresponding to our gauge condition in Eq. (7.4.7), in terms of the coordinates x^a for the spacetime g_{ab} and $\tilde{x}^a$ for the background spacetime $\tilde{g}_{ab}$, is

$$\nabla_b \nabla_a x^b = \tilde{\nabla}_b \tilde{\nabla}_a \tilde{x}^b, \tag{7.4.8}$$

using that $\nabla_b \nabla_a x^b = -\omega_a$. Here, $\tilde{\nabla}_a$ is the covariant derivative on the background metric $\tilde{g}_{ab}$ and ∇_a is the covariant derivative defined with respect to the full perturbed metric g_{ab}.

Armed with the necessary definitions, we are ready to write the gravity action in terms of our field redefinition and gauge-fixing, generalized to an arbitrary background spacetime. First, we note that Eqs. (7.4.2) and (7.4.6) imply that Eq. (7.1.6) is, up to a total derivative,

$$S = \frac{1}{16\pi G} \int \mathrm{d}^D x \, \sqrt{-\tilde{g}} \left[\partial_a \sigma_{ce} \partial_b \sigma^{de} \left(\frac{1}{4} \sigma^{ab} \delta^c_d - \frac{1}{2} \sigma^{cb} \delta^a_d \right) - \sigma^{ab} \partial_a \tilde{\omega}_b \right]. \tag{7.4.9}$$

A useful identity for this simplification is $\sigma^{ab}\partial_c \sigma_{ab} = (2-D)\omega_c + D\tilde{\omega}_c$, which makes use of the fact that $g^{ab}\partial_c g_{ab} = 2\omega_c$. To derive an expression that is manifestly covariant with respect to the background spacetime, we recast

partial derivatives in terms of covariant derivatives and Christoffel symbols of the background metric. We then obtain an action that is a nice generalization of Eq. (7.3.6) to an arbitrary curved background spacetime,

$$S = \frac{1}{16\pi G}\int \mathrm{d}^D x\,\sqrt{-\tilde{g}}\left[\tilde{\nabla}_a\sigma_{ce}\tilde{\nabla}_b\sigma^{de}\left(\frac{1}{4}\sigma^{ab}\delta^c_d - \frac{1}{2}\sigma^{cb}\delta^a_d\right) + \sigma^{ab}\tilde{R}_{ab}\right], \quad (7.4.10)$$

where $\tilde{R}_{ab}$ is the Ricci tensor of the background spacetime. This action reverts back to Eq. (7.3.6) in the flat-spacetime limit.

Note that Eq. (7.4.10) applies independently of the precise field basis for the graviton perturbations, merely requiring the existence of an object σ_{ab} consistent with Eq. (7.4.3), as well as the gauge fixing in Eq. (7.4.6). For concreteness, we now give an explicit field basis for the graviton, for which the required object exists and thus for which Eq. (7.4.10) is the action. Lifting Eqs. (7.2.7) and (7.3.1) to curved spacetime, we define the full metric g_{ab} to be

$$g_{ab} = \tilde{g}_{ab} + \pi_{ab} + \frac{1}{2!}\pi^2_{ab} + \frac{1}{3!}\pi^3_{ab} + \cdots, \text{ where } \pi_{ab} = h_{ab} - \frac{1}{D-2}\tilde{g}_{ab}[h] \quad (7.4.11)$$

and where we have defined

$$\begin{aligned} h^n_{ab} &= h_{ab_1}\tilde{g}^{b_1a_1}h_{a_1b_2}\tilde{g}^{b_2a_2}\cdots h_{a_{n-2}b_{n-1}}\tilde{g}^{b_{n-1}a_{n-1}}h_{a_{n-1}b} \\ &= h_a{}^{a_1}h_{a_1}{}^{a_2}\cdots h_{a_{n-2}}{}^{a_{n-1}}h_{a_{n-1}b}, \end{aligned} \quad (7.4.12)$$

with traces $[h^n] = h^n_{ab}\tilde{g}^{ba}$. We now define an exponential field

$$\sigma_{ab} = \tilde{g}_{ab} + h_{ab} + \frac{1}{2!}h^2_{ab} + \frac{1}{3!}h^3_{ab} + \cdots = (e^h)_{ab} \quad (7.4.13)$$

and similarly redefine $\sigma^{ab} = (e^{-h})^{ab}$ with indices in its expansion contracted using $\tilde{g}_{ab}$. With these definitions, along with the useful relation $\sqrt{-g} = \sqrt{-\tilde{g}}\,e^{-[h]/(D-2)}$, the nonlinear field σ_{ab} satisfies the property desired in Eq. (7.4.3). Hence, the action for the graviton, to all orders in perturbation theory, expanded about an arbitrary background spacetime as in Eq. (7.4.11) and gauge-fixed according to Eq. (7.4.6), is given in Eq. (7.4.10). In this field basis, the gauge condition (7.4.7) becomes $\partial_a[h] = 0$.

Generically, a nonzero value of $\tilde{R}_{ab}$ will violate index factorization since $\sigma^{ab}\tilde{R}_{ab}$ unavoidably contracts left and right indices in all odd powers of h_{ab} in σ^{ab}. For example, this will occur in (anti-)de Sitter space, where $\tilde{R}_{ab} \propto \tilde{g}_{ab}$. However, for a background vacuum solution $\tilde{R}_{ab} = 0$, the action (7.4.10) factorzies when expressed in terms of $2D$-dimensional spacetime, so

$$\begin{aligned} S = \frac{1}{16\pi G} \int \mathrm{d}^D x \, \mathrm{d}^D \bar{x} \, \delta^D(x - \bar{x}) \sqrt{-\tilde{g}} \times \\ \times \tilde{\nabla}_A \Sigma_{CE} \tilde{\nabla}_B \Sigma^{DE} \left(\frac{1}{16} \Sigma^{AB} \delta^C_D - \frac{1}{4} \Sigma^{CB} \delta^A_D \right), \end{aligned} \tag{7.4.14}$$

where $\tilde{\nabla}_A = (\tilde{\nabla}_a, \tilde{\nabla}_{\bar{a}})$. This action applies for any background vacuum solution to the Einstein equations, including the Schwarzschild and Kerr metrics, Taub-NUT space, gravitational wave backgrounds, etc. In all of these cases, Eq. (7.4.14) provides an all-orders factorized representation of the perturbation theory. As a special case, Eq. (7.4.14) can accommodate any background metric on Minkowski spacetime, e.g., curvilinear coordinates, as opposed to the strict Cartesian coordinate system necessary for the formulation in Eq. (7.3.20). In addition to the nice factorization properties, the action is very simple in perturbation theory; indeed, the slow scaling of the number of terms discussed in Sec. 7.3.4 is equally applicable to Eq. (7.4.14). Hence, our result may have applicability to the treatment of black hole perturbations, nonlinear gravitational wave effects, etc.

In general, a nontrivial background energy-momentum tensor $\tilde{T}_{ab}$ will source the Ricci curvature $\tilde{R}_{ab}$, thus violating index factorization. However, it is simple to see that one particular matter source actually remains compatible with twofold Lorentz symmetry: a massless, minimally-coupled free scalar. For a background source $\tilde{\phi}$, the energy-momentum tensor and Ricci tensor are

$$\tilde{T}_{ab} = \partial_a \tilde{\phi} \partial_b \tilde{\phi} - \frac{1}{2} \tilde{g}_{ab} \partial_c \tilde{\phi} \partial^c \tilde{\phi} \qquad \text{and} \qquad \tilde{R}_{ab} = 8\pi G \partial_a \tilde{\phi} \partial_b \tilde{\phi}. \tag{7.4.15}$$

Moreover, the matter action, in the full perturbed spacetime metric with the $\tilde{\phi}$ background, is

$$S_{\text{matt}} = -\frac{1}{2}\int \mathrm{d}^D x\, \sqrt{-g}\, g^{ab}\partial_a\tilde{\phi}\partial_b\tilde{\phi} = -\frac{1}{2}\int \mathrm{d}^D x\, \sqrt{-\tilde{g}}\, \sigma^{ab}\partial_a\tilde{\phi}\partial_b\tilde{\phi}. \quad (7.4.16)$$

In this case, the contribution to the action (7.4.10) from $\sigma^{ab}\tilde{R}_{ab}$ and the matter action S_{matt} are both separately compatible with index factorization and moreover exactly cancel each other. The individual index factorization of the two terms and the cancellation between S_{matt} and $\sigma^{ab}\tilde{R}_{ab}$ both stem from the fact that the matter Lagrangian for the free massless scalar is linear in the metric g^{ab}, which allows for the background value of the scalar action to be equal to $\sqrt{-g}\, g^{ab}\tilde{R}_{ab}$. In this case, the background value of the scalar becomes irrelevant to the gravity action, which in factorized form reduces to that given in Eq. (7.4.14).

7.4.2 Equations of Motion

As we saw previously, the twofold Lorentz invariance of the action is directly manifest in the corresponding Feynman diagrams. Moreover, this property should be exhibited by the equations of motion, i.e., the Einstein equations. In this section, we compute the Einstein equations, to all orders in perturbation theory in our chosen field basis, about an arbitrary curved spacetime background. A priori, one can compute the equations of motion corresponding to field variations of g_{ab}, σ_{ab}, or h_{ab}, but all of these are related to each other by an appropriate Jacobian.

The Einstein equations with respect to g_{ab} are of the standard form,

$$R_{ab} - \frac{1}{2}R\, g_{ab} = 8\pi G\, T_{ab}, \quad (7.4.17)$$

where we have defined the stress-energy tensor

$$T_{ab} = -\frac{2}{\sqrt{-g}} \frac{\delta(\sqrt{-g}\,\mathcal{L}_{\rm matt})}{\delta g^{ab}} \tag{7.4.18}$$

for matter Lagrangian $\mathcal{L}_{\rm matt}$. Now, let us relate the usual Einstein equations to the equation of motion corresponding to the variation of σ_{ab}. The Jacobian relating σ_{ab} and g_{ab} is

$$J^{ab}{}_{cd} = \sqrt{\frac{-g}{-\tilde{g}}} \frac{\delta g^{ab}}{\delta \sigma^{cd}} = \frac{1}{2}\left(\delta^a_c \delta^b_d + \delta^a_d \delta^b_c\right) - \frac{1}{D-2} g^{ab} g_{cd}, \tag{7.4.19}$$

which has the structure of the graviton propagator numerator in harmonic gauge. To obtain the equations of motion for σ_{ab}, we multiply Eq. (7.4.17) by the Jacobian, yielding

$$\begin{aligned} & J^{cd}{}_{ab}\left(R_{cd} - \frac{1}{2} R\, g_{cd}\right) = 8\pi G\, J^{cd}{}_{ab} T_{cd} \\ \Longrightarrow\ & R_{ab} = 8\pi G \left(T_{ab} - \frac{1}{D-2} T g_{ab}\right), \end{aligned} \tag{7.4.20}$$

which are just another form of the Einstein equations. Varying the action (7.4.10) with respect to σ^{ab}, we obtain the equations of motion to all orders in perturbation theory,

$$\begin{aligned} R_{ab} = {} & \frac{1}{2}\tilde{\nabla}_c \left(\sigma^{cd}\tilde{\nabla}_a \sigma_{bd} + \sigma^{cd}\tilde{\nabla}_b \sigma_{ad} - \sigma^{cd}\tilde{\nabla}_d \sigma_{ab}\right) \\ & + \frac{1}{2}\left(\sigma^{ce}\sigma^{df} - \sigma^{cf}\sigma^{de}\right)\tilde{\nabla}_d \sigma_{ac} \tilde{\nabla}_f \sigma_{be} + \frac{1}{4}\tilde{\nabla}_a \sigma_{cd} \tilde{\nabla}_b \sigma^{cd} + \tilde{R}_{ab}. \end{aligned} \tag{7.4.21}$$

A useful trick for handling the inverse matrix σ^{ab} in the equations of motion is to introduce a constraint term $\lambda^a_b(\sigma_{ac}\sigma^{cb} - \delta^b_a)$, where λ is a Lagrange multiplier. As consistency check, we can instead write R_{ab} explicitly in terms of g_{ab} for the perturbation expansion about flat spacetime, substituting in our field redefinition from Eq. (7.3.1). Indeed, in this case we obtain the same result as the flat-background limit of Eq. (7.4.21).

Let us momentarily consider the linearized Einstein equations in the flat-spacetime limit,

$$-\Box h_{ab} + \partial_a \partial_c h^c{}_b + \partial_b \partial_c h^c{}_a = 16\pi G \left(T_{ab} - \frac{1}{D-2} T \eta_{ab} \right), \tag{7.4.22}$$

where $\Box = \eta^{ab}\partial_b\partial_a$. The left-hand side of Eq. (7.4.22) is nearly the same as the general, non-gauge-fixed linearized field equations in the so-called trace-reversed basis [118], but is missing the term $-\eta_{ab}\partial^c\partial^d h_{cd}$, which violates index factorization and which was removed by our gauge-fixing procedure. Note that we have not eliminated all of the available gauge freedom, since we can shift the coordinate functions x^a by a perturbation δx^a satisfying $\nabla_b \nabla_a \delta x^b = 0$. Equivalently, as noted in Sec. 7.3.3, we can send $h_{ab} \to h_{ab} + \partial_a \xi_b + \partial_b \xi_a$, as long as $\partial^a \xi_a = \text{constant}$, so that $\partial_a[h] = 0$ and the gauge condition (7.4.7) remains satisfied. In particular, for a radiative solution in which $T_{ab} = 0$, we can choose ξ^a such that $\Box \xi^a = -\partial_b h^{ab}$, in which case we find that, after the shift, the perturbation satisfies $\partial^a h_{ab} = 0$. Hence, the vacuum equation reduces to the usual wave equation $\Box h_{ab} = 0$.

We now turn back to the general case of the Einstein equation for perturbation theory in an arbitrary background spacetime. To be as general as possible, we will for now ignore the issue of factorization and merely consider some implications of Eq. (7.4.21), the equation of motion for the gravity action (7.4.10). While Eq. (7.4.10) appears with a tadpole in the graviton, $h^{ab}\tilde{R}_{ab}$, this is precisely canceled by additional tadpole terms generated by the matter action. This is mandated by the equations of motion for the background,

$$\tilde{R}_{ab} - \frac{1}{2}\tilde{R}\,\tilde{g}_{ab} = 8\pi G\,\tilde{T}_{ab}, \tag{7.4.23}$$

where $\tilde{R} = \tilde{R}_{ab}\tilde{g}^{ba}$. So as long as the background spacetime satisfies the Einstein equation, the tadpole in the action in Eq. (7.4.10) is canceled.

For backgrounds with vanishing $\tilde{R}_{ab}$, we can also write the equations of motion associated with the action in Eq. (7.4.14) in terms of the fields Σ_{AB}

on the doubled spacetime, so that the factorization of indices in $h_{a\bar{b}}$ occurs automatically. Doing so, we can rewrite Eq. (7.4.21) as

$$\begin{aligned}[R_{AB}]_{x=\bar{x}} = \Bigg[&\frac{1}{4}\tilde{\nabla}_C\left(\Sigma^{CD}\tilde{\nabla}_B\Sigma_{AD} + \Sigma^{DC}\tilde{\nabla}_A\Sigma_{DB} - \frac{1}{2}\Sigma^{CD}\tilde{\nabla}_D\Sigma_{AB}\right)\\ &+\frac{1}{8}\left(\Sigma^{EC}\Sigma^{FD} - \Sigma^{FC}\Sigma^{ED}\right)\tilde{\nabla}_D\Sigma_{AC}\tilde{\nabla}_F\Sigma_{EB}\\ &+\frac{1}{8}\tilde{\nabla}_A\Sigma_{CD}\tilde{\nabla}_B\Sigma^{CD}\Bigg]_{x=\bar{x}},\end{aligned} \tag{7.4.24}$$

where R_{AB} is the lift of the Ricci tensor into the $2D$-dimensional space. We thus have a factorized form of the Einstein equations, valid to all orders in perturbation theory about an arbitrary curved vacuum background spacetime. Note, however, that if one simply varies the doubled-spacetime action in Eq. (7.4.14) with respect to Σ_{AB}, the resulting expression contains various errors in factors of two compared to the correct expression in Eq. (7.4.24), since the Lagrangian is integrated only over the diagonal spacetime $x = \bar{x}$. If one substitutes explicit expressions for Σ_{AB} in terms of $h_{a\bar{b}}$ and then drops all bars, setting $x = \bar{x}$, then Eq. (7.4.24) reduces to Eq. (7.4.21) with $\tilde{R}_{ab} = 0$ as required.

7.5 Conclusions

In this chapter, we have described a systematic search for a pure gravity action exhibiting the twofold Lorentz symmetry suggested by the double copy relations. This property is manifested by two sets of indices, barred and unbarred, that are independently contracted and naturally parameterized by an auxiliary set of extra spacetime dimensions. By exploring the space of nonlinear field redefinitions and local gauge-fixing of the Einstein-Hilbert action discussed in Sec. 7.2, we derived the twofold Lorentz invariant action described in Sec. 7.3. This action extends to an infinite family of actions related by twofold Lorentz

invariant field redefinitions. Some choices, e.g., the case of the Cayley-like basis explored in Sec. 7.3.4, possess enhanced simplicity in terms of the reduced number of Lorentz invariant structures present in the action at each order in perturbation theory.

Because our results for flat-spacetime perturbation theory apply to all orders in the graviton field, they can be extended to curved spacetime. In Sec. 7.4, we derived a simple action for graviton perturbations around an arbitrary curved spacetime background in our field basis. Furthermore, we found that this action exhibited the same factorization properties for arbitrary Ricci-flat background spacetimes. We derived the Einstein equations in index-factorized form to all orders in the graviton about an arbitrary vacuum background and explored several interesting features they possess.

This work leaves a number of promising directions for future research. First of all, while we introduced auxiliary spacetime dimensions simply as a convenient bookkeeping tool, it is likely that these can be derived from a truly extra-dimensional construction. One path would be to understand how our action somehow arises as a truncation of double field theory that lifts the dilaton from the spectrum. Alternatively, one could introduce dynamics governing fluctuations of the D-dimensional region within the doubled spacetime or smear out the delta function in Eq. (7.3.20), modifying the theory in the ultraviolet.

Second, it would be illuminating to study the properties of graviton scattering amplitudes computed with this class of twofold Lorentz invariant actions. Indeed, it has long been known that the properties of on-shell graviton scattering amplitudes enjoy improved high-momentum behavior from the study of BCFW recursion relations [205, 206] for general gauge and gravity theories [207, 208]. As discussed in Ref. [207], these properties can be understand from a "spin Lorentz symmetry" that can be derived from the high-energy limit of

these theories. From this perspective, the results of this chapter are a nonlinear generalization of this property beyond the high-energy limit.

Last but not least, a critical open question is whether and how our results relate directly to the double copy construction for scattering amplitudes in gauge theory and gravity. Here, it would be extraordinary to somehow reformulate our family of twofold Lorentz invariant gravity actions as two bona fide gauge theory copies. The naive prescription—to simply substitute $h_{a\bar{b}} \sim A_a \bar{A}_{\bar{b}}$ at the level of Feynman vertices—is ambiguous since there are an infinite number of pure gravity actions from which one can start. Nevertheless, we believe that a formulation likely exists, in part because the analogous puzzle has been understood for the double copy of effective field theories, where new representations of the nonlinear sigma model and special Galileon theories [209] manifest these dualities as a symmetry of a cubic action. In any case, this chapter represents an initial step towards understanding the gauge and gravity double copy at the level of the action.

Chapter 8
Splitting Spacetime and Cloning Qubits: Linking No-Go Theorems across the ER=EPR Duality

We analyze the no-cloning theorem in quantum mechanics through the lens of the proposed ER=EPR (Einstein-Rosen = Einstein-Podolsky-Rosen) duality between entanglement and wormholes. In particular, we find that the no-cloning theorem is dual on the gravity side to the no-go theorem for topology change, violating the axioms of which allows for wormhole stabilization and causality violation. Such a duality between important no-go theorems elucidates the proposed connection between spacetime geometry and quantum entanglement.

This chapter is from Ref. [7], N. Bao, J. Pollack, and G. N. Remmen, "Splitting Spacetime and Cloning Qubits: Linking No-Go Theorems across the ER=EPR Duality," Fortsch. Phys. ***63*** *(2015) 705,* arXiv:1506.08203 [hep-th].

8.1 Introduction

The connection between entanglement and geometry is an unexpected stepping-stone on the path to an understanding of quantum gravity. Historically originating from black hole thermodynamics [210, 211] and later in the context of the holographic principle [50, 51], the AdS/CFT correspondence [52–54], entropy bounds [71], and the Ryu-Takayanagi formula [132], the relation between quantum entanglement and spacetime geometry is increasingly thought to be an important feature of a consistent theory of quantum gravity. Underscoring this view is recent work on deriving the Einstein equations holographically from entanglement constraints [212] and perhaps even spacetime itself from

qubits [10, 213]. However, significant puzzles remain. The classic black hole information paradox [125, 214] has given way to new questions about black hole interiors and their entanglement with Hawking radiation [215, 216]. One of the most drastic, albeit promising, proposals to arise from these debates is the so-called ER=EPR duality [57].

The ER=EPR correspondence [57] is a compelling [217, 218] proposal for an exact duality between Einstein-Podolsky-Rosen (EPR) pairs [58], that is, qubits entangled in a Bell state [219], and nontraversable wormholes, that is, Einstein-Rosen (ER) bridges [55, 126, 127]. More specifically, the ER=EPR proposal generalizes the notion of entangled black hole pairs at opposite ends of an ER bridge, by asserting that every pair of entangled qubits is connected by a Planck-scale quantum wormhole. The proposal, if true, would have profound implications for AdS/CFT and suggest a solution to the firewall paradox of Ref. [215], not to mention the fundamental shift it would induce in our understanding of both quantum mechanics and general relativity.

The ER=EPR correspondence might allow the exploration of gravitational analogues of fundamental properties of quantum systems (and vice versa). In particular, we can check whether there is a precise correspondence between no-go theorems in quantum mechanics and similar no-go theorems in gravity. Arguably the most celebrated no-go theorem in quantum mechanics is the no-cloning theorem [220], which prohibits the duplication of quantum states.

In this chapter, we investigate the manifestation of the no-cloning theorem on the gravitational side of the ER=EPR duality. In particular, we show that violation of the no-cloning theorem is dual under ER=EPR to topology-changing processes in general relativity, which, via classical topology-conservation theorems [221–227], lead to causal anomalies through violation of the Hausdorff condition (which leads to the breakdown of strong causality), creation of closed

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

8.2 Quantum Cloning

Here, we reconstruct the standard argument for why the no-cloning theorem prohibits superluminal signaling [228]. Assume that cloning of states is allowed, that is, that there exists an operation that takes an arbitrary state $|\Psi\rangle$ in a product state with some $|0\rangle$ state and replaces the $|0\rangle$ state with $|\Psi\rangle$:

$$|\Psi\rangle_A |0\rangle_B \to |\Psi\rangle_A |\Psi\rangle_B . \tag{8.2.1}$$

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

Bob now proceeds to clone his qubit as in Eq. (8.2.1). Note that each of his cloned qubits remains maximally entangled with Alice's qubit, in violation of monogamy of entanglement, while remaining unentangled with each other. By measuring enough of his own qubits in the σ_z basis, Bob can determine, to any desired degree of confidence, whether Alice performed a measurement or not: his measurements will all yield the same result if Alice performed a measurement, but will be equally and randomly split between the two outcomes if she did not. As this experiment does not depend on their separation, Bob's

utilization of cloning and their shared entanglement has allowed Alice to send one classical bit to Bob acausally.

8.3 Black Hole Cloning

In order to geometrically interpret the no-cloning theorem using the ER=EPR proposal, we need a system with both a high level of entanglement (like the EPR pair just considered) and a robust geometric description. One such system is the eternal AdS-Schwarzchild black hole, which is described in AdS/CFT by two noninteracting large-N CFTs in a thermally entangled state on the boundary sphere [56, 131]:

$$|\Psi\rangle = \frac{1}{\sqrt{Z}} \sum_n e^{-\beta E_n/2} |n\rangle_L \otimes |n\rangle_R \,, \tag{8.3.1}$$

where $|n\rangle_{L(R)}$ is the nth eigenstate on the left (respectively, right) CFT with energy E_n, β is the inverse temperature, and Z is the partition function. In this state, the reduced density matrices $\rho_{L,R}$ of either side are identically thermal. If both exterior regions of the geometry are considered [56, 131, 229], this state describes a spacetime consisting of two separate AdS-Schwarzchild regions that are spatially disconnected outside the horizon but linked by an ER bridge between a maximally entangled[8.1] pair of black holes with temperature β^{-1}. This is a concrete realization of ER=EPR: to reiterate, the two black holes are both maximally entangled (EPR) and connected by a nontraversable wormhole (ER). It will be convenient to consider the slight generalization of this setup in which the two black holes share the same asymptotic space. As discussed in Ref. [57], such black hole pairs can be naturally obtained as an instanton solution in a geometry with a constant magnetic field.

[8.1] Strictly speaking, the state is only truly maximally entangled when $\rho_L = \rho_R = \mathbb{1}$, i.e., when $\beta \to 0$, but we adopt the terminological abuse of Ref. [57].

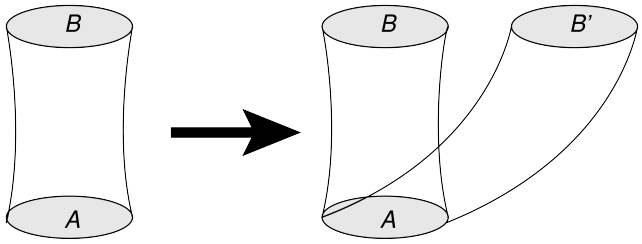

Figure 8.1. Illustration of the black hole cloning thought experiment in the context of the ER=EPR conjecture. If Bob has access to a device that can clone quantum states, he can transform black hole B, which is entangled with A, into two black holes B and B', each connected to A via an ER bridge.

We now consider repeating the experiment in the previous section using entangled black holes instead of qubits, as depicted in Fig. 8.1. Alice and Bob, who live in an asymptotically-AdS spacetime, are each given access to a Schwarzschild black hole, with the two black holes, labeled A and B respectively, maximally entangled and therefore connected by a nontraversable wormhole. If Bob now clones all the degrees of freedom on his stretched horizon [124], he is left with two black holes B and B', each of which is connected by an ER bridge to Alice's black hole. That is, cloning is dual to change of spacetime topology under ER=EPR.

8.4 Changing Spacetime Topology

We now turn to the question of whether the double-wormhole geometry of Fig. 8.1 suffers from any inconsistencies in general relativity. Throughout this chapter, we assume that the Einstein equations hold and that the spacetime can be well described by a semiclassical geometry (which corresponds to a choice of how Bob implements the cloning).

The simplest interpretation of the geometry M in Fig. 8.2 is that, since horizon pairs AB and AB' are each in the thermofield double state (8.3.1), the geometries of both wormholes are the same. In this case, the geometry after Bob performs the cloning simply consists of two separate sheets, each a copy of

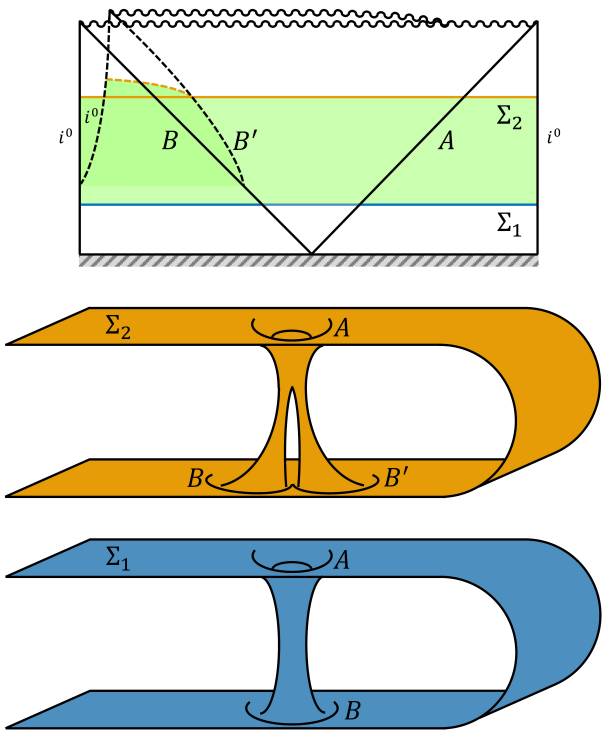

Figure 8.2. Penrose diagram for the topology-change process depicted in Fig. 8.1, with spatial slices Σ_1 (blue) and Σ_2 (orange) shown as embedding diagrams. The spacetime region M (green) is indicated; the compact region K with nontrivial topology is bounded by horizons A, B, and B'. All of the spatial infinities i^0 are identified, as the black holes share the same asymptotically-AdS spacetime. The diagonal stripes at the bottom of the Penrose diagram indicate that the half of the spacetime containing the past horizons is not shown.

the original ER bridge, glued together along horizon A. Note that in this case M contains bifurcate geodesics: any timelike geodesic intersecting horizon A after the cloning occurs will split into two timelike geodesics, one going along the sheet containing B and the other along the sheet containing B'. These timelike bifurcate curves indicate a breakdown of the Hausdorff condition, the requirement that for any two points $x \neq y$, there exist disjoint open sets $X \ni x$ and $Y \ni y$.[8.2] Since the bifurcate timelike curve in question has bounded (being a geodesic, zero) acceleration and moreover the non-Hausdorff boundary of M

[8.2] Bifurcating geodesics imply failure of the Hausdorff condition, but the converse is not necessarily true; see, for example, the discussion of Taub-NUT space in Refs. [68, 225].

(horizon A) is codimension 1, it follows by a theorem of Hájíček [225] that M is not strongly causal. Strong causality is the requirement that for all points $p \in M$ there is an open neighborhood $P \ni p$ such that any timelike curve passing through P does so only once; this is a weaker condition than global hyperbolicity, so the setup depicted in Fig. 8.2 leads, via Hájíček's theorem, to breakdown of Cauchy evolution [230]. Intuitively, this happens because once a timelike curve intersects horizon A it becomes impossible to predict its future. If we wish to avoid immediately abandoning strong causality, we must relax the assumption that the geometry after cloning is merely a two-sheeted copy of the original ER bridge and instead turn to the question of whether the topology change induced by cloning is alone sufficient to guarantee a pathology for a spacetime that remains Hausdorff.

The topology change in question occurs in a localized region of spacetime. Let us define a partial Cauchy surface [223] to be a spacelike slice through the entire spacetime such that any causal (timelike or null) curve intersects the surface at most once. A 3-surface Σ is called externally Euclidean if there exists compact $\Gamma \subset \Sigma$ such that $\Sigma - \Gamma$ is diffeomorphic to Euclidean space minus a 3-ball, i.e., $\Sigma - \Gamma \simeq S^2 \otimes \mathbb{R}$. Given these definitions, we can draw two disjoint externally Euclidean partial Cauchy surfaces Σ_1 and Σ_2, where Σ_1 passes through horizons A and B before the cloning and Σ_2 passes through horizons A, B, and B' after the cloning, as shown in Fig. 8.2. Importantly, Σ_1 and Σ_2 are *not* diffeomorphic, $\Sigma_1 \not\simeq \Sigma_2$. Taking A, B, and B' to be centered on a line on Σ_2 and quotienting by the rotation group $SO(2)$ around this line, $\Sigma_1/SO(2)$ and $\Sigma_2/SO(2)$ are 2-manifolds with genera 1 and 2, respectively, and are therefore not topologically equivalent. The four-dimensional spacetime region whose boundary is $\Sigma_1 \cup \Sigma_2$, called M in Fig. 8.2, is externally Lorentzian: there exists a compact manifold K such that $M - K \simeq S^2 \otimes \mathbb{R} \otimes [0, 1]$, a timelike foliation of

spacelike slices $S^2 \otimes \mathbb{R}$. Then Geroch's topology-conservation theorem [221–223] implies that, since $\Sigma_1 \not\simeq \Sigma_2$, M must contain a CTC.

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

$$k^\alpha k^\beta k_{[\mu} R_{\nu]\alpha\beta[\rho} k_{\sigma]} \neq 0. \tag{8.4.1}$$

This means that every timelike or null geodesic experiences a tidal force at some point.[8.3] Then Tipler's topology-conservation theorem [223, 224] implies that since $\Sigma_1 \not\simeq \Sigma_2$, the NEC[8.4] must fail. That is, the topology change dual to cloning under ER=EPR implies that there must exist fields in the theory for which one can arrange an energy-momentum tensor $T_{\mu\nu}$ such that

$$T_{\mu\nu} k^\mu k^\nu < 0 \tag{8.4.2}$$

along some null vector k^μ.

Although violations of the NEC (see also Ref. [231]) have been shown to occur at a quantum level [232], it has not been shown that such violation is sufficient to allow unusual semiclassical gravitational behavior [57, 129]. However, the NEC violation in the present thought experiment implies macroscopic topology change that results from Bob's cloning procedure with, for example, astrophysical-scale entangled black holes. We conclude that violation of the no-cloning theorem is dual under ER=EPR to topology change and problems

[8.3] If the spacetime under consideration has some special symmetry allowing Eq. (8.4.1) to fail for some geodesic, we can enforce the generic condition by simply adding gravitational waves (that is, nonzero Weyl tensor) sufficiently weak to avoid nonnegligible back-reaction on the rest of our argument.

[8.4] While Ref. [223] states the theorem in terms of the weak energy condition, this can be strengthened to the NEC as stated in Ref. [224].

with causality, leading to CTCs (by Geroch's theorem) or strong violation of the NEC (by Tipler's theorem).

It is worth noting that the topology theorems do not rule out sensible processes like black hole pair production in the context of ER=EPR. If we consider entanglement as a conserved quantity [233], then creation of a pair of entangled black holes does not change the topology, as the ER bridge between them is formed in ER=EPR from the Planckian wormholes connecting the entangled vacuum. Moreover, the process of black hole pair creation is not well described semiclassically, so our results do not apply in that case; in contrast, the cloning process examined in this chapter can be treated in the setting of semiclassical geometry. Unlike pair production, cloning *does* violate the axioms of the topology-conservation theorems precisely because it involves non-unitarily creating entanglement (and therefore wormholes) that did not previously exist.

8.5 Wormholes and Causality

We have shown that violation of the no-cloning theorem is dual under ER=EPR either to immediate breakdown of Cauchy evolution or to severe violation of the NEC [Eq. (8.4.2)]. The latter implies the condition that allows for stabilization of wormholes; specifically, one must have violation of the *averaged* NEC [128, 129]. That is, a traversable ER bridge requires

$$\int_0^\infty T_{\mu\nu} k^\mu k^\nu \mathrm{d}\lambda < 0 \tag{8.5.1}$$

for some null geodesics with affine parameter λ and tangent vector k^μ. Ref. [129] exhibits a construction of a traversable ER bridge that just satisfies Eq. (8.5.1) within the wormhole while retaining non-negative total energy.

The connection between wormhole stabilization and the NEC is highly

relevant in the context of the ER=EPR correspondence, as the argument in Ref. [57] regarding the impossibility of using wormholes (and by duality, entanglement) to transmit information is critically dependent on the ER bridges pinching off too quickly to allow for signal traversal [127]; a stabilized wormhole would falsify this line of reasoning. Said another way, violation of the NEC plus the existence of wormholes leads to traversable wormholes, which would lead to causality violation. In particular, given a traversable ER bridge, one can immediately form a causal paradox (i.e., a closed signal trajectory) by simply moving the wormhole mouths far apart and giving them a small relative boost [2, 129]. The connection between topology change and causality violation in the gravitational sector is now explicit and is satisfyingly analogous to the connection between unitarity/no-cloning and causality on the quantum mechanical side of the ER=EPR duality.

8.6 Perspectives for Future Work

As we have seen, spacetime topology change leads inexorably to violation of causality, via either breakdown of the Hausdorff condition or creation of traversable wormholes. Using ER=EPR to translate this result to quantum mechanics, we find that violation of the axioms of the topology-conservation theorems is dual to violation of monogamy of entanglement (i.e., cloning) and the existence of wormholes is dual to the existence of entanglement entropy. The logical flow of our reasoning is:

$$\begin{array}{ccccccc} \mathrm{C} & \& & \exists\,\mathrm{QE} & \Longrightarrow & \mathrm{SLS} \\ \updownarrow & & \updownarrow & & \updownarrow \\ \Delta\mathrm{T} & \& & \exists\,\mathrm{WH} & \overset{\cancel{\mathrm{NEC}}}{\Longrightarrow} & \mathrm{TWH} \\ \Downarrow & & & & \\ \left(\cancel{\mathrm{NEC}}\ \&\ \exists\ \mathrm{CTCs}\right) & || & \cancel{SC}. & & \end{array} \tag{8.6.1}$$

Here, C denotes "quantum cloning," "QE" quantum entanglement, "SLS" superluminal signaling, "T" topology, "WH" wormholes, "TWH" traversable wormholes, and "SC" strong causality. The single-lined arrows in Eq. (8.6.1) indicate duality of specific statements under ER=EPR, double-lined arrows indicate logical implication, and strikethroughs indicate violation.

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

A promising avenue for future research is the investigation of whether other no-go theorems in quantum mechanics and gravity neatly correspond under ER=EPR. The no-deleting theorem corresponds to the topology theorem in exactly the same way as the no-cloning theorem, while the no-communication theorem is equivalent to the assertion of nontraversability of wormholes. On the gravity side, violation of Hawking's area theorem, i.e., the generalized second law of thermodynamics, requires either breakdown of cosmic censorship or of the null energy condition [234], the latter allowing wormhole traversal [129]. In ER=EPR, this corresponds to violation of the no-communication theorem [57] and, in AdS/CFT, would correspond to violation of unitarity in the dual CFT state of Eq. (8.3.1) [2]. Whether all known gravitational or quantum mechanical no-go theorems map onto each other in this way is a fascinating open question. More generally, the connections among infrared constraints on ultraviolet physics, such as unitarity and causality [2, 31, 134, 235], will continue to play an important role in understanding quantum gravity.

Chapter 9
Wormhole and Entanglement (Non-)Detection in the ER=EPR Correspondence

The recently proposed ER=EPR correspondence postulates the existence of wormholes (Einstein-Rosen bridges) between entangled states (such as EPR pairs). Entanglement is famously known to be unobservable in quantum mechanics, in that there exists no observable (or, equivalently, projector) that can accurately pick out whether a generic state is entangled. Many features of the geometry of spacetime, however, are observables, so one might worry that the presence or absence of a wormhole could identify an entangled state in ER=EPR, violating quantum mechanics, specifically, the property of state-independence of observables. In this note, we establish that this cannot occur: there is no measurement in general relativity that unambiguously detects the presence of a generic wormhole geometry. This statement is the ER=EPR dual of the undetectability of entanglement.

This chapter is from Ref. [8], N. Bao, J. Pollack, and G. N. Remmen, "Wormhole and Entanglement (Non-)Detection in the ER=EPR Correspondence," JHEP ***11*** *(2015) 126,* arXiv:1509.05426 [hep-th].

9.1 Introduction

Black holes are the paradigmatic example of a system where both field-theoretic and gravitational considerations are important. Black hole thermodynamics and the area theorem [210, 211] already provided a relationship between entanglement and geometry, while the classic black hole information paradox

[214] and its potential resolution via complementarity [124, 236, 237] pointed at the subtlety of the needed quantum mechanical description. In the last few years, the firewall paradox [215] has heightened the tension between these two descriptions, prompting a number of proposals to modify the standard picture to a greater or lesser extent.

One set of proposals [238–242] modifies quantum mechanics to allow for state-dependence of the black hole horizon, so that an infalling observer does not encounter a firewall even though the state of the black hole horizon can be written as a superposition of basis states that each have high energy excitations [243]. In order to avoid this problem, the presence or absence of a black hole firewall must become a nonlinear observable, contrary to standard quantum mechanics. In a recent paper, Marolf and Polchinski [244] have pointed out that this nonlinearity cannot be "hidden"; if it is strong enough to remove a firewall from a generic state, it must lead to violations of the Born Rule visible from outside the horizon.

In this chapter, we consider a different idea inspired by the firewall paradox, the ER=EPR correspondence [57], which asserts the existence of an exact duality between Einstein-Podolsky-Rosen (EPR) pairs, i.e., entangled qubits, and Einstein-Rosen (ER) bridges [55, 126, 127], i.e., nontraversable wormholes. This duality is supposed to be contained within quantum gravity, which is in itself meant to be a *bona fide* quantum mechanical theory in the standard sense. The ER=EPR proposal is radical, but it is not obviously excluded by either theory or observation, and indeed has passed a number of nontrivial checks [217, 218, 245–247]; if true, it has the potential to relate previously unconnected statements about entanglement and general relativity in a manner reminiscent of the AdS/CFT correspondence [52–54]. In a previous paper [7], we pointed out that in ER=EPR the no-cloning theorem is dual to the general

relativistic no-go theorem for topology change [221, 223]; violation on either side of the duality, given an ER bridge (two-sided black hole), would lead to causality violation and wormhole traversability.

In light of the result of Ref. [244], one might be worried that ER=EPR is in danger. It is well known that entanglement is not an observable, in the sense we will make precise below; we cannot look at two spins and determine whether they are in an arbitrary, unspecified entangled state with one another. Yet ER=EPR implies that the two spins are connected by a wormhole, so that the geometry of spacetime differs according to whether or not they are entangled. If this difference in geometry could be observed, entanglement would become a (necessarily nonlinear) observable as well and the laws of quantum mechanics would be violated, contrary to the assumption that the latter are obeyed by quantum gravity.

In this chapter, we show that ER=EPR does not have this issue. Unlike the modifications to quantum mechanics considered by Ref. [244], wormhole geometry *can* be hidden. In particular, we show that in general relativity no measurement can detect whether the interior of a black hole has a wormhole geometry. More precisely, observers can check for the presence or absence of specific ER bridge configurations, but there is no projection operator (i.e., observable) onto the entire family of wormhole geometries, just as (and, in ER=EPR, for the same reason that) there is no projection operator onto the family of entangled states.

The remainder of this chapter is organized as follows. We first review the basic quantum mechanical statement that entanglement is not an observable. Next we introduce the maximally-extended AdS-Schwarzschild geometry in general relativity and, using AdS/CFT, on the CFT side. As a warmup, we first show that no single observer can detect the presence of a wormhole geometry.

We then turn to more complicated multiple-observer setups and show, as desired, that they are unable to detect the presence of nontrivial topology in complete generality.

9.2 Entanglement is Not an Observable

The proof that one cannot project onto a basis of entangled states [248] proceeds as follows. Assume the existence of a complete basis set of entangled states $|\psi_{E_i}\rangle$, distinct from the basis set of all states. A projection onto this basis could be written in the form

$$\hat{P}_E = \sum_i |\psi_{E_i}\rangle \langle\psi_{E_i}| . \tag{9.2.1}$$

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

$$|\psi_{E_i}\rangle = \sum_{j \in B_i} |\psi_j\rangle \tag{9.2.2}$$

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

9.3 Setup

We consider the maximally-extended AdS-Schwarzschild geometry [249, 250], which, following Ref. [57], we will interpret as an Einstein-Rosen bridge connecting two black holes. The metric for the AdS-Schwarzschild black hole in D spacetime dimensions is [251, 252]

$$\mathrm{d}s^2 = -f(r)\mathrm{d}t^2 + \frac{\mathrm{d}r^2}{f(r)} + r^2\mathrm{d}\Omega_{D-2}^2, \tag{9.3.1}$$

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

$$f(r) = 1 - \frac{16\pi G_D M}{(D-2)\Omega_{D-2}r^{D-3}} + \frac{r^2}{L^2}, \tag{9.3.2}$$

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

$$\begin{aligned}
&\text{(I)} && U = -e^{-f'(r_{\mathrm{H}})u/2} && V = e^{f'(r_{\mathrm{H}})v/2} \\
&\text{(II)} && U = e^{-f'(r_{\mathrm{H}})u/2} && V = e^{f'(r_{\mathrm{H}})v/2} \\
&\text{(III)} && U = e^{-f'(r_{\mathrm{H}})u/2} && V = -e^{f'(r_{\mathrm{H}})v/2} \\
&\text{(IV)} && U = -e^{-f'(r_{\mathrm{H}})u/2} && V = -e^{f'(r_{\mathrm{H}})v/2}.
\end{aligned} \tag{9.3.3}$$

Regions I through IV are depicted in Fig. 9.1 and define the maximally-extended AdS-Schwarzschild black hole geometry. Defining $T = (U+V)/2$ and $X = (V-U)/2$, the horizon is located at $T = \pm X$, that is, at $UV = 0$, while the singularity is located at $T^2 - X^2 = 1$. The one-sided AdS black hole occupies Region I and half of Region II, i.e., $V > 0, X > 0$. In these coordinates, the metric becomes

$$\begin{aligned}
\mathrm{d}s^2 &= -\frac{4|f(r)|e^{-f'(r_{\mathrm{H}})r^*}}{[f'(r_{\mathrm{H}})]^2}\mathrm{d}U\mathrm{d}V + r^2\mathrm{d}\Omega_{D-2}^2 \\
&= \frac{4|f(r)|e^{-f'(r_{\mathrm{H}})r^*}}{[f'(r_{\mathrm{H}})]^2}(-\mathrm{d}T^2 + \mathrm{d}X^2) + r^2\mathrm{d}\Omega_{D-2}^2,
\end{aligned} \tag{9.3.4}$$

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

$$UV = T^2 - X^2 = \pm e^{f'(r_{\mathrm{H}})r^*}, \tag{9.3.5}$$

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

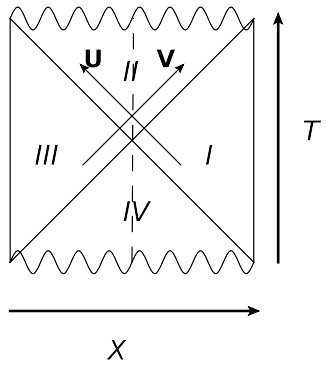

Figure 9.1. The maximally-extended AdS-Schwarzschild geometry, with Kruskal-Szekeres coordinates T, X and lightcone coordinates U, V indicated. Of course, the singularity actually appears as a hyperbola in T, X. This diagram is a conformally-transformed sketch to indicate the general relationship among the coordinates; see Ref. [253] for more discussion. Regions I through IV are defined by Eq. (9.3.3).

We now turn to the CFT interpretation of the geometry. In the Maldacena and Susskind proposal of ER=EPR [57], it is pointed out that, in AdS/CFT, the state $|\psi(t)\rangle$ in the CFT corresponds at different times to different causal diamonds in the eternal, maximally-extended AdS-Schwarzschild geometry. Different spatial slices through a given causal diamond that intersect the boundaries at fixed points are related to each other by the Wheeler-deWitt equation in the bulk. If one is outside a black hole in AdS, without knowing a priori which time slice one is on, then the different $|\psi(t)\rangle$ are simply a one-parameter family of states $|\psi_\alpha\rangle$, where α has replaced t and is now just the label for the state of the CFT at time $t = 0$; all of the $|\psi_\alpha\rangle$ describe pairs of black holes containing some sort of ER bridge. The various geometries are shown in Fig. 9.2.

Famously, the maximally-extended AdS black hole can be described on the CFT side of the AdS/CFT correspondence by the thermofield double state of two noninteracting large-N CFTs on the boundary sphere. We take the interpretation [57] of the state as two entangled black holes that both evolve forward in time, that is,

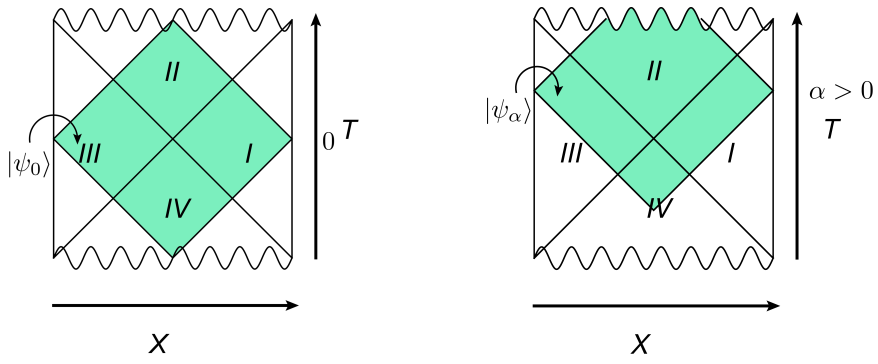

Figure 9.2. (left) The state $|\psi_0\rangle$, corresponding to a wormhole geometry where the ER bridge intersects the boundary at $T = 0$. (right) The family of states $|\psi_\alpha\rangle$, $\alpha > 0$, for which the ER bridge intersects the boundary at $T > 0$.

$$|\psi(t)\rangle = \frac{1}{\sqrt{Z}} \sum_n e^{-\beta E_n/2} e^{-2iE_n t} \, |\bar{n}\rangle_L \otimes |n\rangle_R \,, \tag{9.3.6}$$

where $|n\rangle_L$ and $|n\rangle_R$ are the n^{th} eigenstates of the left and right CFTs, respectively, with eigenvalue E_n, a bar denotes the CPT conjugate, and β is the inverse temperature. We note that the CFT time t in Eq. (9.3.6) is the $r \to \infty$ limit of the Schwarzschild time t that appears in Eq. (9.3.1). By considering the surface of constant Kruskal time T that intersects the $r = \infty$ boundary at Schwarzschild time t, we can instead parameterize the CFT state corresponding to the eternal AdS black hole as $|\psi(T)\rangle$. Equivalently, we can write as $|\psi_T\rangle$ the family of ER bridges indexed by T, which correspond at the fixed Kruskal time $T = 0$ to the CFT state $|\psi(T)\rangle$. The black hole described by $|\psi_{T_0}\rangle$ is given by the metric (9.3.4) with T replaced with $T - T_0$ in Eq. (9.3.5). The analogous states with two one-sided black holes on the boundary CFTs will be called $|\phi_T\rangle$, where

$$|\phi_t\rangle = \frac{1}{\sqrt{Z}} \left(\sum_m e^{-\beta E_m/2} e^{-iE_m t} \, |\bar{m}\rangle_L \right) \otimes \left(\sum_n e^{-\beta E_n/2} e^{-iE_n t} \, |n\rangle_R \right). \tag{9.3.7}$$

9.4 The Single-Observer Case

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

In general relativity, the nontraversability of wormholes follows immediately from a more fundamental result, the topological censorship theorem [226], which is the statement that in a globally hyperbolic, asymptotically flat spacetime satisfying the null energy condition (NEC), any causal curve from past null infinity to future null infinity is diffeomorphic to an infinite causal curve in topologically trivial spacetime (such as Minkowski space). In other words, no causal observer's worldline can ever probe nontriviality of topology of spacetime.[9.1] Probing the nontrivial topology of an ER bridge simply means passing through the wormhole, which is therefore forbidden given the NEC. In a previous paper [7], we showed that violation of the NEC in ER=EPR necessarily leads to violation of the no-cloning theorem and the breakdown

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

of unitary evolution. Traversable wormholes are therefore also forbidden by quantum mechanics given ER=EPR, as they would correspond to a breakdown of unitarity by allowing superluminal signaling.

The next simplest means of verifying the existence of an ER bridge would be to detect the nontrivial topology of the wormhole without traversing it. In the present context, we see that detecting the nontrivial topology is equivalent under ER=EPR to detecting the existence of entanglement—more precisely, to constructing a linear operator that detects if an unknown state is entangled with anything else. But it is well known that such an operator is forbidden by the linearity of quantum mechanics, as Ref. [244] discusses. Briefly, this is because projection operators cannot project onto a subspace unless that subspace is closed under superposition. An attempt to project onto the set of all entangled states will therefore fail due to the set of all entangled states not being closed under superposition; such a projector will inevitably project onto the entire Hilbert space of all states. On the gravity side, this leads to a result stronger than the nontraversability of wormholes: not only does ER=EPR forbid an observer from traversing wormholes, it forbids an isolated observer from verifying their existence even once inside them.

This result can be straightforwardly verified in general relativity by examining the applicable metrics. Importantly, the metric given in Eqs. (9.3.4) and (9.3.5) for the maximally-extended geometry has several isometries: it is invariant under the exchange $(U, V) \leftrightarrow (-U, -V)$ and also under the exchanges $(T, X) \leftrightarrow (T, -X)$ and $(T, X) \leftrightarrow (-T, X)$. That is, Regions I and II in Fig. 9.1 are the same as Regions III and IV, respectively, and moreover the entire metric is symmetric under spatial (X) or temporal (T) reversal. In particular, the regions present in both this geometry and the one-sided black hole geometry (Region I and half of Region II, i.e., $V > 0, X > 0$) are completely identical in

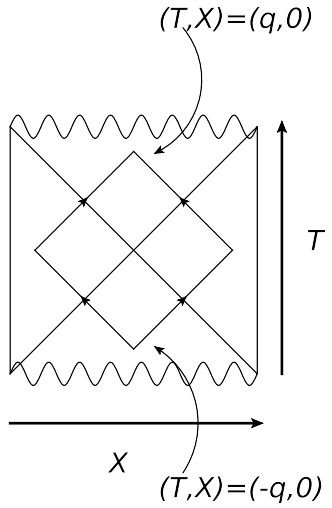

Figure 9.3. The procedure described in the text for detecting a wormhole. Alice and Bob emerge from the white hole portion of the AdS-Schwarzschild geometry, then meet again inside the black hole.

the two cases. It is this property that implies that an observer on a geodesic entering a one-sided black hole cannot distinguish it from a two-sided black hole via any local measurement of curvature.

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

9.5 The Multiple-Observer Case

One can ask the question of whether two (or, for that matter, many) observers can detect the existence of entanglement or, equivalently, of nontrivial topology. The setup of the experiment is as follows. Consider a maximally-extended, eternal AdS-Schwarzschild geometry, as depicted in Fig. 9.3. Allow two observers, Alice and Bob, to initially begin in the white hole portion of the geometry. (We will consider the case of more than two observers later in this section.) Now let

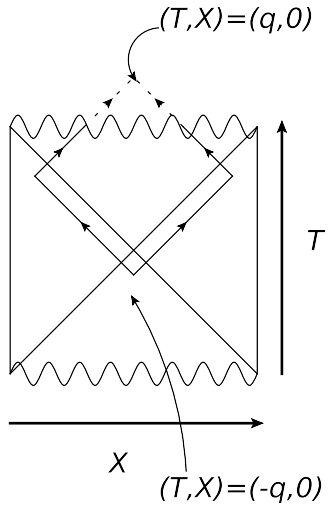

Figure 9.4. Unlike in Fig. 9.3 above, here the geometry is shifted to some $|\psi_\alpha\rangle$ for some sufficiently large $\alpha \neq 0$; Alice and Bob hit the singularity before they can meet and are therefore unable to verify the existence of a wormhole.

the observers exit the white hole[9.2] to the two different asymptotic regions not contained in the black hole. Next, they both jump into their respective black holes and compare notes. In such a way, they could potentially determine if there was entanglement before hitting the singularity.

The problem with this construction is that it doesn't definitively tell the observers if there was entanglement or not. Indeed, Alice and Bob could jump into the $|\psi_0\rangle$ ER bridge at sufficiently late time T that they are unable to communicate (since one or both of them will hit the singularity before being able to do so); equivalently, the geometry could be $|\psi_\alpha\rangle$ for α too large (instead of $|\psi_0\rangle$), as depicted in Fig. 9.4. The same argument that states that no linear operator permits the observers from detecting whether or not there is entanglement precludes this verification procedure from succeeding with

[9.2] We note that the white holes mentioned in our construction are for convenience only; it suffices for Alice and Bob to have communicated at some past time and simply to have moved out of causal contact. Indeed, it is possible for Alice and Bob to both exist in the same asymptotically AdS vacuum, as long as a wormhole exists connecting their locations. It is, however, necessary for them to enter the wormhole in order to attempt to detect information regarding the entanglement in this picture.

probability 1. But how is it possible to reconcile the fact that Alice and Bob can sometimes verify the existence of an ER bridge with the impossibility of projecting onto a generic family of states?

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

However, this thought experiment does not require the existence of a projection operator onto the entire family S_q. Instead, after their geodesics cross, Alice and Bob can actually determine in which of the $|\psi_\alpha\rangle$ they are. All null geodesics from the horizon to the singularity are isomorphic and experience the same pattern of values of the curvature tensor on the way in. That is, a family of null geodesics with, e.g., constant $V = V_0$ can be labeled by the time t at which they cross a surface at fixed proper distance from the horizon in Region I, which is the only difference among the geodesics; since the metric (9.3.1) is independent of t, all of these geodesics experience the same inward journey. Hence, before meeting Bob inside Region II, there is no distinguishing event by which Alice can measure α. However, the value of the Riemann tensor

at the moment Alice's and Bob's geodesics cross is unique for each $|\psi_\alpha\rangle$.

In particular, at the moment their geodesics cross, Alice and Bob can measure the tidal forces acting in their local Lorentz frames by computing some component of the Weyl tensor. At $(T, X) = (q, 0)$, Eq. (9.3.5) implies that, in Region II, r^* and hence r is a monotonic function of $(q - \alpha)$. (Recall that for $|\psi_\alpha\rangle$ all equations describing the metric are shifted by $T \to T - \alpha$.) Let us define a local Lorentz frame in coordinates $(\hat{t}, \hat{r}, \hat{\theta})$, where $\hat{\theta}$ is the orthonormal coordinate in the $D - 2$ angular directions. When their paths cross, Alice and Bob can measure the $\hat{r}\hat{\theta}\hat{r}\hat{\theta}$ component of the Weyl tensor, which is

$$W_{\hat{r}\hat{\theta}\hat{r}\hat{\theta}} = -\frac{1}{L^2} - \left(\frac{D-3}{D-2}\right)\frac{8\pi G_D M}{\Omega_{D-2} r^{D-1}}. \tag{9.5.1}$$

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

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

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

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

9.6 Conclusions

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

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

Many options are available for future investigation. The ER=EPR corre-

spondence has been subjected to some tests [7, 217, 218, 245–247], but the challenge of seeing the duality between wormholes and any arbitrary form of quantum entanglement remains, as does the very definition of what is meant by a "wormhole" in ER=EPR for theories without a weakly-coupled holographic gravity dual. Other open issues include the investigation of whether firewalls are truly nongeneric in ER=EPR [55] and whether the correspondence can be concretely realized outside of asymptotically AdS spacetime. The answers to these questions and others will likely provide important insight in future investigations in the connections between entanglement and spacetime geometry.

Chapter 10
Entanglement Conservation, ER=EPR, and a New Classical Area Theorem for Wormholes

We consider the question of entanglement conservation in the context of the ER=EPR correspondence equating quantum entanglement with wormholes. In quantum mechanics, the entanglement between a system and its complement is conserved under unitary operations that act independently on each; ER=EPR suggests that an analogous statement should hold for wormholes. We accordingly prove a new area theorem in general relativity: for a collection of dynamical wormholes and black holes in a spacetime satisfying the null curvature condition, the maximin area for a subset of the horizons (giving the largest area attained by the minimal cross-section of the multi-wormhole throat separating the subset from its complement) is invariant under classical time evolution along the outermost apparent horizons. The evolution can be completely general, including horizon mergers and the addition of classical matter satisfying the null energy condition. This theorem is the gravitational dual of entanglement conservation and thus constitutes an explicit characterization of the ER=EPR duality in the classical limit.

This chapter is from Ref. [9], G. N. Remmen, N. Bao, and J. Pollack, "Entanglement Conservation, ER=EPR, and a New Classical Area Theorem for Wormholes," JHEP ***07*** *(2016) 048,* arXiv:1604.08217 [hep-th].

10.1 Introduction

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

It is well known that this seeming non-locality does not lead to violations of causality. It cannot be used to send faster-than-light messages [228] and in fact it is impossible for any measurement to determine whether the state is entangled (see, e.g., Ref. [248]). Similarly, it is impossible to alter the entanglement between a system and its environment (that is, to change the entanglement entropy of the reduced density matrix of the system) by acting purely on the degrees of freedom in the system or by adding more unentangled degrees of freedom. A number of well-established properties, such as monogamy [254] and strong subadditivity [255], constrain the entanglement entropy of subsystems created from arbitrary factorizations of the Hilbert space.

Although entanglement entropy is a fundamental quantity, it is typically very difficult to compute in field theories, where working directly with the reduced density matrix can be computationally intractable, although important progress has been made in certain conformal field theories [256, 257] and more generally along lightsheets for interacting quantum field theories [258].

The AdS/CFT correspondence [52–54], however, allows us to transform many field-theoretic questions to a gravitational footing. In particular, the Ryu-Takayanagi formula [132] equates the entanglement entropy of a region for a state in a conformal field theory living on the boundary of an asymptotically AdS spacetime to the area of a minimal surface with the same boundary as that region in the spacetime corresponding to that CFT state. Using this identification of entropy with area, a number of "holographic entanglement inequalities" have been proven [259, 260], some reproducing and some stronger than the purely quantum mechanical entanglement inequalities.

Motivated in part by AdS/CFT, as well as a number of older ideas in black hole thermodynamics [210, 211] and holography [50, 51, 261], Maldacena and Susskind have recently conjectured [57] an ER=EPR correspondence, an exact duality between entangled states (Einstein-Podolsky-Rosen [58] pairs) and so-called "quantum wormholes," which reduce in the classical general relativistic limit to two-sided black holes (Einstein-Rosen [55] bridges, i.e., wormholes). In a series of recent papers, we have considered the implications of this correspondence in the purely classical regime. In this limit, if the ER=EPR duality holds true, certain statements in quantum mechanics about entangled states should match directly with statements in general relativity about black holes and wormholes [262], with the same assumptions required on both sides. We indeed previously found two beautiful and nontrivial detailed correspondences: the no-cloning theorem in quantum mechanics corresponds to the no-go theorem for topology change in general relativity [7] and the unobservability of entanglement corresponds to the undetectability of the presence or absence of a wormhole [8].

In this chapter, we extend this correspondence to a direct equality between the entanglement entropy and a certain invariant area, which we define, of a

geometry containing classical black holes and wormholes. We follow a long tradition of clarifying general relativistic dynamics using area theorems [68, 263–266], which hold that various areas of interest satisfy certain properties under time evolution. Our strategy is to show that the area in question remains unchanged under dynamics constituting the gravitational analogue of applying tensor product operators to an individual system and its complement. We show that, just as entanglement entropy cannot be changed by acting on the subsystem and its complement separately, this area is not altered by merging pairs of black holes or wormholes or by adding classical (unentangled) matter. The area we consider is chosen to be that of a maximin surface [267, 268] for a collection of wormhole horizons, a time-dependent generalization of the Ryu-Takayanagi minimal area, which again establishes that the entanglement entropy is also conserved under these operations. At least for asymptotically AdS spacetimes, our result constitutes an explicit characterization of the ER=EPR correspondence in the classical limit. Moreover, our theorem is additionally interesting from the gravitational perspective alone, as it constitutes a new area law within general relativity.

This chapter is structured as follows. In Sec. 10.2, we review the simple quantum mechanical fact that entanglement is conserved under local operations. In Sec. 10.3, we define the maximin surface and review its properties. In Sec. 10.4, we prove our desired general relativistic theorem. Finally, we discuss the implications of our result and conclude in Sec. 10.5.

10.2 Conservation of Entanglement

Consider a Hilbert space $\mathcal{H}$ that can be written as a tensor product of two factors $\mathcal{H}_{\mathrm{L}}$ and $\mathcal{H}_{\mathrm{R}}$ to which we will refer as "right" and "left," though they need not have any spatial interpretation. For a state $|\psi\rangle \in \mathcal{H}$, let us define

the reduced density matrix associated with $\mathcal{H}_{\mathrm{L}}$ as $\rho_{\mathrm{L}} = \mathrm{Tr}_{\mathcal{H}_{\mathrm{R}}} |\psi\rangle \langle\psi|$ and use this to define the entanglement entropy between the right and left sides of the Hilbert space:

$$S(L) = S(R) = -\mathrm{Tr}_{\mathcal{H}_{\mathrm{L}}} \rho_{\mathrm{L}} \log \rho_{\mathrm{L}}. \tag{10.2.1}$$

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

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

$$S(L) = -\mathrm{Tr}_{\mathcal{H}_{\mathrm{L}}} U_{\mathrm{L}} \rho_{\mathrm{L}} U_{\mathrm{L}}^{\dagger} \log \left(U_{\mathrm{L}} \rho_{\mathrm{L}} U_{\mathrm{L}}^{\dagger} \right). \tag{10.2.2}$$

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

$$S_n(L) = -\mathrm{Tr}_{\mathcal{H}_{\mathrm{L}}} c_n U_{\mathrm{L}} \rho_{\mathrm{L}} U_{\mathrm{L}}^{\dagger} \left(\mathbb{1} - U_{\mathrm{L}} \rho_{\mathrm{L}} U_{\mathrm{L}}^{\dagger} \right)^n \tag{10.2.3}$$

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

10.3 The Maximin Surface

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

[132] holds:

$$S(H) = \frac{A_H}{4G\hbar}. \tag{10.3.1}$$

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

This method of computing entanglement entropy on a constant-time slice for static geometries was generalized by the Hubeny-Rangamani-Takayanagi (HRT) proposal [267]. The key insight here was that in general there do not exist surfaces that have minimal area in time, as small perturbations can decrease the area. The new proposal was that the area now scales as the smallest extremal area surface, as opposed to the minimal area. The homology condition mentioned previously remains in this prescription.

The maximin proposal [268] gives an explicit algorithm for the implementation of the HRT prescription. In the following definitions, we will closely follow the conventions used by Wall [268]. We define $C[H, \Gamma]$ to be the codimension-two surface of minimal area homologous to H anchored to ∂H that lies on any complete achronal (i.e., spacelike or null) slice Γ. Note that $C[H, \Gamma]$ can refer to any minimal area surface that exists on Γ. Next, the maximin surface $C[H]$ is defined as any of the $C[H, \Gamma]$ with the largest area when optimized over all achronal surfaces Γ. When multiple such candidate maximin surfaces exist, we refine the definition of $C[H]$ to mean any such surface that is a local maximum as a functional over achronal surfaces Γ. In the HRT proposal, the entanglement of H with its complement in the boundary is given by $S(H) = \text{area}[C[H]]/4G\hbar$.

As an example, for a wormhole geometry in which we are computing the

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

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

$$R_{\mu\nu} k^\mu k^\nu \geq 0, \tag{10.3.2}$$

where k^μ is any null vector and $R_{\mu\nu}$ is the Ricci tensor.[10.1] As HRT is a covariant method of calculating entanglement entropy, the maximin construction is therefore manifestly covariant as well.

Maximin surfaces in general have some further nice properties, proven in Ref. [268]: they have smaller area than the causal surface (the edge of the causal domain of dependence associated with bulk causality), they move monotonically outward as the boundary region increases in size, they obey strong subadditivity, and they also obey monogamy of mutual information, but not necessarily other inequalities that hold for constant-time slices [259, 260, 268]:

$$\begin{gathered} S(AB) + S(BC) \geq S(B) + S(ABC), \\ S(AB) + S(BC) + S(AC) \geq S(A) + S(B) + S(C) + S(ABC) \end{gathered} \tag{10.3.3}$$

for disjoint regions A, B, and C. The above statements are all proven in detail for maximin surfaces in Ref. [268].

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

10.4 A Multi-Wormhole Area Theorem

We are now ready to find the gravitational statement dual to entanglement conservation. Let us take as our spacetime M the most general possible setup to consider in the context of the ER=EPR correspondence: an arbitrary, dynamical collection of wormholes and black holes in asymptotically AdS spacetime. We work in D spacetime dimensions. Throughout this chapter, we will assume that M obeys the null curvature condition (10.3.2). The degrees of freedom associated with the Hilbert space $\mathcal{H} = \otimes_i \mathcal{H}_i$ can be considered to be localized on the union of the stretched horizons, with each horizon comprising one of the $\mathcal{H}_i$ factors. We choose our spacetime setup such that the wormholes are past-initialized, by which we mean that for $t \leq 0$ the wormholes are far apart and the spacetime around the wormholes is in vacuum, with negligible back-reaction. Suppose we arbitrarily divide this system into two subsystems by labeling each horizon as "left" or "right." The left and right Hilbert spaces factorize as $\mathcal{H}_{\rm L} = \otimes_i \mathcal{H}_{{\rm L},i}$ and $\mathcal{H}_{\rm R} = \otimes_i \mathcal{H}_{{\rm R},i}$, where $H_{{\rm L(R)},i}$ contains the degrees of freedom associated with horizon i in the left (right) set. Now, some of the black holes in the left subset may be entangled with each other and so be described by ER bridges among the left set. A similar statement applies to the right set. Importantly, there may be horizons in the left set entangled with horizons in the right set, describing ER bridges across the left/right boundary. For the sake of tractability, we consider horizons that are only pairwise entangled and that begin in equal-mass pairs in the asymptotically AdS spacetime; this stipulation can be made without loss of generality provided we consider black holes smaller than the AdS length and do not consider changes to the asymptotic structure of the spacetime (see, e.g., Ref. [269]). (To treat wormholes with mouths of unequal masses, we could start in an equal-mass

configuration and add matter into one of the mouths.) We thus take any two horizons i and j that are entangled to be in the thermofield double state at $t = 0$,

$$\Pi_i \Pi_j \left|\psi\right\rangle (t=0) = \left|\psi_{i,j}\right\rangle (t=0) = \frac{1}{\sqrt{Z}} \sum_n e^{-\beta E_n/2} \left|\bar{n}\right\rangle_i \otimes \left|n\right\rangle_j , \qquad (10.4.1)$$

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

Let us define a time slicing of the spacetime M into spacelike codimension-one surfaces Σ_t parameterized by a real number t that smoothly approaches the standard AdS time coordinate in the limit of spacelike infinity, where the metric is asymptotically AdS. The Σ_t are chosen to pass through the wormholes without coordinate singularities along the horizon (cf. Kruskal coordinates); see Fig. 10.1 for an example geometry. For the wormholes spanning the left and right subsets, we write as L_i and R_i the null codimension-one surfaces that form the outermost left and right apparent horizons, respectively, and define $L = \cup_i L_i$ and $R = \cup_i R_i$. Note that, since new apparent horizons can form outside of the initial apparent horizons, L_i and R_i are each not necessarily connected, but are the piecewise-connected union of the outermost connected components of the apparent horizons. On a given spacelike slice, an apparent horizon is a boundary between regions in which the outgoing orthogonal null congruences are diverging (untrapped) or converging (trapped) [68]. Of course, the indexing i may become redundant if horizons merge among the L_i or R_i. Let us define the restriction of the outermost apparent horizons to the constant-time slice Σ_t as the spacelike codimension-two surfaces $L_{t,i} = L_i \cap \Sigma_t$ and $R_{t,i} = R_i \cap \Sigma_t$ and similarly $L_t = L \cap \Sigma_t$ and $R_t = R \cap \Sigma_t$. Without loss of generality, we will use the initial spatial separation of the wormholes along with

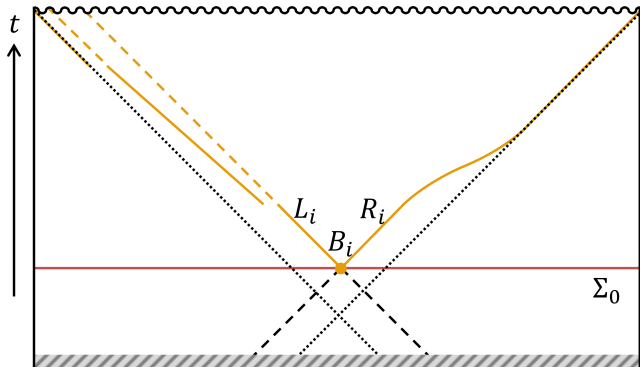

Figure 10.1. Penrose diagram, for an example spacetime M, of a slice through a particular wormhole i joining a left and right horizon. (Showing the full geometry would require a multi-sheeted Penrose diagram to accommodate the multiple wormholes.) The spacelike codimension-one surface Σ_0 is shown in burgundy. The initial bifurcation codimension-two surface B_i is illustrated by the orange dot. Apparent horizons are denoted by the orange lines, with the outermost apparent horizons L_i and R_i being the solid lines. For $t \leq 0$, the setup is past-initialized and the metric is given to good approximation by the eternal black hole in AdS, where the past event horizon of the white hole is indicated by the dashed black lines. The dotted black lines denote the future event horizon of M. As the spacetime at negative t is known, we do not show the entire Penrose diagram in this region, as indicated by the diagonal gray lines.

diffeomorphism invariance to choose the Σ_t and the parameterization of t such that Σ_0 intersects the codimension-two bifurcation surfaces $B_i \equiv L_{0,i} = R_{0,i}$ at which all the wormholes have zero length. The past-initialization condition then means that the wormholes are far apart in the white hole portion of the spacetime, which corresponds to $t \leq 0$. Throughout this chapter, we will assume that $M \cup \partial M$ is globally hyperbolic; equivalently [230], we will assume that the closure of Σ_0 is a Cauchy surface for $M \cup \partial M$.

Now, for each $t > 0$, let us define a D-dimensional region of spacetime W_t as the union over all achronal surfaces with boundary $L_t \cup R_t$; that is, W_t is the causal diamond associated with $L_t \cup R_t$. A single wormhole has topology $S^{D-2} \otimes \mathbb{R}$ when restricted to Σ_t. The initial spacetime W_0 is special: it is a codimension-two surface that is just the union over all the B_i, with topology $(S^{D-2})^{\otimes N}$, where N is the number of wormholes connecting the left and right

subsets.

For a given W_t, let us define a slicing of W_t, parameterized by α, with achronal codimension-one surfaces $\Gamma_t(\alpha)$, where the boundary of $\Gamma_t(\alpha)$ is anchored at $L_t \cup R_t$ for all α and where α increases monotonically as we move from the past to the future boundary of W_t. Now, we can imagine slicing $\Gamma_t(\alpha)$ into codimension-two surfaces and write as $C_t(\alpha)$ the surface with minimal area [i.e., the minimal cross-sectional area of $\Gamma_t(\alpha)$]; see Fig. 10.2. We can now define the maximin surface C_t for W_t as a surface for which the area of $C_t(\alpha)$ attains its maximum under our achronal slicing $\Gamma_t(\alpha)$, maximized over all possible such slicings. That is, C_t is a codimension-two surface with the maximum area, among the set of the surfaces of minimal cross-sectional area, for all achronal slices through W_t.

The main result that we will prove is that the area of the maximin surface C_t is actually independent of t, equaling just the sum of the areas of the initial bifurcation surfaces B_i.[10.2] In most cases, the maximin surface C_t will actually be the union of the initial bifurcation surfaces B_i, independent of t. In other words, the maximin area is invariant among all of the different causal diamonds W_t. Interpreting the area of the maximin surface as an entropy, this is the gravitational analogue of entanglement conservation. We will first prove a few intermediate results.

Proposition 10.1. *The area of the maximin surface C_t is upper bounded by the sum of the areas of the initial bifurcation surfaces B_i.*

Proof. Consider the rightward outgoing orthogonal null congruence $\tilde{B}_i$, a null codimension-one surface starting on B_i and satisfying the geodesic equation.

[10.2] In Ref. [270] it was shown for the special cases of the Schwarzschild-AdS and the single, symmetric, Vaidya-Schwarzschild-AdS geometries that the initial bifurcation surface is the extremal surface in the HRT prescription. Our theorem in this chapter generalizes this result to an arbitrary, dynamical, multi-wormhole geometry in asymptotically AdS spacetime that is past-initialized and that obeys the null curvature condition.

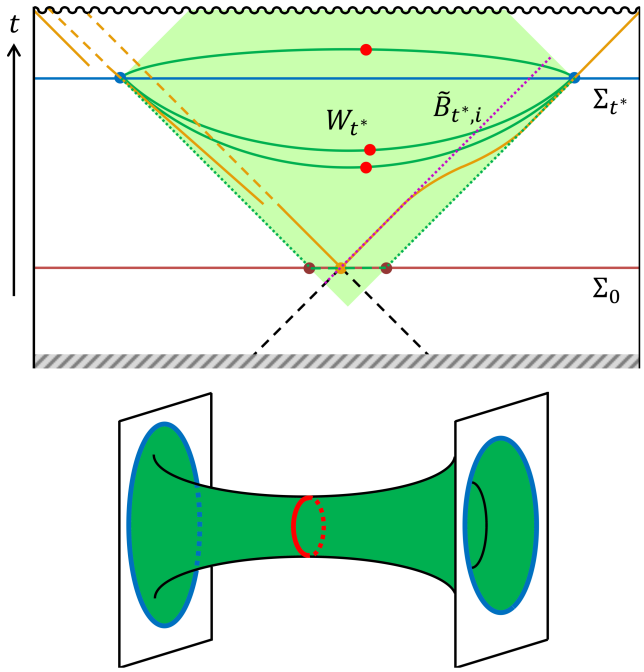

Figure 10.2. Penrose diagram (top), for the example geometry of Fig. 10.1, of the segment of the region W_{t^*} (green shading), for some t^*, that passes through a particular wormhole i joining a left and right horizon. The apparent horizons (orange lines, with solid lines for the outermost apparent horizons L_i and R_i), bifurcation surface B_i (orange dot), spacelike codimension-one surface Σ_0 (burgundy line), and past event horizons for the white hole (dashed black lines) are illustrated as in Fig. 10.1. The spacelike codimension-one surface Σ_{t^*} is shown as a blue line. The purple dotted line denotes the truncated null surface $\tilde{B}_{t^*,i}$ formed from the rightward outgoing orthogonal null congruence $\tilde{B}_i$ originating on B_i, used in Proposition 10.1. The codimension-two boundaries of W_{t^*} along wormhole i, $L_{t^*,i}$ and $R_{t^*,i}$, are indicated by the blue dots. The achronal codimension-one surfaces $\Gamma_{t^*}(\alpha)$ foliating W_{t^*} are indicated within wormhole i by the green lines; the codimension-two surfaces $C_{t^*}(\alpha)$ of minimal area for some slices $\Gamma_{t^*}(\alpha)$ are indicated within wormhole i by red dots. The particular surface $\Gamma_{t^*}(0)$, constructed in Eq. (10.4.8), is shown (for the portion restricted to wormhole i) by the dashed and dotted green lines, corresponding to $\Sigma_0 \cap W_{t^*}$ (the horizontal section) and $M_+ \cap \dot{J}^-[\Sigma_{t^*} \backslash W_{t^*}] = \tilde{L} \cup \tilde{R}$ (the diagonal sections), respectively. The burgundy dots denote the pieces of $\tilde{L}_0$ and $\tilde{R}_0$ in the vicinity of wormhole i. The embedding diagram (bottom) shows a particular slice $\Gamma_{t^*}(\alpha)$ through W_{t^*} for some α, where, as in the Penrose diagram, the codimension-two boundaries $L_{t^*,i}$ and $R_{t^*,i}$ are shown in blue and the surface $C_{t^*}(\alpha)$ of minimal cross-sectional area, restricted to wormhole i, is shown in red.

Choosing some particular t^* arbitrarily, we truncate the null geodesics generating $\tilde{B}_i$ whenever a caustic is reached or when they intersect either the future singularity or the future null boundary of W_{t^*}; we further extend the

null geodesics into the past until they intersect the past null boundary of W_{t^*}. We will hereafter write the truncated null surface as $\tilde{B}_{t^*,i}$. Let λ be an affine parameter for $\tilde{B}_{t^*,i}$ that increases toward the future and vanishes on B_i; let us write $\tilde{B}_{t^*,i}(\lambda)$ for the spatial codimension-two surface at fixed λ. The rotation $\hat{\omega}_{\mu\nu}$ in a space orthogonal to the tangent vector $k^\mu = (\mathrm{d}/\mathrm{d}\lambda)^\mu$ satisfies [271]

$$\frac{\mathrm{D}\hat{\omega}_{\mu\nu}}{\mathrm{d}\lambda} = -\theta\hat{\omega}_{\mu\nu}, \tag{10.4.2}$$

where $\theta = \nabla_\mu k^\mu$ is the expansion. Since θ vanishes on B_i, $\hat{\omega}_{\mu\nu}$ vanishes identically on $\tilde{B}_{t^*,i}$. The Raychaudhuri equation is therefore

$$\frac{\mathrm{d}\theta}{\mathrm{d}\lambda} = -\frac{1}{D-2}\theta^2 - \hat{\sigma}_{\mu\nu}\hat{\sigma}^{\mu\nu} - R_{\mu\nu}k^\mu k^\nu, \tag{10.4.3}$$

where $\hat{\sigma}_{\mu\nu}$ is the shear and $R_{\mu\nu}$ is the Ricci tensor. We note that if the null curvature condition (10.3.2) is satisfied, then θ is non-increasing, as $\hat{\sigma}_{\mu\nu}\hat{\sigma}^{\mu\nu}$ is always non-negative. Since the apparent horizon consists of marginally outer trapped surfaces (i.e., surfaces for which the outgoing orthogonal null geodesics have $\theta = 0$), it must be either null or spacelike, so any orthogonal null congruence starting on the apparent horizon remains either on or inside the apparent horizon in the future [68]. In particular, $\tilde{B}_{t^*,i} \subset W_{t^*}$.

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

$$\mathrm{area}[\tilde{B}_{t^*,i}(\lambda)] \leq \mathrm{area}[B_i]. \tag{10.4.4}$$

By the past-initialization condition, there are no caustics to the past of B_i. Further, by definition, the wormhole does not pinch off until the singularity

is reached, so some subset of the generators of $\tilde{B}_i$ must extend all the way through W_{t^*} without encountering caustics. Writing $\Gamma_{t^*}(\alpha)$ as a foliation of W_{t^*} by achronal slices, we thus have that $\tilde{B}_{t^*,i}(\lambda) \cap \Gamma_{t^*}(\alpha)$ is never an empty set for all α, i.e., for all λ there exists α such that $\tilde{B}_{t^*,i}(\lambda) \subset \Gamma_{t^*}(\alpha)$. Moreover, we can reparameterize and identify the affine parameters for each i of the $\tilde{B}_{t^*,i}$ such that for each λ there exists α for which $\cup_i \tilde{B}_{t^*,i}(\lambda) \subset \Gamma_{t^*}(\alpha)$; for such α, $\cup_i \tilde{B}_{t^*,i}$ is a complete cross-section of $\Gamma_{t^*}(\alpha)$, possibly with redundancy due to merging horizons. We choose our slicing $\Gamma_{t^*}(\alpha)$ such that there exists some α^* for which $\Gamma_{t^*}(\alpha^*)$ contains the maximin surface C_{t^*} for W_{t^*}, so

$$C_{t^*} = C_{t^*}(\alpha^*) \text{ such that } \text{area}[C_{t^*}(\alpha^*)] = \max_{\alpha} \text{area}[C_{t^*}(\alpha)], \tag{10.4.5}$$

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

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

$$\text{area}[C_{t^*}(\alpha)] \leq \sum_i \text{area}[\tilde{B}_{t^*,i}(\lambda)]. \tag{10.4.6}$$

Putting together Eqs. (10.4.4) and (10.4.6), taking the maximum over λ and α on both sides, applying Eq. (10.4.5), and using the fact that t^* was chosen arbitrarily, we have a t-independent upper bound on the area of the maximin surface C_t:

$$\text{area}[C_t] \leq \sum_i \text{area}[B_i]. \tag{10.4.7}$$

□

Let us now construct a lower bound on the area of the maximin surface C_t. We can do this by examining an achronal codimension-one surface through W_t and computing its minimal cross-sectional area; judiciously choosing the achronal surface optimizes the bound. In particular, for some arbitrary t^*,

consider $\Gamma_{t^*}(0)$ passing through $\cup_i B_i$, where we choose the slicing such that

$$\Gamma_{t^*}(0) = (\Sigma_0 \cap W_{t^*}) \cup \left(M_+ \cap \dot{J}^-[\Sigma_{t^*} \backslash W_{t^*}]\right), \tag{10.4.8}$$

where M_+ is the restriction of M to $t \geq 0$, $J^-[A]$ denotes the causal past of a set A, and the dot denotes its boundary. That is, $\Gamma_{t^*}(0)$ consists of the codimension-one null surfaces forming the $t \geq 0$ portion of the boundary of W_{t^*} towards the past, plus a codimension-one segment of Σ_0 containing $\cup_i B_i$; see Fig. 10.2. Let us label the left and right boundaries of $\Sigma_0 \cup W_{t^*}$ (equivalently, the left and right portions of the intersection of Σ_0 and $\dot{J}^-[\Sigma_{t^*} \backslash W_{t^*}]$) as $\tilde{L}_0$ and $\tilde{R}_0$, respectively.

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

Proposition 10.2. *The minimal cross-sectional area of* $\Sigma_0 \cap W_{t^*}$ *is* $\sum_i \text{area}[B_i]$.

Proof. By the requirement that the wormholes be past-initialized, the metric on Σ_0 is, up to negligible back-reaction, just a number of copies of the metric on the $t = 0$ slice of the single maximally-extended AdS-Schwarzschild black hole; for this metric the $t_{\text{KS}} = 0$ and $t_{\text{S}} = 0$ slices are the same, where t_{KS} is the Kruskal-Szekeres time coordinate and t_{S} is the Schwarzschild time coordinate [8]. Taking the t-slicing to correspond to the Kruskal-Szekeres coordinates in the vicinity of each wormhole, therefore, the metric on $\Sigma_0 \cap W_{t^*}$ is

$$\mathrm{d}s^2_{\Sigma_0 \cap W_t} = \frac{4|f(r)|e^{-f'(r_{\text{H}})r^*}}{[f'(r_{\text{H}})]^2}\mathrm{d}X^2 + r^2\mathrm{d}\Omega^2_{D-2} = \frac{\mathrm{d}r^2}{f(r)} + r^2\mathrm{d}\Omega^2_{D-2}, \tag{10.4.9}$$

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

function $f(r)$ is

$$f(r) = 1 - \frac{16\pi G_D M}{(D-2)\Omega_{D-2} r^{D-3}} + \frac{r^2}{\ell^2}, \tag{10.4.10}$$

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

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

Proposition 10.3. *The cross-sectional area of $M_+ \cap \dot{J}^-[\Sigma_{t^*} \backslash W_{t^*}]$ is nondecreasing towards the future.*

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

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

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

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

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

Proposition 10.4. *The area of C_t is lower bounded by the sum of the areas of the initial bifurcation surfaces B_i.*

Proof. To prove a lower bound on the maximin area, area$[C_{t^*}]$, it suffices to exhibit an achronal surface through W_{t^*} for which the minimal cross-sectional area is equal to the desired lower bound. Such a surface is given by $\Gamma_{t^*}(0)$ in Eq. (10.4.8): by Proposition 10.2, $\sum_i \text{area}[B_i]$ is the minimal cross-sectional area of $\Sigma_0 \cap W_{t^*}$ and, in particular, $\sum_i \text{area}[B_i] \leq \text{area}[\tilde{L}_0] + \text{area}[\tilde{R}_0]$. By Proposition 10.3, the minimal cross-sectional area of $M_+ \cap \dot{J}^-[\Sigma_{t^*} \backslash W_{t^*}]$ is $\text{area}[\tilde{L}_0] + \text{area}[\tilde{R}_0]$. Thus, $\Gamma_{t^*}(0)$ is an achronal slice through W_{t^*} with minimal cross-sectional area equal to $\sum_i \text{area}[B_i]$. □

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

Theorem 10.1. *For the family of spacetime regions W_t defined as the causal diamonds anchored on the piecewise-connected outermost apparent horizons L_t and R_t for an arbitrary set of dynamical, past-initialized wormholes and black holes satisfying the null curvature condition, the corresponding maximin surface C_t dividing the left and right collections of wormholes has an area independent of t, equaling the sum of the areas of the initial bifurcation surfaces for the wormholes linking the left and right sets of horizons.*

Proof. By Proposition 10.1, $\text{area}[C_t] \leq \sum_i \text{area}[B_i]$, while by Proposition 10.4, $\text{area}[C_t] \geq \sum_i \text{area}[B_i]$. Hence,

$$\text{area}[C_t] = \sum_i \text{area}[B_i]. \tag{10.4.11}$$

□

Thus, the maximin surface dividing one collection of wormhole mouths from another has an area that is conserved under arbitrary spacetime evolution and horizon mergers as well as arbitrary addition of matter satisfying the null energy condition. Viewing the maximin surface area as the entanglement entropy associated with the left and right sets of horizons in accordance with the HRT prescription, we have proven a statement in general relativity that is a precise analogue of the statement in Sec. 10.2 of conservation of entanglement under evolution of a state with a tensor product unitary operator.

10.5 Conclusions

The proposed ER=EPR correspondence is surprising insofar as it identifies a generic feature (entanglement) of any quantum mechanical theory with a specific geometric and topological structure (wormholes) in a specific theory with both gravity and spacetime (quantum gravity). Until an understanding is reached of the geometrical nature of the "quantum wormholes" that should be dual to, e.g., individual entangled qubits, it will be difficult to directly establish the validity of the ER=EPR correspondence as a general statement about quantum gravity. In a special limiting case of quantum gravity—namely, the classical limit, which gives general relativity—this task is more tractable. In this chapter, we have provided a general and explicit elucidation of the ER=EPR correspondence in this limit. For a spacetime geometry with an arbitrary set of wormholes and black holes, we have constructed the maximin

area of the multi-wormhole throat separating a subset of the wormholes from the rest of the geometry, the analogue of the entanglement entropy of a reduced density matrix constructed from a subset of the degrees of freedom of a quantum mechanical state. We then proved that the maximin area is unchanged under all operations that preserve the relation between the subset and the rest of the geometry, the equivalent of quantum mechanical operations that leave the entanglement entropy invariant. We have therefore completely characterized the ER=EPR relation in the general relativistic limit: the entanglement entropy and area (in the sense defined above) of wormholes obey precisely the same rules.

In addition to providing an examination of the ER=EPR duality, our result constitutes a new area theorem within general relativity. The maximin area of the wormhole throat is invariant under dynamical spacetime evolution and the addition of classical matter satisfying the null energy condition. The dynamics of wormhole evolution were already constrained topologically (see Ref. [7] and references therein), but this result goes further by constraining them geometrically. Note that throughout this chapter we have worked in asymptotically AdS spacetimes in order to relate our results to a boundary theory using the language of the AdS/CFT correspondence, but our area theorem is independent of this asymptotic choice provided that all of the black holes are smaller than the asymptotic curvature scale.

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

Chapter 11
Consistency Conditions for an AdS/MERA Correspondence

The Multiscale 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 is from Ref. [10], N. Bao, C. Cao, S. M. Carroll, A. Chatwin-Davies, N. Hunter-Jones, J. Pollack, and G. N. Remmen, "Consistency Conditions for an AdS Multiscale Entanglement Renormalization Ansatz Correspondence," Phys. Rev. ***D91*** *(2015) 125036,* arXiv:1504.06632 [hep-th].

11.1 Introduction

The idea that spacetime might emerge from more fundamental degrees of freedom has long fascinated physicists. The holographic principle suggests that a $(D+1)$-dimensional spacetime might emerge from degrees of freedom in a D-dimensional theory without gravity [50, 51]. 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 [52–54]. An intriguing aspect of this duality is the Ryu-Takayanagi formula [132, 272], 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:

$$S(B) = \frac{\text{area}(\tilde{B})}{4G} + \text{corrections}. \tag{11.1.1}$$

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

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

Meanwhile, from a very different perspective, tensor networks have arisen

as a useful way to calculate quantum states in strongly-interacting many-body systems [273]. One significant example is the Multiscale Entanglement Renormalization Ansatz (MERA) [274], which is relevant for critical (gapless) systems, i.e., CFTs. Starting from a simple state in a low-dimensional Hilbert space, acting repeatedly with fixed tensors living on a network lattice produces an entangled wave function for the quantum system of interest; varying with respect to the tensor parameters efficiently computes the system's ground state.

Working "backwards" in the MERA, starting with the ground state and gradually removing entanglement, produces a set of consecutively renormalized quantum states. This process reveals a renormalization direction along the graph, which may be thought of as an emergent radial direction of space. As pointed out by Swingle [275], 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 [275]. 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 [276–279].

In this chapter, 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. 11.3 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. 11.4 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. 11.5 we apply the covariant entropy (Bousso) bound to regions of the bulk lattice. In the most naive 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.

11.2 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.

11.2.1 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. [273]. Detailed analyses of the MERA are given in [274, 280, 281] and references therein.) In this chapter, we restrict our attention to the case $D = 1 + 1$.

The MERA tensor network is shown in Fig. 11.1. The quantum system being modeled by the MERA lives at the bottom of the diagram, henceforth

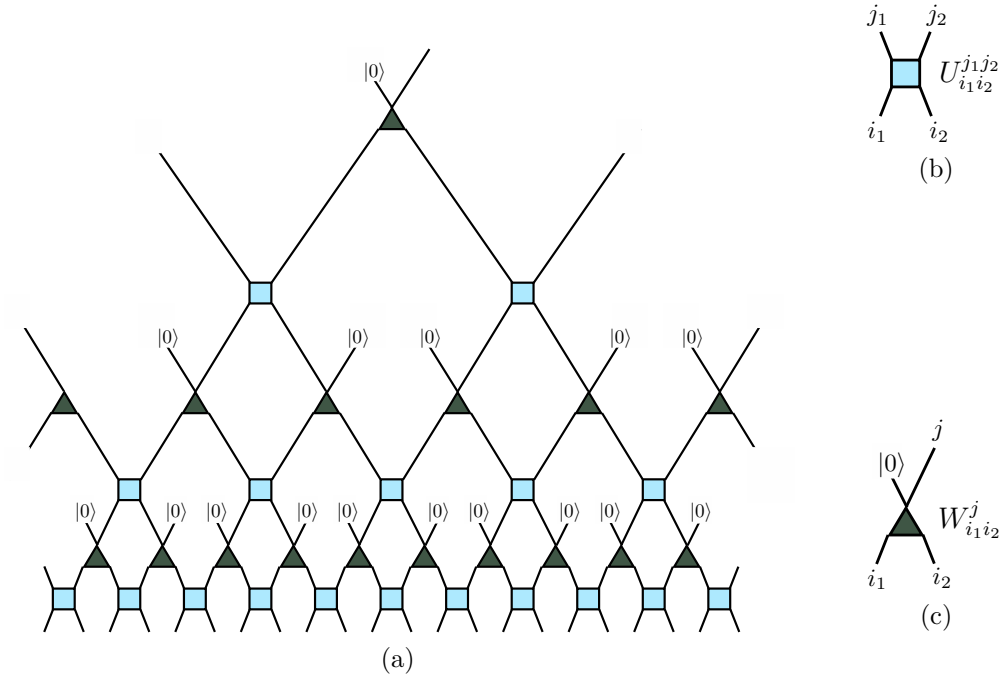

Figure 11.1. (a) Basic construction of a $k = 2$ MERA (2 sites renormalized to 1). (b) The squares represent disentanglers: unitary maps that, from the moving-upward perspective, remove entanglement between two adjacent sites. (c) The triangles represent isometries: linear maps that, again from the moving-upward 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.

"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. 11.1 we have portrayed the case $k = 2$. The squares and triangles are the tensors: multilinear maps between direct products of vector spaces. Each line represents an index i of the corresponding tensor, ranging over values from 1 to the "bond dimension" χ. The boundary Hilbert space $\mathcal{H}_{\text{boundary}} = V^{\otimes \mathcal{N}_{\text{boundary}}}$ is given by

a tensor product of $\mathcal{N}_{\text{boundary}}$ individual spaces V, each of dimension χ. (In principle the dimension of the factors in the boundary could be different from the bond dimension of the MERA, and indeed the bond dimensions could vary over the different tensors. We will assume these are all equal.)

As its name promises, the MERA serves to renormalize the initial boundary state via coarse-graining. If we were to implement the MERA for only a few levels, we would end up with a quantum state in a smaller Hilbert space (defined on a fixed level of the tensor network), retaining some features of the original state but with some of the entanglement removed. However, we can also run the MERA backwards, to obtain a boundary state from a simple initial input. By varying the parameters in the individual tensors, we can look for an approximation of the ground state of the CFT on the boundary. Numerical evidence indicates that this process provides a computationally efficient method of constructing such ground states [281, 282].

The tensors, or gates, of the MERA come in two types. The first type are the disentanglers, represented by squares in Fig. 11.1. These are unitary maps $U : V \otimes V \to V \otimes V$, as in Fig. 11.1b. The name comes from thinking of moving upward through the network, in the direction of coarse-graining, where the disentanglers serve to remove local entanglement; as we move downward, of course, they take product states and entangle them. The second type of tensors are the isometries, represented by triangles. From the moving-downward perspective these are linear maps $W : V \to V \otimes V$; moving upward, they implement the coarse-graining, see Fig. 11.1c. The isometries are subject to the further requirement that $W^\dagger W = I_V$, where I_V is the identity map on V, and $WW^\dagger = P_A$, where P_A is a projector onto some subspace $A \subset V \otimes V$. From the top-down perspective, we can also think of the isometries as bijective unitary operators $W_U : V \otimes V \to V \otimes V$, for which a fixed "ancilla" state (typically the

ground state $|0\rangle$) is inserted in one of the input factors, as shown in Fig. 11.1c. More generally, isometries could map $q < k$ sites onto k sites, $W : V^{\otimes q} \to V^{\otimes k}$.

The MERA is not the simplest tensor network which implements coarse-graining. For instance, the tree tensor network [283] (also considered in a holographic context in Ref. [276]), 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 [284]. In $1 + 1$ dimensions, the model describes a chain of spins with nearest-neighbor interactions subject to a transverse magnetic field. Its Hamiltonian is

$$\hat{H} = -J \sum_i \hat{\sigma}_i^z \hat{\sigma}_{i+1}^z - h \sum_i \hat{\sigma}_i^x \,, \tag{11.2.1}$$

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

In general, much information is required to describe an arbitrary MERA. In principle, the Hilbert spaces, the disentanglers, and the isometries could all be different. Also, for $k > 2$, there is no canonical way of laying out the disentanglers and isometries; the circuit itself must be specified. We will restrict ourselves to the case $q = 1$, so that isometries have 1 upward-going leg and k downward-going legs. Further, without loss of generality, we take the

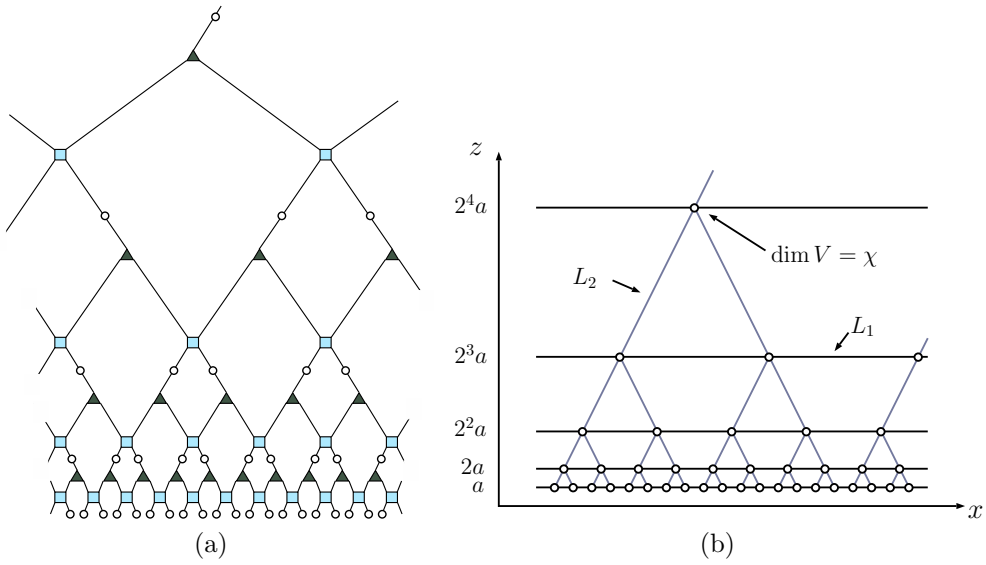

Figure 11.2. (a) A $k = 2$ MERA, and (b) the same MERA with its disentanglers and isometries suppressed. The horizontal lines in the graph on the right indicate lattice connectivity at different renormalization depths, and the vertical lines indicate which sites at different depths are related via coarse-graining 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.

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. 11.2. 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.

11.2.2 An AdS/MERA Correspondence?

The possibility of a correspondence between AdS and the MERA was first proposed by Swingle in Ref. [275], 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 [275, 285].

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 [213, 275]. Moreover, this phenomenon makes it possible to understand the computation of CFT entanglement entropy using a MERA as a discrete realization of the Ryu-Takayanagi formula [286]. These and other examples [213, 275] 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. 11.5, 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 [276], 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 chapter 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.

11.3 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 [213, 275] 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.

11.3.1 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. 11.2. 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. [275] 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

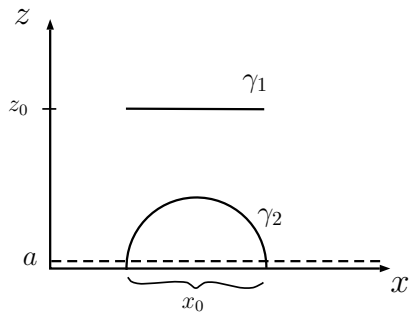

Figure 11.3. A horizontal line (γ_1) and a geodesic (γ_2) in a spatial slice of AdS_3.

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:

$$ds^2 = \frac{L^2}{z^2}(dz^2 + dx^2). \tag{11.3.1}$$

We will compare the proper lengths of straight horizontal lines and geodesics in the AdS slice to the proper lengths of the corresponding paths in the MERA graph. In the AdS slice, let γ_1 be a straight horizontal line ($dz = 0$) 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. 11.3). It is a straightforward computation to show that the proper lengths of these curves are

$$|\gamma_1|_{\text{AdS}} = \frac{L}{z_0}x_0 \qquad \text{and} \qquad |\gamma_2|_{\text{AdS}} = 2L\log\left(\frac{x_0}{a}\right). \tag{11.3.2}$$

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

We fix L_1 and L_2 by comparing γ_1 and γ_2 to horizontal lines and "geodesics" in the MERA, respectively. Consider two sites in a horizontal lattice at depth m (i.e., m renormalizations of the UV-most lattice) and separated by a coordinate distance x_0 in the coordinate system shown in Fig. 11.2. 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. 11.2 for the case $k = 2$). It follows that the proper length of the line connecting the two points is just

$$\begin{aligned} |\gamma_1|_{\text{MERA}} &= L_1 \cdot (\text{number of bonds between endpoints}) \\ &= L_1 \left.\frac{x_0}{z_0}\right|_{z_0 = k^m a} . \end{aligned} \tag{11.3.3}$$

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

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

$$\begin{aligned} |\gamma_2|_{\text{MERA}} &= L_2 \cdot (\text{number of bonds in the geodesic}) \\ &= 2L_2 \log_k \left(\frac{x_0}{a}\right) . \end{aligned} \tag{11.3.4}$$

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

11.3.2 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 parameterize the proper distance between lattice sites are of order the AdS scale L or larger, as was also noted in Refs. [275, 285]. 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

[213].

One could try to evade this difficulty by attempting to redefine the MERA coordinates $(x,z)^{\mathrm{MERA}}$ (those of Fig. 11.2) as functions of the AdS coordinates $(x,z)^{\mathrm{AdS}}$ (those of Fig. 11.3) and taking a continuum limit; above, we assumed that the two sets of coordinates were simply identified. That is, suppose $x^{\mathrm{MERA}} = f(x^{\mathrm{AdS}})$ and $z^{\mathrm{MERA}} = g(z^{\mathrm{AdS}})$. (For example, one could consider $f(x) = \varepsilon x$ for small ε and imagine taking the continuum limit, with the aim of making L_1 much smaller than the AdS scale.) If a is still the UV cutoff on the AdS side, then in the MERA we have $f(a)$ as the UV-most lattice spacing and $g(a)$ as the UV cutoff in the holographic direction. Consider the computation of $|\gamma_1|$. From the AdS side, we have $|\gamma_1|_{\mathrm{AdS}} = Lx_0^{\mathrm{AdS}}/z_0^{\mathrm{AdS}}$. On the MERA side, the number of sites spanned by $x_0^{\mathrm{MERA}} = f(x_0^{\mathrm{AdS}})$ is $x_0^{\mathrm{MERA}}/k^m f(a)$, while the holographic coordinate is $z_0^{\mathrm{MERA}} = k^m g(a)$. Hence,

$$|\gamma_1|_{\mathrm{MERA}} = L_1 \frac{f(x_0^{\mathrm{AdS}})}{f(a)} \frac{g(a)}{g(z_0^{\mathrm{AdS}})}. \tag{11.3.5}$$

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

$$g(z_0^{\mathrm{AdS}}) \frac{\partial}{\partial x_0^{\mathrm{AdS}}} |\gamma_1| = L_1 \frac{f'(x_0^{\mathrm{AdS}})}{f(a)} g(a) = L \frac{g(z_0^{\mathrm{AdS}})}{z_0^{\mathrm{AdS}}}. \tag{11.3.6}$$

Since the right side of the first equality only depends on x_0^{AdS} and the second equality only depends on z_0^{AdS}, but we can vary both parameters independently, both expressions must be independent of both AdS coordinates. Hence, we must have $f(x) = \varepsilon_x x$ and $g(z) = \varepsilon_z z$ for some constants ε_x and ε_z. Plugging everything back into Eq. (11.3.5) and comparing with $|\gamma_1|_{\mathrm{AdS}}$, we again find that $L_1 = L$, so no continuum limit is possible. Similarly, in computing $|\gamma_2|$, we note that the number of bonds between the endpoints on the UV-most lattice level is $x_0^{\mathrm{MERA}}/f(a)$, so the geodesic connecting the endpoints has $2\log_k(x_0^{\mathrm{MERA}}/\varepsilon_x a)$ bonds. On the other hand, we have $|\gamma_2|_{\mathrm{AdS}} = 2L\log(x_0^{\mathrm{AdS}}/a) = 2L\log(x_0^{\mathrm{MERA}}/\varepsilon_x a)$. That is, in equating $|\gamma_2|_{\mathrm{AdS}}$

and $|\gamma_2|_{\text{MERA}}$, we must again set $L_2 = L \log k$. 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) [287–289] 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.

11.4 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.

11.4.1 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 [256, 257]

$$S(B) = \frac{c}{3} \log \ell_0 \,, \tag{11.4.1}$$

where the length of the interval is much smaller than the system size. Here, ℓ_0 is the length of the interval in units of the UV cutoff. In the notation of the last section, we have $\ell_0 = x_0/a$. In the special case that the CFT is dual to AdS in $2+1$ dimensions, the central charge is set by the Brown-Henneaux formula [290],

$$c = \frac{3L}{2G}. \tag{11.4.2}$$

Also note that the length of the geodesic that connects the two ends of B (the curve γ_2 in Fig. 11.3) is given in Eq. (11.3.2) by $|\gamma_2| = 2L \log \ell_0$. The Brown-Henneaux relation allows us to reproduce the Ryu-Takayanagi formula [132, 291] from the entanglement entropy,

$$S(B) = \frac{\text{area}(\tilde{B})}{4G} \,, \tag{11.4.3}$$

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

The MERA calculation of the entanglement entropy of ℓ_0 sites in the CFT has an analogous geometric interpretation. Suppose one is given the MERA representation of a lattice CFT ground state, i.e., one uses a MERA to generate the CFT state. Denote by $S_{\text{MERA}}(\ell_0)$ the entanglement entropy of the resulting state restricted to ℓ_0 sites. In Ref. [286], it is shown that for a specific, optimal

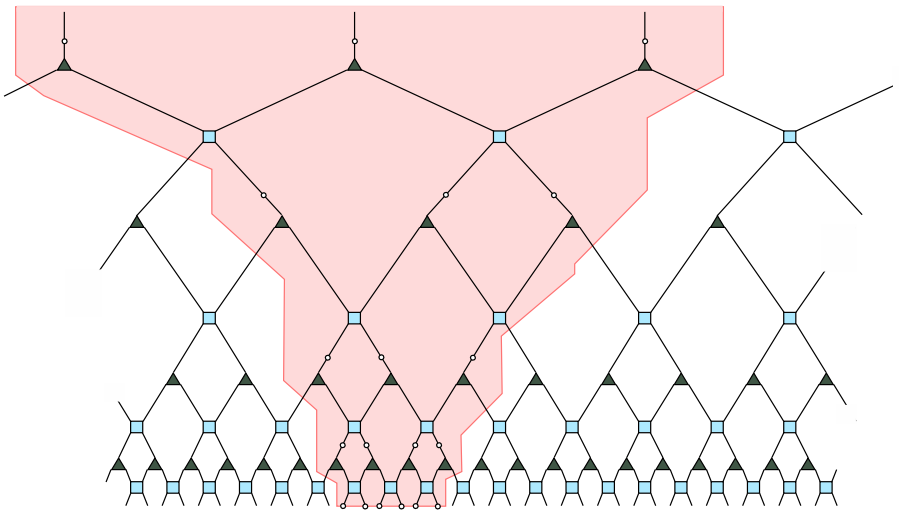

Figure 11.4. Causal cone (shaded) for a set of $\ell_0 = 6$ sites in a MERA with $k = 2$. The width ℓ_m of the causal cone at depth m is $\ell_1 = 4$, $\ell_2 = 3$, $\ell_3 = 3$, $\ell_4 = 3$, etc. The crossover scale for this causal cone occurs at $\bar{m} = 2$. Between the zeroth and first layer, $n_1^{\text{tr}} = 2$ bonds are cut by the causal cone. Similarly, $n_2^{\text{tr}} = 2$, $n_3^{\text{tr}} = 3$, etc.

choice of ℓ_0 sites, for ℓ_0 parametrically large, the following bound is placed on $S_{\text{MERA}}(\ell_0)$ for a MERA with $k = 2$:

$$S_{\text{MERA}}(\ell_0) \leq 2 \log_2 \ell_0 \log \chi. \tag{11.4.4}$$

Parsing the equation above, this bound essentially counts the number of bonds that the *causal cone* of the ℓ_0 sites in question crosses ($\sim 2 \log_2 \ell_0$) and $\log \chi$ 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. 11.4.

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. (11.4.4) 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. 11.4). The scale $\bar{m}$ is called the *crossover scale*.

For general k, it is also possible to formulate a bound similar to Eq. (11.4.4) for the entanglement entropy of ℓ_0 sites. For parametrically large ℓ_0, we find that

$$S_{\text{MERA}}(\ell_0; B) \leq 4(k-1) \log_k \ell_0 \log \chi \,. \tag{11.4.5}$$

We demonstrate this bound in App. 11.A using techniques that are similar to those developed in Ref. [286]. 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. (11.4.5) is more conservative than the optimal bound given in Eq. (11.4.4).

11.4.2 Constraining S_{MERA}

Let us examine Eq. (11.4.5) a bit more closely. As discussed in App. 11.A, $4(k-1)$ is an upper bound on the number of bonds that the causal cone could cut at any given depth m below the crossover scale $\bar{m}$. (The crossover scale $\bar{m}$ is attained after roughly $\log_k \ell_0$ renormalization steps.) For a given causal cone, i.e., for ℓ_0 sites at a given location with respect to the MERA, let us parameterize our ignorance by writing

$$S_{\text{MERA}}(\ell_0; B) \leq 4 f_B(k) \log_k \ell_0 \log \chi \,, \tag{11.4.6}$$

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

$$f_B(k) \equiv \frac{1}{4\bar{m}} \sum_{m=0}^{\bar{m}-1} n_m^{\mathrm{tr}}\,, \tag{11.4.7}$$

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 $\log \chi$ 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 (11.4.6) as an equality:

$$S_{\mathrm{MERA}}(\ell_0; B) = 4 f_B(k) \log_k \ell_0 \cdot \eta_B \log \chi. \tag{11.4.8}$$

The quantity $\eta_B \log \chi$ is the average entanglement entropy per cut bond in the causal cone of B. Equivalently, Eq. (11.4.8) may be taken as the definition of η_B.

This definition of η_B of course depends on the location of B and only applies to bonds that are cut by the causal cone of B. In what follows, it will be advantageous to have a notion of average entanglement entropy per bond that applies to *all* bonds in the MERA. To this end, start with a lattice consisting of ℓ_{tot} sites in total and consider the limit in which the size of a region B is unbounded but where the ratio $\ell_0/\ell_{\mathrm{tot}}$ 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. (11.4.1). Let us consequently define the average entanglement entropy per bond in the MERA:

$$\eta \log \chi = \lim_{\ell_0 \to \infty} \frac{S_{\mathrm{MERA}}(\ell_0)}{4 f(k) \log_k(\ell_0)}\,. \tag{11.4.9}$$

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

Intuitively, one would not expect each individual bond in the MERA to be maximally entangled and so it should be possible to constrain η more

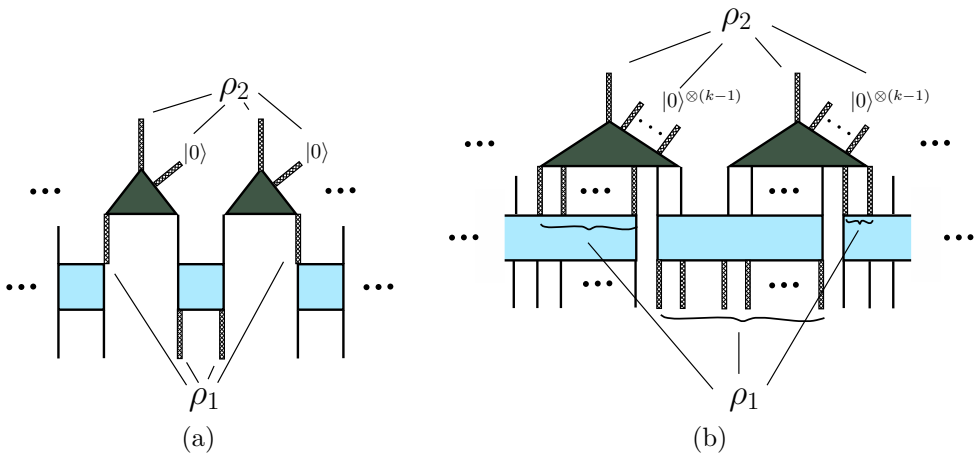

Figure 11.5. A pair of isometries with their ancillae explicitly indicated for a MERA with (a) $k = 2$ and (b) general k. The thick bonds below the isometries, the state of which is denoted by ρ_1, are unitarily related to the bonds that exit the isometries and the ancillae, the state of which is denoted by ρ_2.

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. 11.5a, let ρ_2 denote the density matrix of the bonds and ancillae emanating from the two isometries and let ρ_1 denote the density matrix of the four highlighted bonds below the isometries. We again assume that the ancillae are initialized to the pure product state composed of factors of $|0\rangle$. Taking into account the ancillae, or in other words promoting the isometries to unitaries, we see that ρ_1 and ρ_2 are related by a unitary transformation, so $S(\rho_1) = S(\rho_2)$. By assumption, the state of each ancilla is $|0\rangle$, so $\rho_2 = \tilde{\rho}_2 \otimes |0\rangle \langle 0| \otimes |0\rangle \langle 0|$ for some density matrix $\tilde{\rho}_2$. This in turn implies that $S(\rho_2) = S(\tilde{\rho}_2) \leq 2 \log \chi$. From the definition of η above, the entanglement entropy of a single bond is asymptotically given by $\eta \log \chi$, so $S(\rho_1) \simeq 4\eta \log \chi$. It therefore follows that $\eta \leq 1/2$.

For general k, the argument is nearly identical. We again begin by considering a pair of isometries at a given level m (see Fig. 11.5b). Analogously with the $k = 2$ case, let ρ_2 denote the density matrix of the two bonds and

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

$$\eta f(k) \leq \frac{k-1}{k}\,. \tag{11.4.10}$$

We note that, in computational practice, one typically does not use the "worst-case scenario" construction explored in App. 11.A; a more conventional construction would result in a tighter bound on $f(k)$ and hence a stricter inequality than Eq. (11.4.10). For our purposes, however, we will remain as conservative as possible and therefore use the inequality (11.4.10) in our subsequent bounds.

11.4.3 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 $S(\ell_0) = (c/3)\log\ell_0$ with the MERA entropy (11.4.8). Doing so, the central charge is given by

$$c = \frac{3L}{2G} = 12\eta\, f(k)\frac{\log\chi}{\log k}\,. \tag{11.4.11}$$

Then in light of Eq. (11.4.10), we find that

$$c \leq 12 \left(\frac{k-1}{k \log k} \right) \log \chi \,. \tag{11.4.12}$$

To recapitulate, given a CFT with central charge c and a MERA representation of its ground state, a necessary condition for a consistent AdS/MERA correspondence is that the MERA parameters k and χ satisfy the constraint (11.4.12). Importantly, this implies that, for a well-defined semiclassical spacetime (for which $c \gg 1$), the bond dimension χ must be exponentially large in the size of the AdS scale compared to the Planck scale.

Let us also note that we can still obtain a bound from Eq. (11.4.11), albeit a weaker one, without using the result of Eq. (11.4.10). Recall that this latter result relies on having unentangled ancillae in the MERA. This is not necessarily the case for other tensor network bulk constructions, as we will subsequently discuss. As such, if we disregard the result of Eq. (11.4.10), we still have by virtue of their definitions that $f(k) \leq k - 1$ and $\eta \leq 1$. The following weaker but more general bound on the central charge therefore follows from Eq. (11.4.11) for such generalized tensor networks:

$$c \leq 12 \left(\frac{k-1}{\log k} \right) \log \chi. \tag{11.4.13}$$

11.5 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.

11.5.1 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 [59] and the Bekenstein bound [292] and later in more general holographic contexts, culminating in the covariant entropy bound, i.e., the Bousso bound [258, 265].

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

$$S(\mathcal{B}) \leq \frac{A}{4G}\,. \tag{11.5.1}$$

It is natural to cast the Bousso bound as a constraint on the dimension of the bulk Hilbert space. As argued in Ref. [71], 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 naive Hilbert space in quantum field theory), which the Bousso

bound implies is less than $A/4G$.[11.1] As such, if we denote the Hilbert space of $\mathcal{B}$ by $\mathcal{H}_\mathcal{B}$, let us replace Eq. (11.5.1) with the slightly more concrete statement

$$\log \dim \mathcal{H}_\mathcal{B} \leq \frac{A}{4G}. \tag{11.5.2}$$

11.5.2 A MERA Version of the Bousso Bound

Our aim is to compute both sides of the inequality (11.5.2) using the MERA. For this calculation, it is instructive to change our parameterization 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., [295]; 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. 11.6a for $k = 2$ and $m = 4$. Note 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.

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. 11.6b. Let us also assume periodic boundary conditions for this MERA and hence identify $x = 0$ and $x = k^m a$.

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

$$\begin{aligned} \rho &= \frac{k^m a - z}{k^m a}, \\ \theta &= 2\pi \frac{x}{k^m a}. \end{aligned} \tag{11.5.3}$$

[11.1] Moreover, it is known that there exists an asymptotically-AdS bulk configuration that saturates the Bousso bound, namely, the BTZ black hole [293, 294], which further implies that $\log \dim \mathcal{H}_\mathcal{B}$ 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. [261].

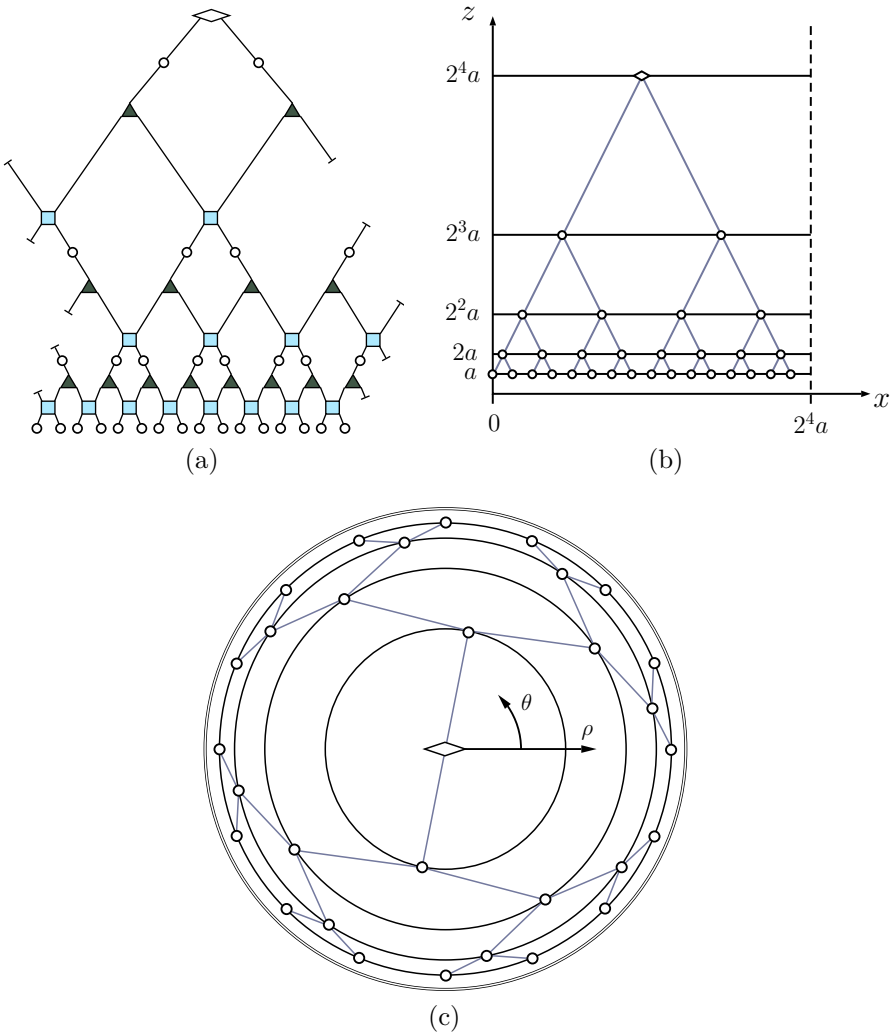

Figure 11.6. (a) A $k = 2$ MERA consisting of $m = 4$ layers and with periodic boundary conditions, (b) the corresponding embedding in (x, z) coordinates, and (c) the embedding in (ρ, θ) coordinates.

In these coordinates, the metric reads

$$\mathrm{d}s^2 = \frac{L^2}{(1-\rho)^2}\left[\mathrm{d}\rho^2 + \left(\frac{\mathrm{d}\theta}{2\pi}\right)^2\right], \tag{11.5.4}$$

cf. Eq. (11.3.1). This embedding of the MERA is shown in Fig. 11.6c; the top-level tensor always sits at $\rho = 0$ and the lower layers of the MERA are equally spaced on circles of radii 1/2, 3/4, 7/8, ... that are centered at $\rho = 0$.

More generally, one could construct a top-level tensor that has T legs, each of which begets a tree of sites. In this case, $x = 0$ and $x = Tk^{m-1}a$ are

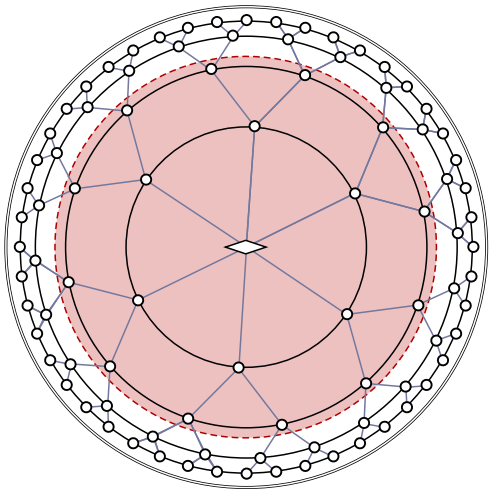

Figure 11.7. Disk parameterization of the Poincaré patch of AdS in which a MERA has been embedded. The top tensor of the MERA shown has $T = 6$. The shaded region is a ball $\mathcal{B}$, which is this case contains $N_{\mathcal{B}} = 1$ generation.

identified, so one should define the angular variable as $\theta \equiv 2\pi x/(Tk^{m-1}a)$. The metric (11.5.4) is correspondingly modified and reads

$$\mathrm{d}s^2 = \frac{L^2}{(1-\rho)^2}\left[\mathrm{d}\rho^2 + \frac{T^2}{k^2}\left(\frac{\mathrm{d}\theta}{2\pi}\right)^2\right] . \tag{11.5.5}$$

This situation is depicted in Fig. 11.7. (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. (11.5.2). Let the ball $\mathcal{B}$ be centered about $\rho = 0$, and suppose $\mathcal{B}$ contains the top-level tensor, the sites at the top tensor's legs, and then the first $N_{\mathcal{B}}$ generations of the MERA emanating from these sites, as indicated in Fig. 11.7. 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

$$\frac{A}{4G} = \frac{Tk^{N_{\mathcal{B}}}L}{4G} = \frac{Tk^{N_{\mathcal{B}}}c}{6} , \tag{11.5.6}$$

where in the second equality we used the Brown-Henneaux relation, Eq. (11.4.2).

How one evaluates the left-hand side of Eq. (11.5.2) using the MERA is not as immediate. Recall that $\mathcal{H}_\mathcal{B}$ is the Hilbert space of *bulk states*. The MERA, however, does not directly prescribe the quantum-gravitational state in the bulk; it is not by itself a bulk-boundary dictionary. As we mentioned in Sec. 11.2.2, 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

$$\mathcal{N}_\mathcal{B} = T\sum_{i=0}^{N_\mathcal{B}} k^i = T\left(\frac{k^{N_\mathcal{B}+1}-1}{k-1}\right). \tag{11.5.7}$$

As such, the Hilbert space of bulk states restricted to $\mathcal{B}$ is $\mathcal{H}_\mathcal{B} = (V_\text{bulk})^{\otimes\mathcal{N}_\mathcal{B}} \otimes V_\text{T}$. Next, suppose that $\dim V_\text{bulk} = \tilde{\chi}$ and that $\dim V_\text{T} = \tilde{\chi}_\text{T}$, where, like χ, $\tilde{\chi}$ and $\tilde{\chi}_\text{T}$ are some fixed, $N_\mathcal{B}$-independent numbers. Then $\dim \mathcal{H}_\mathcal{B} = \tilde{\chi}_\text{T}(\tilde{\chi}^{\mathcal{N}_\mathcal{B}})$. 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. (11.5.6) and (11.5.7), the dimensionality of $\mathcal{H}_\mathcal{B}$ is upper bounded as follows:

$$\log \dim \mathcal{H}_{\mathcal{B}} \leq \frac{A}{4G} \quad \Longrightarrow \quad T\left(\frac{k^{N_{\mathcal{B}}+1}-1}{k-1}\right)\log\tilde{\chi}+\log\tilde{\chi}_{\mathrm{T}} \leq \frac{Tk_{\mathcal{B}}^{N}c}{6}\,. \quad (11.5.8)$$

After isolating c in Eq. (11.5.8) and using the result of Eq. (11.4.11), we find that

$$c = 12\eta f(k)\frac{\log\chi}{\log k} \geq 6\left(\frac{k^{N_{\mathcal{B}}+1}-1}{k^{N_{\mathcal{B}}}(k-1)}\log\tilde{\chi} + \frac{1}{Tk^{N_{\mathcal{B}}}}\log\tilde{\chi}_{\mathrm{T}}\right). \quad (11.5.9)$$

Next, let us consider this inequality in the limit of large $N_{\mathcal{B}}$. A motivation for this limit is the fact that the natural scale of validity of an AdS/MERA correspondence is super-AdS, as was established in Sec. 11.3. 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. (11.5.9) reduces to

$$\eta f(k) \geq \frac{k\log k}{2(k-1)}\left(\frac{\log\tilde{\chi}}{\log\chi}\right). \quad (11.5.10)$$

By using the bound on $\eta f(k)$ given by Eq. (11.4.10), we arrive at a constraint on k, χ, and $\tilde{\chi}$:

$$\frac{k^2\log k}{2(k-1)^2}\left(\frac{\log\tilde{\chi}}{\log\chi}\right) \leq 1. \quad (11.5.11)$$

In principle, the above inequality could be satisfied for any k, provided that the dimension $\tilde{\chi}$ of the factors V_{bulk} can be arbitrarily chosen with respect to the bond dimension χ. However, the essence of holography, that the bulk and boundary are dual descriptions of the same degrees of freedom and therefore have isomorphic Hilbert spaces [54], implies a relation between χ and $\tilde{\chi}$. Namely, for a MERA with a total of N levels of sites in the bulk strictly between the UV-most level and the top-level tensor, the number of bulk sites $\mathcal{N}_{\mathrm{bulk}}$ that are not located at the top tensor is given by Eq. (11.5.7) 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[11.2] the dimension of the bulk and boundary Hilbert spaces then yields

$$\frac{\log\tilde{\chi}}{\log\chi} = \frac{1}{\mathcal{N}_{\text{bulk}}}\left(Tk^{N+1} - \frac{\log\tilde{\chi}_{\text{T}}}{\log\chi}\right) \overset{N\,\text{large}}{\to} k-1, \tag{11.5.12}$$

where we took the limit of N large, consistent with Eq. (11.5.10) and in keeping with the expectation that the UV cutoff be parametrically close to the boundary at $\rho = 1$. Putting together Eqs. (11.5.11) and (11.5.12), we obtain a constraint on k alone:

$$\frac{k^2\log k}{2(k-1)} \leq 1. \tag{11.5.13}$$

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

If we relax this bound and, instead of Eq. (11.4.10), only observe the weaker, natural bounds $\eta \leq 1$ and $f(k) \leq k-1$ as discussed at the end of Sec. 11.4.3, the constraint (11.5.13) is correspondingly modified:

$$\frac{k\log k}{2(k-1)} \leq 1. \tag{11.5.14}$$

In contrast to Eq. (11.5.13), this latter bound can be satisfied, but only for

[11.2] We recognize that there are other proposals [277, 296] 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.

$k = 2$, 3, or 4. As such, other AdS/tensor network correspondences, in which the ancillae are perhaps entangled and therefore do not describe a conventional MERA, are not ruled out. Note that we never needed to compute bulk entanglement entropy explicitly—and therefore did not need to treat separately the possibility of entanglement among ancillae—because we cast the Bousso bound as a constraint on the size of the bulk Hilbert space itself. The appearance of η in Eq. (11.5.10) corresponds to entanglement in the boundary theory as computed by the tensor network; Eqs. (11.5.10) and (11.5.12) still apply.

11.6 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. 11.3, 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. (11.4.12) and (11.4.13)], which implies that the bond dimension χ must be exponentially large in the ratio of the AdS scale to the Planck scale.

3. In the natural construction of a bulk Hilbert space ($\mathcal{H}_{\text{bulk}}$) using the MERA, we used the Bousso bound to constrain the dimension of $\mathcal{H}_{\text{bulk}}$. When combined with our previous results, we found that any strict AdS/MERA correspondence cannot satisfy the resulting constraint, Eq. (11.5.13). Upon relaxing the definition of the MERA or allowing for additional structure, however, we obtained a looser constraint, Eq. (11.5.14), which may not rule out some other AdS/tensor network correspondences.

In particular, more general correspondences between AdS and MERA-like tensor networks, in which we allow the ancillae to be entangled when reproducing the CFT ground state [and for which Eq. (11.5.14) applies in place of Eq. (11.5.13)] are not ruled out by our bounds, provided that the rescaling factor $k = 2$, 3, or 4. Further, it is interesting to note that our bounds extend to states other than the vacuum that are described by a MERA. One such example, namely, states at finite temperature dual to black holes in AdS, is discussed in App. 11.B below.

While the consistency conditions that we found are specific to the MERA tensor network, many of the ideas and techniques that we used apply equally well to other tensor networks. In the EHM, for instance, the type of bulk Hilbert space dimensionality arguments that we made based on the covariant entropy bound may be directly transferred to the EHM. The same stringent final constraints that we derived do not apply to the EHM, however, since it is unclear to what extent the EHM reproduces the Ryu-Takayanagi formula

(which renders the results of Sec. 11.4 inapplicable). Our bulk Hilbert space arguments similarly apply to the holographic error-correcting code proposal in Ref. [277], 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 chapter, are important for elucidating the subtleties in this quest and providing guidance along the way.

11.A Entropy Bound for General MERAs

Following the method presented in Ref. [286], 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. [286] throughout this chapter.

First, recall the result from Ref. [286] that the entanglement entropy of a region consisting of ℓ_0 sites is bounded by

$$S_{\mathrm{MERA}}(\ell_0; B) \leq (\ell_{m'} + N^{\mathrm{tr}}_{m'}) \log \chi. \tag{11.A.1}$$

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. 11.4). The quantity $\log \chi$ is the maximum entanglement entropy that each site that is traced out could contribute to $S_{\mathrm{MERA}}(\ell_0; B)$. Note that Eq. (11.A.1) holds for all $m' \geq 0$.

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

Nevertheless, we can compute an upper bound for $S_{\mathrm{MERA}}(\ell_0; B)$ by consid-

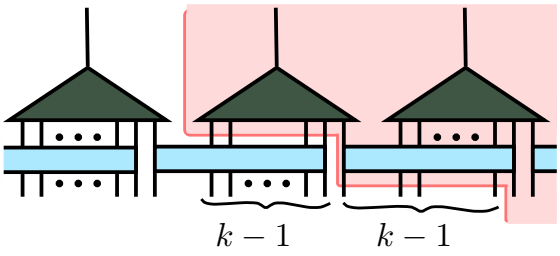

Figure 11.8. Left side of a causal cone that cuts the maximum possible number of bonds over the course of one renormalization step. The rectangles are disentanglers that accept k bonds as input and the triangles are isometries that coarse-grain k bonds into one. The causal cone is the shaded region. If this situation is mirrored on the right side of the causal cone, then $4(k-1)$ bonds are cut in this renormalization step.

ering 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. 11.8. 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., $n^{\mathrm{tr}}_{m'} = 4(k-1)$.

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. 11.8 occurs at every step up until the crossover scale. Then, by Eq. (11.A.1), the entropy

bound is given by

$$S_{\mathrm{MERA}}(\ell_0; B) \leq (\ell_{\bar{m}} + 4(k-1)\bar{m}) \log \chi, \tag{11.A.2}$$

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

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

$$\ell_{m'} \;\leq\; \frac{\ell_0}{k^{m'}} + 5 \sum_{m=1}^{m'} \frac{1}{k^m} \;\leq\; \frac{\ell_0}{k^{m'}} + 5\,. \tag{11.A.3}$$

Setting $m' = \bar{m}$, it trivially follows that the crossover scale obeys $\bar{m} \leq \log_k \ell_0$. Furthermore, we notice that this is the scale at which the entanglement entropy is minimized if we trace over the remaining sites. In other words, the number of bonds cut by going deeper into the renormalization direction is no less than the bonds cut horizontally, so $4(k-1) \geq \ell_{\bar{m}}$.[11.3] Applying the bounds for $\bar{m}$ and $\ell_{\bar{m}}$ on Eq. (11.A.2), we arrive at an upper bound on $S_{\mathrm{MERA}}(\ell_0; B)$ for a k-to-one MERA,

$$S_{\mathrm{MERA}}(\ell_0; B) \leq 4(k-1)(1 + \log_k \ell_0) \log \chi. \tag{11.A.4}$$

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. (11.4.5).

[11.3] 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, $\ell_{\bar{m}} \lesssim 4 \leq 4(k-1)$.

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

Thus far, we have found constraints on the structure of a MERA that can describe CFT states dual to the AdS_3 vacuum. One might ask whether these results extend to other constructions that exist in three-dimensional gravity. Although pure gravity in AdS_3 has no local or propagating degrees of freedom, there exist interesting nonperturbative objects, namely, BTZ black holes [293]. In this appendix, we extend our constraints on boundary entanglement entropy to these objects.

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

$$\mathrm{d}s^2 = -\frac{(r^2 - r_+^2)}{L^2}\mathrm{d}t^2 + \frac{L^2}{(r^2 - r_+^2)}\mathrm{d}r^2 + r^2\mathrm{d}\phi^2\,, \tag{11.B.1}$$

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

$$S_{\mathrm{BH}} = \frac{\mathrm{Area}}{4G} = \frac{\pi r_+}{2G}\,. \tag{11.B.2}$$

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

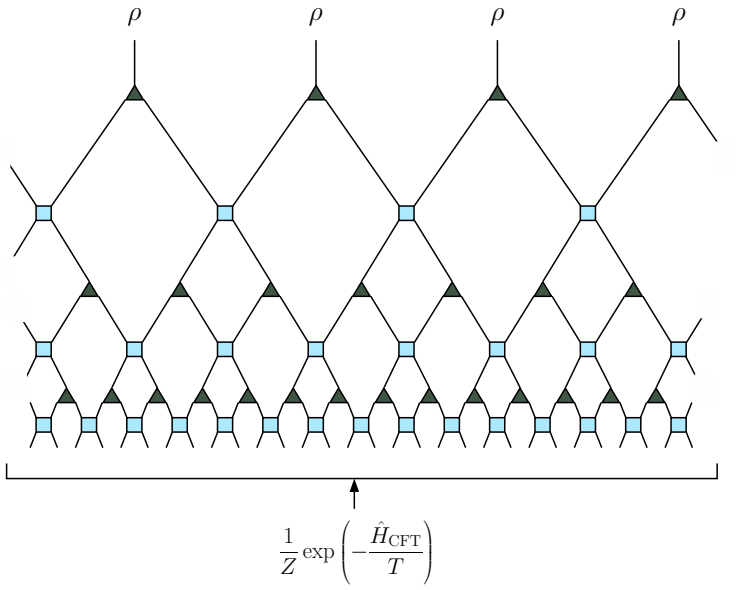

Figure 11.9. The MERA, when applied to a thermal CFT state $Z^{-1}\exp(-\hat{H}_{\mathrm{CFT}}/T)$, where $Z = \mathrm{tr}(\exp(-\hat{H}_{\mathrm{CFT}}/T))$, truncates after a finite number of layers. The boundary state at the top of the truncated MERA effectively factorizes into a product of maximally mixed states $\rho = I/\chi$.

extensive contribution, truncates at a level with multiple sites. The schematic entanglement renormalization process is illustrated in Fig. 11.9. The state at the top level effectively factorizes, where each factor appears maximally mixed [213, 275]. A tractable realization of this tensor network structure recently appeared in Ref. [297], which found a MERA representation of a thermal state.

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 [275]. 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. [297], 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 (11.B.1), we see that $T x_{\mathrm{sys}} = r_+/L$, as was pointed out in Refs. [132, 291]. 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. 11.9 as a discretization of a BTZ spacetime and repeat the analysis of Sec. 11.3. In this discretization, the layers of the MERA lie along circles of fixed radius r in the coordinates of Eq. (11.B.1). Again, we ask what proper length L_1 separates sites in any given layer of the MERA.

First, note that a path at fixed r_0 that subtends an angle ϕ_0 has proper length $r_0\phi_0$. At the boundary of the MERA, we consider a region defined by $0 \leq \phi \leq \phi_0 = 2\pi x_0/x_{\mathrm{sys}}$, 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 $r_{\mathrm{b}} = L^2/a$ [132]. 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

$$m = \log_k \left(\frac{n_{\mathrm{b}}}{n_{\mathrm{h}}}\right) = \log_k \frac{1}{2\pi T a}\,. \tag{11.B.3}$$

This coincides with the conclusion in Refs. [297, 298] 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 $\log\chi$ 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 $\log\chi$. Combining the contributions, we find

$$S_{\mathrm{MERA}}(B) = 4\eta_B f_B(k) \log_k \left(\frac{\ell_0}{\ell_{\mathrm{h}}}\right) \log\chi + \ell_{\mathrm{h}} \log\chi\,. \tag{11.B.4}$$

Recall that the entanglement entropy of a single interval B of coordinate length x_0 in a CFT at finite temperature [257] is given, up to a non-universal constant, by

$$S_{\mathrm{CFT}}(B) = \frac{c}{3} \log\left(\frac{1}{\pi a T} \sinh \pi x_0 T\right)\,, \tag{11.B.5}$$

where x_0 is much smaller than the total system size $x_{\rm sys}$. 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. (11.B.5), we recover the usual result (11.4.1). In the $T \to \infty$ limit, the von Neumann entropy gives the usual thermal entropy as entanglement vanishes. Taking $Tx_0 \gg 1$, the leading and subleading contributions to the entanglement entropy are

$$S_{\rm CFT} = \frac{c}{3}\pi x_0 T + \frac{c}{3}\log\frac{1}{2\pi a T}\,, \tag{11.B.6}$$

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

Now let us consider a finite-temperature CFT that is dual to a BTZ black hole with horizon temperature $T = r_+/2\pi L^2$. In terms of geometric MERA parameters, we find that Eq. (11.B.6) becomes

$$S_{\rm CFT} = \frac{c}{6}\,\ell_{\rm h} + \frac{c}{3} m \log k\,. \tag{11.B.7}$$

Here we used the fact that $\ell_{\rm h} = x_0 r_+/L^2$ as well as Eq. (11.B.3), where we note that m can also be written as $\log_k(\ell_{\rm b}/\ell_{\rm h})$. The result (11.B.7) coincides precisely with the extensive and subextensive contributions calculated using the MERA in Eq. (11.B.4) provided that $c/\log\chi \sim \mathcal{O}(1)$. 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. [285, 298].

As a check of the claim that c and $\log\chi$ 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 (11.B.2) of a BTZ black hole.

There are n_{h} sites comprising the horizon, each with Hilbert space dimension χ. The system is in the infinite-temperature limit—and hence described by a maximally mixed density matrix, with entropy contribution $\log \chi$ from each site—so

$$S_{\mathrm{horizon}} = n_{\mathrm{h}} \log \chi \,. \tag{11.B.8}$$

Making use of the Brown-Henneaux relation and requiring that the entropy (11.B.8) coincide with the Beckenstein-Hawking entropy, we again find that $c/\log \chi \sim \mathcal{O}(1)$. More specifically, taking the counting to be precise, we find that

$$c/\log \chi = 6 \,, \tag{11.B.9}$$

which is qualitatively in agreement with the previous conclusion (11.4.12) that the Hilbert space dimension must be exponentially large in c.

With this relation, the extensive terms in Eqs. (11.B.4) and (11.B.7) agree precisely. Further identifying the subextensive terms, we find $\eta_B f_B(k) = (\log k)/2$. If we then impose the constraint (11.4.10), we find that

$$\frac{k \log k}{2(k-1)} \leq 1 \,. \tag{11.B.10}$$

This last inequality exactly reproduces Eq. (11.5.14) and thus constrains k to be 2, 3, or 4. Interestingly, we have found the weaker of the two bounds derived in Sec. 11.5, 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. 11.4 further hold in BTZ spacetimes might not be too surprising given that such spacetimes are quotients of pure AdS_3.

Chapter 12
What is the Entropy in Entropic Gravity?

We investigate theories in which gravity arises as a consequence of entropy. We distinguish between two approaches to this idea: *holographic gravity*, in which Einstein's equation arises from keeping entropy stationary in equilibrium under variations of the geometry and quantum state of a small region, and *thermodynamic gravity*, in which Einstein's equation emerges as a local equation of state from constraints on the area of a dynamical lightsheet in a fixed spacetime background. Examining holographic gravity, we argue that its underlying assumptions can be justified in part using recent results on the form of the modular energy in quantum field theory. For thermodynamic gravity, on the other hand, we find that it is difficult to formulate a self-consistent definition of the entropy, which represents an obstacle for this approach. This investigation points the way forward in understanding the connections between gravity and entanglement.

This chapter is from Ref. [11], S. M. Carroll and G. N. Remmen, "What is the Entropy in Entropic Gravity?," Phys. Rev. ***D93*** *(2016) 124052,* arXiv:1601.07558 [hep-th].

12.1 Introduction

The existence of a profound relationship between gravity and entropy has been recognized since the formulation of the laws of black hole mechanics [210] and the derivation of the Bekenstein-Hawking entropy [211, 299]. More recently, ideas such as the holographic principle [50, 51], black hole complementarity [124],

the gauge/gravity correspondence [52–54], and the firewall puzzle [215, 216] have provided further hints that a deep relationship between gravitation and entropy will be present in the ultimate theory of quantum gravity.

In the quest to explore this connection and further our understanding of quantum gravity, there have been several proposals for directly linking gravity and entanglement. These proposals fall essentially into two distinct types, which we dub *holographic gravity* (HG) and *thermodynamic gravity* (TG). The labels are not perfect, as HG is related to thermodynamics and TG is related to holography, but they will serve as a useful shorthand for the two approaches.

In holographic gravity, one considers variations of the spacetime geometry and quantum state within a region, posits a relationship between the change in entanglement entropy and the change in the area of the boundary, and then uses these constraints to derive the Einstein equation in a bulk spacetime. This approach was used successfully in Refs. [212, 300] in an AdS/CFT context (see also Ref. [301]) and in Ref. [60] in a more general setup based on local causal diamonds. In holographic gravity, gravity emerges as a dual description of the entanglement entropy of the degrees of freedom in a local region.

In thermodynamic gravity, there is no variation over different states. Rather, one fixes a dynamical spacetime and a particular energy-momentum background. One then posits a relationship between some entropy flux (defined using the energy-momentum tensor) and some cross-sectional area (e.g., of a given null surface). Using this area-entropy relation, one derives the Einstein equations. This was the method of Ref. [59], as well as Refs. [302–305]. While these approaches are similar in spirit, Verlinde [302] emphasizes the existence of an entropic force from the gradient of the entropy, while Jacobson [59] derives the Einstein equation directly as a local equation of state.

Open questions are present in both HG and TG approaches. For definiteness,

we will focus on Jacobson's version of HG in Ref. [60] and of TG in Ref. [59]. In the HG case, we clarify the underlying assumptions of the theory and present arguments in their favor. In particular, we show how new results on entanglement entropy and the modular Hamiltonian in quantum field theory [258, 265] can be used to justify a crucial infrared assumption in HG. On the other hand, we find that TG exhibits a tension related to the fact that the "entropy" is not well defined in this theory. We will argue that it is difficult to find a self-consistent definition of the entropy in TG approaches. Our results indicate that holographic gravity is successful and points the way toward promising future results; reassuringly, holographic gravity is most closely related to AdS/CFT, in that it makes gravity in the bulk of a region dual to entanglement constraints on the boundary, in a sense that we will explore later.

The remainder of this chapter is organized as follows. In Sec. 12.2 we first review the holographic formulation of entropic gravity, identify its axioms, and examine its derivation of the Einstein equation. Afterwards, we demonstrate that the axioms of this theory can be justified in part using recent results in quantum field theory. In Sec. 12.3 we examine the formulation of the thermodynamic approach to entropic gravity and demonstrate the origin of the difficulties it experiences in defining the entropy. Finally, we summarize and discuss future directions in Sec. 12.4.

12.2 Holographic Gravity

After reviewing the motivation for relating entropy, particularly that of entanglement, with gravitation, we codify the axioms of holographic gravity and demonstrate the derivation of the Einstein equation. We then investigate how to justify and make rigorous each of the postulates underlying HG.

12.2.1 Motivation

We start with the underlying motivation of the holographic approach to entropic gravity. One of the most important facts that we (believe we) know about quantum gravity is the proportionality relationship between entropy S and horizon area A [299, 306],[12.1]

$$S = \frac{A}{4G\hbar}. \tag{12.2.1}$$

The derivation of this fact is phenomenological: the energy E of the black hole is given by its mass, Hawking used quantum field theory in curved spacetime to calculate the temperature $T = \hbar/8\pi GM$, and then we can use the thermodynamic relation $1/T = \partial S/\partial E$ to define the entropy. One expects that this entropy represents the logarithm of the underlying degrees of freedom in the true theory of quantum gravity; this expectation has been successfully borne out in certain stringy models of black holes [307, 308].

In the black hole case, it is clear what the entropy is actually the entropy of: the black hole, or at least the degrees of freedom that macroscopically appear to us as a black hole. That is a system that can be objectively defined in a way upon which all observers would agree. But the same formula (12.2.1) applies to the horizon of de Sitter space, as shown by Gibbons and Hawking [309]. The de Sitter horizon is an observer-dependent notion; given any worldline extended to future infinity, the horizon separates events within the causal diamond of that worldline from those outside. This suggests that the identification of the entropy as belonging to the system described by the horizon applies more universally than to fixed objects like black holes and indeed may apply to horizons in general.

Another clue comes from the existence of Rindler horizons in Minkowski

[12.1] Throughout this chapter, we leave $\hbar$ explicit in expressions leading to the derivation of the Einstein equation, as a bookkeeping device for semiclassicality.

space. Starting from the vacuum state of a general interacting quantum field theory, the Bisognano-Wichmann theorem guarantees that the density matrix restricted to the wedge $z > |t|$ is that of a thermal state with respect to the boost Hamiltonian [310, 311]. The boundary of the wedge acts as a horizon for observers who are moving with a constant acceleration along the z-axis. In 3+1 dimensions, the area of this horizon is infinite, so it is unsurprising that the von Neumann entropy of the corresponding density matrix is also infinite. We can ask, however, about the entropy density per unit horizon area. This is also infinite, which can be attributed to the contributions of ultraviolet modes of the field. Imposing an arbitrary short-distance cutoff, we find that there is a constant, fixed amount of entropy per unit horizon area. Since the original calculation was carried out in flat-spacetime quantum field theory, it is natural to suppose that the true entropy density would be finite in a quantum theory of gravity.[12.2]

Together, these facts suggest that there is a universal relationship: to any horizon, we can associate an entropy proportional to its area. This observation was the inspiration for entropic gravity in Refs. [59, 60]. It remains to formulate a precise prescription for what kind of entropy is actually involved. The natural candidate in the quantum context is the von Neumann entanglement entropy, $-\mathrm{Tr}\,\rho \log \rho$ for some density matrix ρ. Taking a vacuum spacetime region and cutting off modes at some fixed short distance, we obtain an entanglement entropy that is proportional to the area of the boundary of the region being considered [261, 312]. The entanglement entropy further appears in the recent proofs of versions of the covariant entropy bound within quantum field theory [258, 265]. Moreover, the Ryu-Takayanagi formula [132, 272] in AdS/CFT

[12.2] As noted in Ref. [60], we can say that gravity cuts off the number of degrees of freedom and renders the entropy finite or that demanding a finite horizon entropy implies the existence of gravity. Requiring finite entropy at least implies some ultraviolet cutoff for the applicability of quantum field theory.

relates the entanglement entropy in a boundary region with the area of an extremal surface in the bulk. The conjectured ER=EPR duality [57] (see also Refs. [7–9]) further underscores this connection. Taking these results as motivation, holographic gravity seeks to relate the Einstein equations themselves to constraints on entanglement entropy and areas in a sense that we will make precise.

12.2.2 Formulation of Holographic Gravity

Let us now review the approach to HG laid out in Ref. [60]. Fix an arbitrary background D-dimensional spacetime geometry M and a spacelike slice Σ. Choose a point $p \in \Sigma$ and define a ball B as the set of points $p' \in \Sigma$ such that the geodesic distance in Σ between p and p' is less than ℓ. Next, define the *causal diamond* $D(B)$ associated with B as the union of the past and future domains of dependence of B; that is, the set of all points $x \in M$ such that all inextendible timelike curves through x necessarily intersect B; see Fig. 12.1. We write as V the volume of B and A the area of ∂B. For a sufficiently small causal diamond, the background metric approaches the Minkowski form. There is a unique conformal isometry generated by the Killing vector

$$\zeta = \frac{1}{2\ell}[(\ell^2 - u^2)\partial_u + (\ell^2 - v^2)\partial_v], \tag{12.2.2}$$

where $u = t - r$ and $v = t + r$ for time coordinate t and radial coordinate r.

Writing the quantum state of the system on Σ as $|\psi\rangle$, we can define the reduced density matrix on B as

$$\rho_B = \mathrm{Tr}_{\,\Sigma - B}\, |\psi\rangle \langle\psi| \,. \tag{12.2.3}$$

We define the entanglement entropy associated with B as

$$S_B = -\mathrm{Tr}\, \rho_B \, \log \rho_B, \tag{12.2.4}$$

i.e., the entanglement of the state on B with that on $\Sigma - B$. We posit that

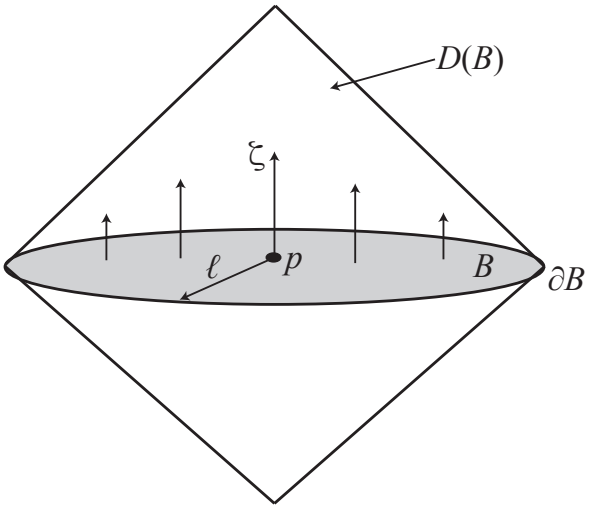

Figure 12.1. A small causal diamond $D(B)$ for a spacelike ball B with boundary ∂B. The ball is defined as all points in some spacelike surface that are less than or equal to a distance ℓ from some point p. The vector field ζ generates a conformal isometry within $D(B)$, assumed to be approximated by a maximally symmetric spacetime.

the Hilbert space of states on B can be factorized into infrared and ultraviolet contributions,

$$\mathcal{H}_B = \mathcal{H}_{\rm UV} \otimes \mathcal{H}_{\rm IR}. \tag{12.2.5}$$

The infrared states are ordinary field-theory states in a spacetime background (including semiclassical gravitational perturbations), while the ultraviolet contributions represent short-distance physics, including specifically quantum-gravitational degrees of freedom. Writing $\Lambda_{\rm UV}$ for the scale of the UV completion, which we take to be below the Planck scale, then $\mathcal{H}_{\rm IR}$ and $\mathcal{H}_{\rm UV}$ contain degrees of freedom with energies below and above $\Lambda_{\rm UV}$, respectively. The size ℓ of the causal diamond is taken to be larger than the Planck length but smaller than $1/\Lambda_{\rm UV}$. Tracing out the UV degrees of freedom, we are left with an infrared density matrix

$$\rho_{\rm IR} = {\rm Tr}_{\rm UV} \rho_B. \tag{12.2.6}$$

We then define the (field-theoretic) modular Hamiltonian K on B via the

implicit relation

$$\rho_{\rm IR} = \frac{e^{-K}}{{\rm Tr}\, e^{-K}}. \tag{12.2.7}$$

In Minkowski space, the causal diamond $D(B)$ can be mapped via a conformal transformation to the Rindler wedge [313]; writing $x^\mu = (t, r, \vec{y})$ for the (radial) coordinates of the ball, $X^\mu = (X^0, X^1, \vec{Y})$ for the (Cartesian) coordinates of the Rindler wedge $X^1 > 0$, and defining $B^\mu = (0, 1, 0, ..., 0)/2\ell$, the conformal transformation is

$$x^\mu = \frac{X^\mu - B^\mu X^2}{1 - 2X\cdot B + B^2 X^2} + 2\ell^2 B^\mu, \tag{12.2.8}$$

where $X^2 = X^\mu X_\mu$ and similarly for B^2. With $U = X^0 - X^1$ and $V = X^0 + X^1$, the Rindler wedge corresponds to the intersection of $V > 0$ and $U < 0$, which maps to the causal diamond $D(B) = \{v < \ell\} \cap \{u > -\ell\}$. For the Rindler wedge, the Bisognano-Wichmann theorem [311] guarantees that the density matrix is thermal with respect to the Hamiltonian generating time translation. Thus, for a conformal field theory (CFT), which is invariant under this transformation of the geometry, the modular Hamiltonian K is just the Hamiltonian generating flow along ζ from Eq. (12.2.2), namely,

$$K_{\rm CFT} = \frac{2\pi}{\hbar}\int_B T^{\mu\nu}\zeta_\mu {\rm d}\Sigma_\nu, \tag{12.2.9}$$

where ${\rm d}\Sigma_\mu$ is the surface element orthogonal to B and $T_{\mu\nu}$ is the energy-momentum tensor.

We now consider a variation of the spacetime M and of the quantum state ρ_B on B. We will write this variation via

$$\begin{aligned} \delta_{g,\rho}: \quad & \text{variation of state } \rho_B \text{ and geometry } g \\ & \text{that keeps the volume } V \text{ of } B \text{ fixed.} \end{aligned} \tag{12.2.10}$$

Under this variation, the area A at fixed V changes by $\delta_{g,\rho}A|_V$ and the quantum state ρ_B on B changes by $\delta_{g,\rho}\rho_B$. Moreover, there is a change in the entanglement entropy $\delta_{g,\rho}S_B$ as well as a change in the expectation value of the

modular Hamiltonian, $\delta_{g,\rho}\langle K\rangle$, which is in general a highly non-local quantity. The modular Hamiltonian does not correspond a priori to any intuitive sense of energy; it is just an operator one can define using the reduced density matrix. Note that all of the above variations are *not* dynamical variations that occur with time; rather, we are considering varying the entire history of the configuration, examining the consequences for various quantities for infinitesimally separated configurations of geometry and fields. For example, for a CFT, plugging in the Killing field (12.2.2) into Eq. (12.2.9) and requiring a sufficiently small causal diamond $\ell \ll L_T$, where L_T is the characteristic length scale of changes in $T_{\mu\nu}$, we have the modular energy

$$\delta_{g,\rho}\langle K_{\rm CFT}\rangle = \frac{2\pi}{\hbar}\frac{\Omega_{D-2}\ell^D}{D^2-1}\delta_{g,\rho}\langle T_{00}\rangle, \tag{12.2.11}$$

where $\Omega_{D-2} = 2\pi^{(D-1)/2}/\Gamma[(D-1)/2]$ is the area of the unit $(D-2)$-sphere.

We are now ready to state the postulates of the holographic gravity theory given in Ref. [60]. They are as follows:

1. **Entanglement separability.** The entropy S_B can be written as a simple sum $S_{\rm UV} + S_{\rm IR}$, where UV and IR denote the entanglement entropies in the UV (quantum gravitational) and IR (quantum field-theoretic) degrees of freedom. Equivalently, the quantum mutual information $I_B = S_{\rm UV} + S_{\rm IR} - S_B$ is negligible. That is, there is minimal entanglement among degrees of freedom at widely separated energy scales.[12.3]

2. **Equilibrium condition.** The entanglement entropy of the causal diamond is stationary with respect to variations of the state and metric, i.e.,

[12.3] This formulation of postulate 1. is actually somewhat stronger than necessary; for holographic gravity it is sufficient that merely the entropy variation $\delta_{g,\rho}S_B$ factorize as in Eq. (12.2.12). However, the justification for this weaker version of postulate 1. will ultimately be the same as the stronger version we state above.

$$\delta_{g,\rho} S_B = \delta_{g,\rho} S_{\rm UV} + \delta_{g,\rho} S_{\rm IR} = 0, \tag{12.2.12}$$

and the geometry of the causal diamond is that of a maximally symmetric spacetime (Minkowski, de Sitter, or anti-de Sitter).

3. **Area-entropy relation.** The variation of the UV entropy of the causal diamond is proportional to its area change at fixed volume,

$$\delta_{g,\rho} S_{\rm UV} = \eta\, \delta_{g,\rho} A|_V, \tag{12.2.13}$$

for some universal constant η. That is, δS satisfies a local, bulk version of holography. This is Jacobson's generalization of the area law for black hole entropy and is the crucial substantive assumption underlying holographic gravity.

4. **Modular energy: CFT form.** The modular energy, defined to be the variation in the expectation value of the modular Hamiltonian, for an arbitrary quantum field theory is given by the form in Eq. (12.2.11), possibly modified by some scalar operator X,

$$\delta_{g,\rho} \langle K \rangle = \frac{2\pi}{\hbar} \frac{\Omega_{D-2} \ell^D}{D^2 - 1} \delta_{g,\rho} \left(\langle T_{00} \rangle + \langle X \rangle g_{00} \right). \tag{12.2.14}$$

While the first three postulates are assumptions about ultraviolet behavior, the fourth is strictly an infrared statement and we will argue that, in its null-limit form, it can be derived rather than postulated. Note that in postulate 3., we expect $\eta = 1/4G\hbar$, the same constant that appears in the Bekenstein-Hawking formula [299].

Reference [60] shows how postulates 1. through 4. can be used to derive the Einstein equations. Our purpose in this section is to illustrate how some of these postulates can be justified rigorously, rather than taken as assumptions.

While we leave the geometric details of how the postulates imply the Einstein equations to Ref. [60], we sketch the main points. First, writing as usual

$$S_{\rm IR} = -\mathrm{Tr}\,\rho_{\rm IR}\log\rho_{\rm IR}, \tag{12.2.15}$$

we have the entanglement first law [314]

$$\begin{aligned}\delta_{g,\rho}S_{\rm IR} &= -\mathrm{Tr}\left[(\delta_{g,\rho}\rho_{\rm IR})\log\rho_{\rm IR}\right] - \mathrm{Tr}\left(\rho_{\rm IR}\rho_{\rm IR}^{-1}\delta_{g,\rho}\rho_{\rm IR}\right)\\ &= \mathrm{Tr}\left(K\delta_{g,\rho}\rho_{\rm IR}\right)\\ &= \delta_{g,\rho}\langle K\rangle.\end{aligned} \tag{12.2.16}$$

Further, the area variation at constant V for a maximally symmetric spacetime in which $G_{\mu\nu} = -f\,g_{\mu\nu}$ for some arbitrary constant f is

$$\delta_{g,\rho}A|_V = -\frac{\Omega_{D-2}\ell^D}{D^2-1}(G_{00} + f\,g_{00}). \tag{12.2.17}$$

Equating Eqs. (12.2.16) and (12.2.14) via postulate 4., setting Eq. (12.2.17) to $\delta S_{\rm UV}/\eta$ via postulate 3., and then putting everything together via postulates 1. and 2., we have

$$0 = \delta_{g,\rho}S_B = \frac{\Omega_{D-2}\ell^D}{D^2-1}\left[-\eta(G_{00} + f\,g_{00}) + \frac{2\pi}{\hbar}\delta_{g,\rho}\left(\langle T_{00}\rangle + \langle X\rangle g_{00}\right)\right]. \tag{12.2.18}$$

Rearranging and requiring that this relation hold for all possible spatial slicings (i.e., in arbitrary reference frames) requires

$$R_{\mu\nu} - \frac{1}{2}Rg_{\mu\nu} + f\,g_{\mu\nu} = \frac{2\pi}{\hbar\eta}\delta_{g,\rho}\left(\langle T_{\mu\nu}\rangle + \langle X\rangle g_{\mu\nu}\right). \tag{12.2.19}$$

Now, since we must have $\nabla^\mu T_{\mu\nu} = 0$ for energy-momentum conservation, but $\nabla^\mu R_{\mu\nu} = \nabla_\nu R/2$ by the Bianchi identity, f can be identified as $2\pi\delta_{g,\rho}\langle X\rangle/\hbar\eta + \Lambda$ for arbitrary constant Λ, yielding Einstein's equation in semiclassical terms,

$$R_{\mu\nu} - \frac{1}{2}Rg_{\mu\nu} + \Lambda g_{\mu\nu} = \frac{2\pi}{\hbar\eta}\,\delta_{g,\rho}\langle T_{\mu\nu}\rangle = 8\pi G\,\delta_{g,\rho}\langle T_{\mu\nu}\rangle, \tag{12.2.20}$$

where in the final equality we plugged in $\eta = 1/4G\hbar$ as expected for consistency with the Bekenstein-Hawking formula. Note that the $\delta_{g,\rho}\langle T_{\mu\nu}\rangle$ appearing on the right-hand side is really just the expectation value of the energy-momentum

tensor under consideration, since without the variation, i.e., in vacuum, the causal diamond is assumed to be described by a maximally symmetric spacetime with vanishing $T_{\mu\nu}$.

The way in which the Einstein equation arose in the above derivation was by the imposition of a relationship between the change of entanglement entropy and area for variations over the spacetime configuration and quantum state. It is not a dynamical constraint within a single solution, but rather a relationship between infinitesimally separated spacetime histories and geometries. Mathematically, how this constraint leads to the Einstein equation is the same as how the Einstein equation was derived [212, 300] in the context of AdS/CFT via the Ryu-Takayanagi formula [132, 272]. That is, AdS/CFT itself, in Refs. [212, 300], provides another realization of holographic gravity. The version of the theory in Ref. [60] attempts wider applicability, by applying holographic formulas to causal diamonds in an arbitrary spacetime. It is therefore crucial to investigate the extent to which the postulates of the theory can be justified. We conduct such an investigation in the next subsection, providing a nontrivial check of the health of HG.

12.2.3 Justifying the Assumptions of Holographic Gravity

Postulates 1. through 3. above deal with the ultraviolet degrees of freedom in the ultimate theory of quantum gravity. Hence, they either must be taken as axioms of the theory or shown to be true in a more general ultraviolet completion of gravity (e.g., through holography and string theory). Despite this ultraviolet character, there are motivations for postulates 1. through 3., which we will briefly mention. More importantly, we offer a derivation of a null-limit version of postulate 4., allowing it to be removed as an independent assumption in HG.

Postulate 1., requiring minimal entanglement between infrared and ultraviolet degrees of freedom, is a basic feature of effective field theory [315], so the first postulate amounts to the assertion that effective field theory is (at least approximately) valid for the field-theoretic degrees of freedom. That is, for renormalization group flow to work in the usual manner, we require a decoupling between the low- and high-momentum states. We do not expect significant mutual information between the low-energy degrees of freedom in a Wilsonian effective action and those in the ultraviolet completion. This was explicitly found to be the case for interacting scalar quantum field theories in Ref. [315].

Postulate 2. is really the entropic foundation of the theory, being the assertion of a condition on the spacetime geometry that will ultimately lead to the Einstein equation. In essence, postulate 2. is the assertion that the vacuum should look as simple as possible, namely, that a small region should be well described by a Gibbs state. For a fixed energy expectation value, the Gibbs state has the maximum entropy, so $\delta S_B = 0$. Moreover, for the Gibbs distribution, expectation values of quantum mechanical quantities related to the entanglement entropy map onto those from classical thermodynamics [316]. Viewed in this sense, the causal diamond represents a canonical ensemble [60], with fixed degrees of freedom and volume. Hence, classically, its entropy for a given expectation value of $T_{\mu\nu}$ is maximized in equilibrium. The requirement that the causal diamond be described by a maximally symmetric spacetime means that there is not power in spacetime fluctuations at arbitrarily small scales. If this were not the case, then introducing fluctuations would produce a large backreaction that would spoil the equilibrium condition. The content of postulate 2. is therefore the assertion that the semiclassical Einstein equations hold if and only if the causal diamond is in thermodynamic equilibrium.

Postulate 3. is related to the Ryu-Takayanagi relation [132, 272], with which Refs. [212, 300] derived the Einstein equations in a holographic context in a manner closely related to that of Ref. [60]. References [212, 300] can be regarded as another example of HG, in which bulk gravitation is again found to be dual to a constraint on entanglement entropy in some boundary degrees of freedom. The boundaries of the causal diamond can be viewed as the Rindler horizons of a set of appropriately accelerating observers. The area of ∂B is just the area of this horizon. Postulate 3. does not require assigning a change in entropy with time to a dynamical change in area. Rather, it just requires identifying the area of the causal diamond with the entanglement entropy and then doing this for an entire family of infinitesimally separated causal diamond configurations. The motivations for assigning an entropy to an area for this apparent horizon in the first place were discussed in Sec. 12.2.1.

Postulate 4. is of a different character. Unlike the ultraviolet-dependent postulates 1. through 3., postulate 4. is an assertion about the form of the modular Hamiltonian for the field-theoretic degrees of freedom. Thus, postulate 4. is amenable to analysis and, as a consistency test of the holographic gravity of Ref. [60], we can investigate whether postulate 4. can be justified, rather than taken as an assumption. A holographic justification of postulate 4. for spacelike slicing was considered in Ref. [317], in which the subtleties of the construction in Ref. [60] for operators of particular conformal dimensions is discussed in detail. However, we will show that postulate 4. may be justified more simply in the null limit by using the conformal symmetry of the causal diamond and the lightsheet results of Ref. [258].

A priori, postulate 4. suffers from two potential weaknesses. First, it is unclear why the modular energy of a generic quantum field theory should take the form appropriate for a CFT. Second, Ref. [60] derived the Einstein equations

only for small variations about the vacuum in the field-theoretic density matrix ρ_{IR}, for which the entanglement first law (12.2.16) holds. However, small variations to the geometry, in which gravitational backreaction is negligible, do not necessarily correspond to small variations in ρ_{IR}. For example, two massive particles in a Bell pair state certainly gravitate, but their long-range entanglement does not correspond to a small perturbation about the vacuum state of ρ_{IR}. Thus, the question remains of how to retain the success of HG in obtaining the Einstein equation for large changes to the quantum state without using the entanglement first law. Both of these challenges can be addressed using recent results proven in quantum field theory.

To address the second issue, we consider the computation of the modular energy and entanglement entropy for an interacting CFT in $D > 2$, which was computed for a null slab in Refs. [258, 265]. For an arbitrary state ρ_{IR} defined on a spatial region (for example, one of the spatial slices of our causal diamond), we can define the Casini entropy

$$\Delta S = -\mathrm{Tr}\,\rho_{\mathrm{IR}} \log \rho_{\mathrm{IR}} + \mathrm{Tr}\,\sigma_{\mathrm{IR}} \log \sigma_{\mathrm{IR}}, \tag{12.2.21}$$

which is just the vacuum-subtracted von Neumann entropy, and the modular energy,

$$\Delta K = \mathrm{Tr}\, K\rho_{\mathrm{IR}} - \mathrm{Tr}\, K\sigma_{\mathrm{IR}}, \tag{12.2.22}$$

where the modular Hamiltonian K is defined as in Eq. (12.2.7) but with respect to the vacuum density matrix σ_{IR}. Note that in the limit in which the field-theoretic density matrix for this state is infinitesimally close to the vacuum state σ_{IR}, we have $\Delta S \to \delta S$ and $\Delta K \to \delta K$ and the entanglement first law guarantees $\delta S = \delta K$. Using the replica trick [318, 319] to compute the nth Rényi entropy for an arbitrary spatial region by inserting defect operators on the boundaries, Ref. [258] shows through an argument involving the operator

product expansion in the null limit that the only operators that can contribute to ΔK or ΔS are single-copy scalar operators with twist τ in the range

$$\frac{1}{2}(D-2) < \tau \leq D-2. \tag{12.2.23}$$

For spin-zero operators, τ is just the scaling dimension. By single-copy, we mean that the operator appears inside just one of the copies of the CFT in the replica trick; in that case, the contribution of this operator to the entanglement entropy is proportional to the expectation value of the operator inside a single copy of the CFT [258]. That is, single-copy operators contribute linearly in the density matrix to $S_{\rm IR}$.

Finally, the modular Hamiltonian is the unique operator on B that matches $S_{\rm IR}$ at linear order for arbitrary perturbations of the density matrix. Thus, single-copy operators contribute equally to ΔS and ΔK, so taking the null limit of any spatial surface and computing ΔS and ΔK, we have

$$\Delta S = \Delta K \qquad \text{[null limit]}. \tag{12.2.24}$$

One can show that, evaluated on any fixed spatial slice, $\Delta K - \Delta S = D(\rho_{\rm IR}|\sigma_{\rm IR})$, the relative entropy between the state and the vacuum, which is always non-negative. However, in the null limit, Ref. [258] showed that in an interacting conformal field theory, no operators in the algebra can be localized to a null surface, which allows the excited state and the vacuum to differ while remaining indistinguishable. Moreover, the null limit is sensitive only to the UV structure of the theory. For quantum field theories with an interacting UV fixed point, Ref. [258] thus showed that the $\Delta S = \Delta K$ result of Eq. (12.2.24) continues to hold (provided the quantum field theory does not have finite wave function renormalization, as for, e.g., superrenormalizable theories). The result is therefore quite general. Equation (12.2.24) applies to *any* quantum state that backreacts weakly on the geometry and thus strengthens the argument for

HG in Ref. [60]. No longer is it necessary to rely on the entanglement first law (12.2.16) and consider only small perturbations about the vacuum density matrix in postulate 4.; we are now free to consider arbitrary states.

To mitigate the first issue raised regarding postulate 4., namely the question of why the modular energy for a generic quantum field theory should be related to that of a CFT, Ref. [60] considered a quantum field theory with a UV fixed point and required that the size of the causal diamond be smaller than every length scale in the quantum field theory, i.e.,

$$\ell \ll \frac{1}{\max_i m_i}, \tag{12.2.25}$$

where m_i are the masses of states in the quantum field theory. That is, we are required to take the causal diamond to be smaller than the cutoff of the quantum field theory, $\Lambda_{\text{UV}} \gg \ell$. Naively, this leads to doubtful consistency of treating the spacetime semiclassically; we do not want to be required to take the causal diamond to be Planck-scale. However, we typically expect the scale of a perturbative UV completion of gravity to be parametrically smaller than the Planck scale, as indeed is the case in string theory [30].

In any case, we can dramatically relax the stipulation of Eq. (12.2.25) by evaluating ΔK in the null limit. Let us choose a sequence of spacelike slices B_ξ through the diamond, $\xi \in [0, 1]$, defined by the orbit of ζ in Eq. (12.2.2), where we start with $B_0 = B$ and end with B_1, the upper null surface of the diamond. Now, ζ is not a member of the Poincaré group; it is a conformal Killing vector and in particular contains a dilation. The proper distance across B_ξ tends to zero as we send $\xi \to 1$, that is, as we flow along ζ. Acting with ζ on a given field configuration with small momentum on B_0 takes us to larger and larger momenta and we experience renormalization group flow as we move through different values of ξ. For a CFT, ζ acts trivially, but for a general interacting quantum field theory with a UV fixed point [60, 258], flow along ζ means are

probing higher and higher energy scales within the theory, eventually reaching a regime in which the CFT approximation is valid. Thus, Eqs. (12.2.9) and (12.2.14) become in the null limit

$$\Delta K = \frac{2\pi}{\hbar}\frac{\Omega_{D-2}\ell^D}{D^2-1}T_{uu}, \tag{12.2.26}$$

where we still assume that $T_{\mu\nu}$ varies with a length scale L_T larger than ℓ. Moreover, we can write the area variation (12.2.17) in terms of its null components as ζ lines up with ∂_u in the null limit. Hence, Eq. (12.2.18) still applies, but for the uu components. Following the logic through as before, we again obtain the Einstein equation

$$G_{\mu\nu} + \Lambda g_{\mu\nu} = 8\pi G T_{\mu\nu}. \tag{12.2.27}$$

We therefore see the infrared assumption underlying holographic gravity, that the modular energy takes the CFT form given in Eq. (12.2.14), need not be separately postulated, but can be justified by examining the null limit. The null surfaces themselves are not special or preferred, but the use of the null limit rendered tractable the explicit computation of the entanglement entropy and modular energy for generic interacting quantum field theories with an ultraviolet fixed point. Moreover, we seem to have a specific and self-consistent formulation of what kind of entropy we are talking about in holographic gravity: the Casini entropy evaluated on the null boundary of a small causal diamond.

12.3 Thermodynamic Gravity

In Sec. 12.2, we demonstrated that the holographic gravity of Ref. [60] can be made well defined, putting its axioms on a more solid footing. In this section, we turn to the question of whether the same can be done for the thermodynamic gravity of Ref. [59]. We will argue that there does not exist any self-consistent definition of entropy in this approach.

In the cases relating entropy and area that we discussed in Sec. 12.2.1, the area is a constant along the horizon. To work our way toward a truly dynamical theory of gravity, we must be able to handle more general cases, including time-dependent spacetimes. Holographic gravity accommodates this requirement by varying the spacetime history: in that case, general relativity can be shown to be equivalent to constraints relating the variation in entanglement entropy and area of a small causal diamond. However, the HG approach does not allow gravity to truly emerge as an equation of state, since the area and entropy variations are not dynamical changes within a single background spacetime. TG takes the other approach. That is, we can start with the null generators of a local Rindler horizon, but the corresponding cross-sectional area will generally change with time. Thermodynamic gravity [59] therefore posits that the *change* in entropy behind such a horizon is proportional to the *change* in that area. This is a natural generalization of the area law itself. We will argue that it is then hard to associate this quantity with a well-defined entropy of any particular local system.

12.3.1 Formulation of Thermodynamic Gravity

Consider an arbitrary spacetime and identify some point p. Restrict to a sufficiently small region such that we can define a spacelike foliation with respect to a time coordinate t. Our point p is located at time coordinate t_1 on a spacelike codimension-one hypersurface Σ_1. Choose a codimension-two approximately-flat spacelike surface $\mathcal{P}_1$ containing p. Approximate flatness means that the null congruences normal to $\mathcal{P}_1$ have vanishing expansion θ and shear $\sigma_{\mu\nu}$ at p to first order in the distance from p. Fix a closed orientable smooth spacelike codimension-two surface $\mathcal{B}_1$ containing $\mathcal{P}_1$ and choose a future-directed inward null direction normal to $\mathcal{B}_1$, which defines a null congruence originating from $\mathcal{B}_1$. Denote the spacelike region of Σ_1 that lies inside $\mathcal{B}_1$

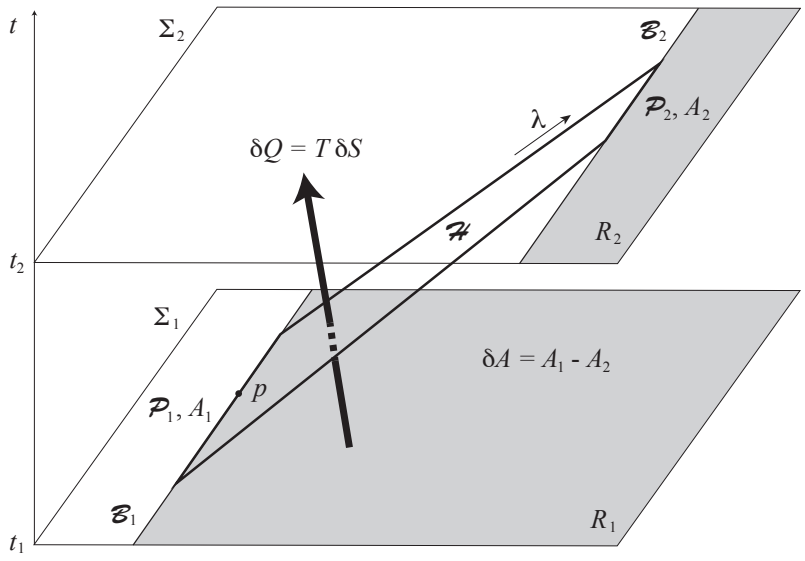

Figure 12.2. Spacetime diagram of the flux through a segment $\mathcal{H}$ of a lightsheet. Starting with some point p, we fix an approximately-flat spacelike surface $\mathcal{P}_1 \ni p$ and a boundaryless surface $\mathcal{B}_1 \supset \mathcal{P}_1$. Consider future-directed inward null geodesics orthogonal to $\mathcal{P}_1$, with affine parameter λ. Flowing along the geodesics by some fixed parameter value, the area A_1 of the initial surface element $\mathcal{P}_1$ evolves into a new area A_2, which we can use to define the area decrement $\delta\mathcal{A} \equiv A_1 - A_2$. An amount of heat δQ passes through $\mathcal{H}$, which by the Clausius relation is equal to $T\delta S$. The regions R_1 and R_2 denote the parts of the spacelike hypersurfaces Σ_1 and Σ_2 that lie inside the spacelike codimension-two surfaces $\mathcal{B}_1$ and $\mathcal{B}_2$, respectively.

by R_1. Choose an affine parameter λ along the congruence, with tangent vector $k^\mu = (\mathrm{d}/\mathrm{d}\lambda)^\mu$, letting λ equal zero at p and increase toward the future. Points in the congruence make up the "lightsheet" $\mathcal{H}$ emanating from $\mathcal{P}_1$. At a not-much-later time t_2, the intersection of the null congruence from $\mathcal{B}_1$ with a spacelike hypersurface Σ_2 defines a spacelike codimension-two surface $\mathcal{B}_2$, such that $\mathcal{P}_1$ evolves to $\mathcal{P}_2$. The region inside $\mathcal{B}_2$ is denoted by R_2. The setup is portrayed in Fig. 12.2.

The lightsheet $\mathcal{H}$ is a horizon in the sense that it serves as a local Rindler horizon for appropriately accelerating observers. (In Ref. [59], the construction was formulated over the past horizon instead of the future horizon, but this distinction makes no difference to our arguments.) We can define an

approximate boost Killing vector $\chi^\mu = \kappa\lambda k^\mu$, where κ is the acceleration of the associated Rindler trajectory. The surface element for the local Rindler horizon is $\mathrm{d}\Sigma^\mu = k^\mu \mathrm{d}\lambda\, \mathrm{d}\mathcal{A}$, where $\mathrm{d}\mathcal{A}$ is the codimension-two spacelike cross-sectional area element. This can be used to define a heat flux across the lightsheet

$$\delta Q \equiv \int_{\mathcal{H}} T_{\mu\nu}\chi^\mu \mathrm{d}\Sigma^\nu = \kappa \int_{\mathcal{H}} T_{\mu\nu} k^\mu k^\nu \lambda \,\mathrm{d}\lambda\, \mathrm{d}\mathcal{A}. \tag{12.3.1}$$

Viewing our system as the set of degrees of freedom on R_1 in Fig. 12.2, δQ defines the heat leaving the system through $\mathcal{H}$. The temperature associated with this process is just the Unruh temperature [310] for the Rindler trajectory, $T = \hbar\kappa/2\pi$.

The area decrement of the lightsheet as δQ flows through it is

$$\delta\mathcal{A} \equiv A_1 - A_2 = -\int_{\mathcal{H}} \theta \,\mathrm{d}\lambda\, \mathrm{d}\mathcal{A}, \tag{12.3.2}$$

where A_1 is the initial area of the codimension-two surface $\mathcal{P}_1$, $A_1 = \int_{\mathcal{P}_1} \mathrm{d}\mathcal{A}$, and A_2 is the area of the codimension-two surface $\mathcal{P}_2$ at the other end of $\mathcal{H}$. The expansion θ is defined to be $\theta \equiv \nabla_\mu k^\mu$. Carefully treating the range of integration of λ will play an important role in our discussion in Sec. 12.3.2.

Having made these preliminary definitions, we are ready to state the assumptions of thermodynamic gravity. They are the following:

1. **Clausius relation.** There exists an entropy change δS associated with the flow of heat through the lightsheet $\mathcal{H}$, which in local thermodynamic equilibrium is given by

$$\delta S = \delta Q/T. \tag{12.3.3}$$

2. **Local holography.** For any lightsheet $\mathcal{H}$ of the form shown in Fig. 12.2, the entropy change δS is proportional to the change in area $\delta\mathcal{A}$ with some universal constant η,

$$\delta S = \eta\, \delta\mathcal{A}. \tag{12.3.4}$$

Note that the use of the Clausius relation (12.3.3) implies that the entropy δS under consideration should correspond to some notion of entropy for a system that can be locally defined. The local holography assumption, meanwhile, is motivated by black hole thermodynamics, upon which entropic gravity is based. Therefore, we should expect that η is the same coefficient as in the Bekenstein-Hawking formula [299], $1/4G\hbar$. Were we to find that $\eta \neq 1/4G\hbar$ is required for consistency with Eqs. (12.3.3) and (12.3.4) and Einstein's equation, this would undermine the original motivation for TG. This is the problem we will uncover in Sec. 12.3.2.

Putting these assumptions together, we can derive Einstein's equation. First, from the Raychaudhuri equation,

$$\frac{\mathrm{d}\theta}{\mathrm{d}\lambda} = -\frac{1}{D-2}\theta^2 - \sigma_{\mu\nu}\sigma^{\mu\nu} - R_{\mu\nu}k^\mu k^\nu, \tag{12.3.5}$$

which is just a geometric statement in terms of the expansion, shear, and Ricci tensor $R_{\mu\nu}$, Ref. [59] writes $\theta = -\lambda R_{\mu\nu}k^\mu k^\nu$ for a small segment of the lightsheet and inserts this result into Eq. (12.3.2) to obtain

$$\delta\mathcal{A} = \int_{\mathcal{H}} R_{\mu\nu}k^\mu k^\nu \lambda \,\mathrm{d}\lambda \,\mathrm{d}\mathcal{A}. \tag{12.3.6}$$

Using local holography (12.3.4) and the Clausius relation (12.3.3) to equate this to $\delta Q/T$, one can invoke the freedom in the choice of k^μ to equate the integrands, obtaining

$$\eta(R_{\mu\nu} + f\, g_{\mu\nu}) = \frac{2\pi}{\hbar}T_{\mu\nu} \tag{12.3.7}$$

for some scalar quantity f. Since we must have $\nabla^\mu T_{\mu\nu} = 0$ for energy-momentum conservation, but $\nabla^\mu R_{\mu\nu} = \nabla_\nu R/2$ by the Bianchi identity, f can be identified, yielding Einstein's equation,

$$R_{\mu\nu} - \frac{1}{2}R\, g_{\mu\nu} + \Lambda g_{\mu\nu} = \frac{2\pi}{\hbar\eta}T_{\mu\nu} = 8\pi G\, T_{\mu\nu}, \tag{12.3.8}$$

where Λ is the cosmological constant. We find that η must indeed be equal to

$1/4G\hbar$, as expected for consistency with the Bekenstein-Hawking formula.

We see that the assumptions of the Clausius relation (12.3.3) and local holography (12.3.4) are, together, sufficient to derive Einstein's equation, at least up to a normalization. Less clear is the nature of the quantity δS—in particular, precisely what this is supposed to be the entropy *of*. Formally, the only role of δS in this derivation is to motivate equating $\eta\,\delta\mathcal{A}$ with $\delta Q/T$; once that happens, δS disappears from the discussion. But if we were simply to assume $\eta\,\delta\mathcal{A} = \delta Q/T$ from the start, that would be tantamount to assuming Einstein's equation. The substantive content of TG, therefore, rests on the existence of a consistent and well-defined local construction for the entropy δS associated with lightsheet segments anywhere in spacetime. We now turn to an investigation of what that construction might be.

12.3.2 Entanglement Entropy of a Null Region

Some form of the von Neumann entanglement entropy is a natural candidate for the quantity δS that plays a crucial role in TG. We first need to specify the precise system whose entanglement entropy we are calculating. Factors of Hilbert space are usually associated with regions of spacelike surfaces, but local holography refers to the entropy associated with part of a null surface. The simplest option would be to introduce some spacelike slicing, zoom in on a small neighborhood so that the spacetime looks approximately static, and compute the von Neumann entropy on the small spacelike region; subsequently, one could enforce local holography on the small lightsheet through which the orthogonal timelike congruence originating from the small spacelike region passes. However, this prescription does not prove suitable: while the von Neumann entropy is subextensive, energy-momentum is extensive. That is, considering two adjacent regions A and B, we have $S_{AB} \leq S_A + S_B$, with strict inequality if A and B are entangled; however, the masses of A and B, and hence the concomitant

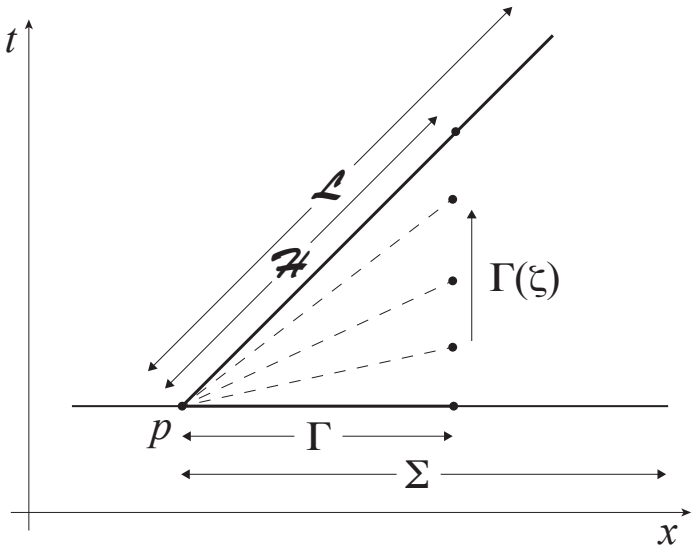

Figure 12.3. A finite lightsheet $\mathcal{H}$ considered as the null limit of a parameterized collection of spacelike regions $\Gamma(\zeta)$. The large spacelike region Σ maps to the large null surface $\mathcal{L}$. The affine parameter generating $\mathcal{H}$ runs from 0 to ϵ.

first-order area decrements of a lightsheet passing through them, add linearly. Thus, the use of the von Neumann entropy on spacelike surfaces cannot provide a consistent formulation of thermodynamic gravity.

We therefore turn to the null limit. Consider a spacelike region Σ, with a point $p \in \mathcal{P}_1$ on its boundary, as shown in Fig. 12.3. It contains a smaller spacelike region Γ with p also on its boundary. The large null surface to the future of Σ is labeled $\mathcal{L}$ and a small lightsheet $\mathcal{H}$, as defined in Sec. 12.3.1, can be thought of as the null limit of a series of spacelike regions $\Gamma(\zeta)$. There are then two different ways to associate an entropy with $\mathcal{H}$: *i*) the entanglement entropy associated with the region itself and *ii*) the difference in entanglement entropies between those of the large null surfaces $\mathcal{L}$ and $\mathcal{L} - \mathcal{H}$, which emanate from $\mathcal{P}_1$ and $\mathcal{P}_2$, respectively. We will consider each possibility in turn.

Let us first see whether the entropy appearing in TG could be the entanglement entropy associated with the region $\mathcal{H}$. Let ρ_Σ be the density matrix of the system on the spacelike region Σ and let σ_Σ be the vacuum density

matrix. Let $\sigma_\Gamma \equiv \mathrm{Tr}_{\Sigma-\Gamma}\sigma_\Sigma$ and $\rho_\Gamma \equiv \mathrm{Tr}_{\Sigma-\Gamma}\rho_\Sigma$. We are immediately forced to identify some way to regulate the von Neumann entropy, which naively diverges. Consider the vacuum von Neumann entropy in the null limit, $\lim_{\Gamma\to\mathcal{H}} S(\sigma_\Gamma)$. If we simply impose an ultraviolet cutoff, the entanglement entropy $S(\sigma_\Gamma)$ associated with a vacuum region is still large [261, 312], going as $\mathcal{A}/\epsilon^2$, where $\mathcal{A}$ is the area of the boundary of Γ and ϵ is the cutoff length. By the local holography postulate (12.3.4), we must have $\delta S = \eta\,\delta\mathcal{A}$, where $\delta\mathcal{A}$ is the area decrement along $\mathcal{H}$, which must vanish in the Minkowski vacuum. While the details of a UV cutoff may have bearing on the renormalization of Newton's constant (see Ref. [320] and references therein) and therefore of η, no such effect could reconcile a finite value of δS with an exactly vanishing $\delta\mathcal{A}$. Thus, we cannot use the UV-regulated von Neumann entropy in the null limit as δS in entropic gravity, since doing so would require violation of either the postulate of local holography or flatness of the vacuum spacetime. We must therefore adopt the prescription of Casini [61], subtracting the entanglement entropy associated with the vacuum as in Eq. (12.2.21), producing the appropriate regulated version of the von Neumann entropy that vanishes in vacuum.

We compute the Casini entropy ΔS_Γ of the small spacelike region as the difference of the von Neumann entropies for ρ_Γ and σ_Γ as in Eq. (12.2.21), $\Delta S_\Gamma \equiv S(\rho_\Gamma) - S(\sigma_\Gamma)$, and then take the null limit to define the entropy on the small lightsheet, $\Delta S_\mathcal{H} \equiv \lim_{\Gamma\to\mathcal{H}}\ \Delta S_\Gamma$. Next, let us define a modular Hamiltonian K_Γ on Γ via

$$\sigma_\Gamma \equiv \frac{e^{-K_\Gamma}}{\mathrm{Tr}\ e^{-K_\Gamma}} \tag{12.3.9}$$

and use this to define ΔK_Γ as in Eq. (12.2.22). Despite the non-locality of K, the modular energy becomes more tractable in the null limit, $\Delta K_\mathcal{H} \equiv \lim_{\Gamma\to\mathcal{H}} \Delta K_\Gamma$, as we saw in Sec. 12.2.3.

It was shown in Refs. [258, 265] for interacting quantum field theories that

ΔS_Γ and ΔK_Γ become equal as the null limit is taken and, in particular,

$$\Delta S_{\mathcal{H}} = \Delta K_{\mathcal{H}} = \frac{2\pi}{\hbar} \int \mathrm{d}\mathcal{A} \int_0^\epsilon \mathrm{d}\lambda \, g(\lambda, \epsilon) T_{\mu\nu} k^\mu k^\nu, \tag{12.3.10}$$

where $g(\lambda, \epsilon)$ is a real function whose precise values depend on the interacting quantum field theory being considered. Note that $g(\lambda, \epsilon)$ is not automatically theory-independent, as in the causal diamond case: the causal diamond was related by a global conformal transformation (12.2.8) to the Rindler wedge, while this is not so for the lightsheet $\mathcal{H}$. However, Ref. [258] showed that $g(\lambda, \epsilon)$ is computable in particular cases and moreover satisfies certain general properties for all interacting quantum field theories, which will be sufficient for our purposes.

The function $g(\lambda, \epsilon)$, whose properties we discuss in detail below, plays a crucial role here. Equations analogous to Eq. (12.3.10) appear as expressions for the heat transfer in Refs. [59, 321], but with $g(\lambda, \epsilon)$ replaced simply by λ. [This similarity suggests that we should view Eq. (12.3.10) as corresponding to the Clausius relation, indicating that this formulation of the entropy is appropriate for application to TG.] For the Rindler Hamiltonian, which inspires this form, λ is perfectly appropriate for a semi-infinite lightsheet, but we are now computing the entropy for the *finite* segment of lightsheet $\mathcal{H}$, for which the entropy takes the form of Eq. (12.3.10), as shown in Refs. [258, 265]. That makes all the difference: $g(\lambda, \epsilon)$ initially increases as λ, but then decreases as $\epsilon - \lambda$ at the other end of the segment. As a result, the integral in Eq. (12.3.10) differs from the Rindler Hamiltonian by a theory-dependent constant factor of order unity. This discrepancy implies that we cannot simultaneously choose our normalization so as to correctly recover Newton's constant in both Einstein's equation and in the area-entropy formula.

Reference [258] derived a number of properties that the function $g(\lambda, \epsilon)$ appearing in Eq. (12.3.10) must obey, amounting essentially to the requirement

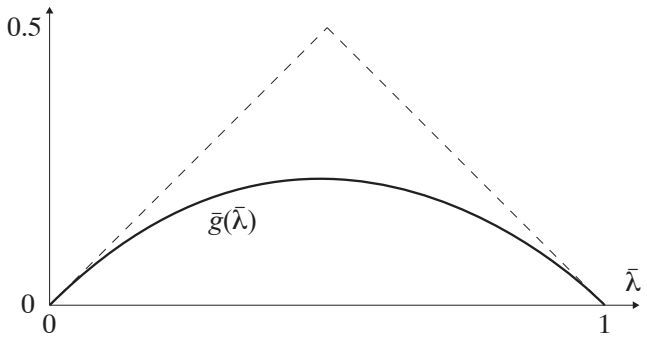

Figure 12.4. Schematic form of the function $\bar{g}(\bar{\lambda})$, proportional to the $g(\lambda, \epsilon)$ used in the expression for the null Casini entropy in Eq. (12.3.10). It is symmetric between $\bar{\lambda} = 0$ and $\bar{\lambda} = 1$, with slope between 1 and -1 and a negative second derivative everywhere.

that it have the form illustrated in Fig. 12.4. More specifically, defining $\bar{\lambda} \equiv \lambda/\epsilon \in [0, 1]$, we have $g(\lambda, \epsilon) = \epsilon \bar{g}(\bar{\lambda})$, with $\bar{g}(\bar{\lambda}) = \bar{g}(1 - \bar{\lambda})$, and

$$\begin{aligned} \bar{g}(\bar{\lambda}) &\to \bar{\lambda} \qquad \text{for } \bar{\lambda} \to 0, \\ \bar{g}(\bar{\lambda}) &\to 1 - \bar{\lambda} \quad \text{for } \bar{\lambda} \to 1. \end{aligned} \tag{12.3.11}$$

Putting together the required properties of $\bar{g}$, Ref. [258] showed that $\left|\mathrm{d}\bar{g}/\mathrm{d}\bar{\lambda}\right| \leq 1$. Note in particular that the integral $\int_0^1 \mathrm{d}\bar{\lambda}\, \bar{g}(\bar{\lambda})$ is less than $1/4$.

Now let us consider the area variation of $\mathcal{H}$. As in Ref. [59], we can choose $\mathcal{H}$ such that θ and $\sigma_{\mu\nu}$ vanish at first order near p. We can evaluate the change in the cross-sectional area of $\mathcal{H}$ by integrating the Raychaudhuri equation (12.3.5) for a finite lightsheet, keeping careful track of the ranges of integration. We find that the area decrement along $\mathcal{H}$ is

$$\Delta\mathcal{A} = -\int \mathrm{d}\mathcal{A} \int_0^\epsilon \mathrm{d}\lambda\, \theta(\lambda) = \int \mathrm{d}\mathcal{A} \int_0^\epsilon \mathrm{d}\lambda \int_0^\lambda \mathrm{d}\hat{\lambda}\, R_{\mu\nu}(\hat{\lambda}) \hat{k}^\mu \hat{k}^\nu. \tag{12.3.12}$$

We can now test whether the null Casini entropy $\Delta S_{\mathcal{H}}$, which is the regularized von Neumann entropy from Eq. (12.2.21) evaluated in the null limit, can be the basis of a consistent formulation of TG. First, we need only consider the limit of a very small lightsheet, since we wish only to recover the local equations of motion, i.e., Einstein's equation. That is, we can take

$\Delta\mathcal{A}$ and $\Delta S_{\mathcal{H}}$ in Eqs. (12.3.12) and (12.3.10) in the limit of very small ϵ and cross-sectional area $\mathcal{A}$ to define $\delta\mathcal{A}$ and δS for use in the assumption of local holography in Eq. (12.3.4). From Eq. (12.3.12), we have

$$\delta\mathcal{A} \equiv \lim_{\epsilon\to\text{small}} \lim_{\mathcal{A}\to\text{small}} \Delta\mathcal{A} = \frac{1}{2}\epsilon^2 \mathcal{A} R_{\mu\nu}(p) k^\mu k^\nu, \tag{12.3.13}$$

where we used the fact that in the limit of a small lightsheet the Ricci tensor could be taken to be a constant evaluated at p. Similarly, using Eq. (12.3.10), we find for the entropy that

$$\delta S \equiv \lim_{\epsilon\to\text{small}} \lim_{\mathcal{A}\to\text{small}} \Delta S_{\mathcal{H}} = \frac{2\pi}{\hbar}\epsilon^2 \mathcal{A} T_{\mu\nu}(p) k^\mu k^\nu \int_0^1 \mathrm{d}\bar{\lambda}\, \bar{g}(\bar{\lambda}). \tag{12.3.14}$$

Local holography posits that $\delta S = \eta\, \delta\mathcal{A}$ for some constant η. For consistency with the Bekenstein-Hawking formula, we expect η to equal $1/4G\hbar$, but for now we will keep it undetermined. Setting Eq. (12.3.13) proportional to Eq. (12.3.14) implies

$$\left[\frac{4\pi}{\hbar\eta}\int_0^1 \mathrm{d}\bar{\lambda}\, \bar{g}(\bar{\lambda})\right] T_{\mu\nu}(p) k^\mu k^\nu = R_{\mu\nu}(p) k^\mu k^\nu. \tag{12.3.15}$$

Let us write $\eta = 1/4G_{\mathrm{S}}\hbar$ and write Newton's constant in Einstein's equation as G_{N}. Then requiring consistency of Eq. (12.3.15) with Einstein's equation and rearranging, we have

$$G_{\mathrm{S}} = \frac{G_{\mathrm{N}}}{2\int_0^1 \mathrm{d}\bar{\lambda}\, \bar{g}(\bar{\lambda})} \geq 2G_{\mathrm{N}}, \tag{12.3.16}$$

noting, as we previously observed, that the integral over $\bar{g}(\bar{\lambda})$ is less than $1/4$. That is, in terms of the constant in Einstein's equation, we have

$$\eta \leq \frac{1}{8G_{\mathrm{N}}\hbar}. \tag{12.3.17}$$

This is inconsistent, by an order-unity factor, with the area-entropy coefficient from black hole thermodynamics, which would be $\eta = 1/4G_{\mathrm{N}}\hbar$. So we see that, while thermodynamic gravity is motivated by the area-entropy equivalence for black holes, enforcing $\delta S = \delta\mathcal{A}/4G\hbar$ would lead to the wrong constant in

Einstein's equation. Moreover, this constant appears in a theory-dependent way via the function $\bar{g}$. On the other hand, one could insist that the correct coefficient be obtained in Einstein's equation. By Eq. (12.3.16), this would require $\delta S = \delta\mathcal{A}\int_0^1 \mathrm{d}\bar{\lambda}\,\bar{g}(\bar{\lambda})/2G\hbar$, which would constitute a theory-dependent modification of the local holography postulate with a coefficient that now no longer corresponds to the area-entropy relation from black hole thermodynamics. In other words, one could require that Einstein's equation and the $1/4G\hbar$ coefficient in the local holography postulate have Newton's constants that differ by the order-unity factor given in Eq. (12.3.16). A question for future work on TG would then be the identification of a justification, independent of Einstein's equation, of why the local holography postulate must take precisely this modified form.

The reason for the inconsistency of Einstein's equation and the expected area-entropy ratio in the formulation of TG we have considered here stems from the fact that, despite the similarity between Eqs. (12.3.1) and (12.3.10), there is a crucial factor-of-g difference. In Ref. [59], the heat transfer was taken to be given by the Rindler form (12.3.1), where g is just λ; interpreted as a modular Hamiltonian, this is the appropriate form for a semi-infinite lightsheet. However, only finite lightsheets [59, 321] can be considered in the formulation of TG, so that θ and $\sigma_{\mu\nu}$ remain subdominant in the Raychaudhuri equation.

There is an important distinction between the formulation of TG here and the causal-diamond derivation of HG in the previous section. The transformation (12.2.8) that brings a Rindler wedge to the causal diamond is a true conformal transformation for the spacetime. In contrast, to bring a semi-infinite lightsheet to a finite segment requires a transformation $\lambda \to 1/\lambda$ that is conformal on two-dimensional subspaces, but not on the spacetime as a whole. For general theories (in particular, those that are not ultralocal), this leads to the

need for the function $g(\lambda, \epsilon)$, which was not present for the causal-diamond formulation.

12.3.3 Loopholes and Alternatives

A possible concern about this analysis might be that the Casini entropy (12.3.10) is calculated in terms of the field-theoretic degrees of freedom alone. That is, one might imagine positing the existence of hidden, quantum-gravitational degrees of freedom that would provide additional entropy so that δS equals $\delta\mathcal{A}/4G\hbar$, with the aim of getting both the correct coefficients in the area-entropy relation and in Einstein's equation.

However, this proves to not be possible. The general form of the Casini entropy must be given by a relation of the form (12.3.10), linear in the energy-momentum tensor, if we are to use $\delta S \propto \delta\mathcal{A}$ to derive Einstein's equation with $T_{\mu\nu}$ on the right-hand side. Positing new degrees of freedom can only affect the calculation of the theory-dependent coefficient $g(\lambda, \epsilon)$. But attaining $\eta = 1/4G\hbar$ would require $|\mathrm{d}\bar{g}/\mathrm{d}\bar{\lambda}|$ to exceed unity. It is shown in Ref. [258] that this is impossible on very general grounds, regardless of any details about quantum field theory: exceeding this limit would violate strong subadditivity of von Neumann entropy or monotonicity of quantum relative entropy. Hence, positing non-field-theoretic degrees of freedom in the density matrix describing the lightsheet system is insufficient to simultaneously recover Einstein's equation and rectify the contradiction with the area-entropy formula we derived in Eq. (12.3.17). We are forced to conclude that the entropy in thermodynamic gravity cannot be the vacuum-subtracted von Neumann (i.e., Casini) entropy of the lightsheet segment $\mathcal{H}$.

An alternative tack for formulating TG would be to use the Casini entropy, but define the quantity δS in a slightly different way. Rather than associating it directly with the quantum state on the null region $\mathcal{H}$, we could let it be the

difference in Casini entropies between the large lightsheet $\mathcal{L}$ emanating from p and the lightsheet with $\mathcal{H}$ removed, $\delta S = \Delta S_{\mathcal{L}} - \Delta S_{\mathcal{L}-\mathcal{H}}$. Note that this is in general a distinctly different quantity from that investigated above, since $\Delta S_{\mathcal{H}} \geq \Delta S_{\mathcal{L}} - \Delta S_{\mathcal{L}-\mathcal{H}}$ by subadditivity. For convenience, we will take $\mathcal{L}$ to be a semi-infinite null surface; because we are only interested in an entropy difference, the conclusions in this section are the same for any $\mathcal{L}$ much longer than $\mathcal{H}$. One might imagine that this alternate formulation, with semi-infinite lightsheets, would allow the Rindler form of the integrand in the expression for the entropy and possibly rescue thermodynamic gravity; however, this will prove to not be the case.

Let us specialize to spacetimes in which gravitational backreaction is small. (Including corrections to the Rindler Hamiltonian induced by spacetime curvature would only be relevant at higher order in Newton's constant.) Generalizing the arguments of Ref. [258] to semi-infinite null surfaces, with affine parameter $\hat{\lambda}$ going from λ_0 to infinity, we have

$$\Delta S(\lambda_0) = \Delta K(\lambda_0) = \frac{2\pi}{\hbar} \int \mathrm{d}\mathcal{A} \int_{\lambda_0}^{\infty} (\hat{\lambda} - \lambda_0) T_{\mu\nu} \hat{k}^{\mu} \hat{k}^{\nu} \mathrm{d}\hat{\lambda}, \tag{12.3.18}$$

where $\hat{k}^{\mu} = (\mathrm{d}/\mathrm{d}\hat{\lambda})^{\mu}$. Then we can define the change in the null Casini entropy, $\delta S = \Delta S_{\mathcal{L}} - \Delta S_{\mathcal{L}-\mathcal{H}} = \Delta S(\alpha) - \Delta S(\beta)$. Here, we have labeled the null regions by the value of the affine parameter from which they emanate, where in the $\hat{\lambda}$ parameterization, $\mathcal{H}$ is defined as $\hat{\lambda} \in [\alpha, \beta]$.

However, this final formulation of the entropy as the null Casini entropy cannot be the correct definition of entropy in thermodynamic gravity. Let us define an affine parameterization that starts at $\lambda = 1$ at $\hat{\lambda} = \alpha$, so $\lambda = \hat{\lambda}/\alpha$. Defining k^{μ} as the tangent four-vector to λ, $(\mathrm{d}/\mathrm{d}\lambda)^{\mu}$, we have

$$\Delta S(\beta) = \frac{2\pi}{\hbar} \int_{\mathcal{A}(\beta\lambda'/\alpha)} \mathrm{d}\mathcal{A} \int_{1}^{\infty} (\lambda' - 1) T_{\mu\nu}(\beta\lambda'/\alpha) k'^{\mu} k'^{\nu} \mathrm{d}\lambda', \tag{12.3.19}$$

where $\lambda' = \alpha\lambda/\beta$ and $k'^{\mu} = (\mathrm{d}/\mathrm{d}\lambda')^{\mu}$. We can make the approximation that $T_{\mu\nu}$ changes slowly with the affine parameter and that α and β are close, so that $T_{\mu\nu}(\beta\lambda'/\alpha) \simeq T_{\mu\nu}(\lambda')$. Further, we can take the cross-sectional area of the lightsheet to be small, so that $T_{\mu\nu}$ is approximately constant over the cross-section at a fixed affine parameter. We thus have

$$\delta S \simeq \frac{2\pi}{\hbar}\int_1^\infty (\lambda-1)[\mathcal{A}(\lambda) - \mathcal{A}(\beta\lambda/\alpha)]\, T_{\mu\nu}(\lambda)k^\mu k^\nu \mathrm{d}\lambda, \tag{12.3.20}$$

where in the final line we dropped the primes, since λ is a dummy variable. Now, from Eq. (12.3.13), we have

$$\mathcal{A}(\lambda) - \mathcal{A}(\beta\lambda/\alpha) \simeq \frac{1}{2}\left(\frac{\beta}{\alpha}-1\right)^2 \mathcal{A}(\lambda) R_{\mu\nu}(\lambda)k^\mu k^\nu, \tag{12.3.21}$$

Plugging this result into Eq. (12.3.20) and then substituting in Einstein's equation (12.3.8), which implies $R_{\mu\nu}k^\mu k^\nu = 8\pi G T_{\mu\nu}k^\mu k^\nu$, we obtain

$$\delta S = \frac{1}{8G\hbar}\left(\frac{\beta}{\alpha}-1\right)^2 \int_1^\infty (\lambda-1)\mathcal{A}(\lambda)[R_{\mu\nu}(\lambda)k^\mu k^\nu]^2 \mathrm{d}\lambda. \tag{12.3.22}$$

We see that Eq. (12.3.22) cannot be arranged in a form that looks like $\delta S = \eta\, \delta\mathcal{A}$ as in Eq. (12.3.4). In particular, Eq. (12.3.22) is second-order rather than linear in the curvature and therefore in the area decrement $\delta\mathcal{A}$. Though Eq. (12.3.18) looks similar to Eq. (12.3.1), one cannot naively conclude that the difference between the values of $\Delta S(\lambda_0)$ for $\lambda_0 = \alpha$ versus β in Eq. (12.3.18) can be taken as simply an integral over $\lambda \in [\alpha, \beta]$; such an operation is not valid when the integrand itself has explicit dependence on its end points, as is the case in Eq. (12.3.18). We have found that by taking δS to be the difference in the null Casini entropies of overlapping null surfaces, we obtain an expression (12.3.22) for δS that is fundamentally incompatible with the local holographic postulate (12.3.4) that is one of the axioms of TG.

Hence, neither the null-limit Casini entropy of a small null region nor the

difference in null-limit Casini entropies of two large null regions provides an acceptable definition of entropy in the thermodynamic formulation of entropic gravity.

12.4 Conclusions

The idea that gravity can be thought of as an entropic force is an attractive one. In this chapter we have distinguished between two different ways of implementing this idea: holographic gravity, which derives the Einstein equation from constraints on the boundary entanglement after varying over different states in the theory, and thermodynamic gravity, which relates the time evolution of a cross-sectional area to the entropy passing through a null surface in a specified spacetime. We argued that holographic gravity is a consistent formulation and indeed that recent work on the modular Hamiltonian in quantum field theory provides additional support for its underlying assumptions. The thermodynamic approach, on the other hand, seems to suffer from a difficulty in providing a self-consistent definition for what the appropriate entropy is going to be.

In the title of this chapter, we asked, "What is the entropy in entropic gravity?" We are now equipped to answer this question. In what we have called "holographic gravity," the vacuum-subtracted von Neumann entanglement entropy (the Casini entropy), evaluated on the null surfaces of the causal diamond, provides an appropriate formulation for an entropic treatment of gravitation. This can help guide further attempts to understand the underlying microscopic degrees of freedom giving rise to gravitation in general spacetime backgrounds.

Chapter 13
Attractor Solutions in Scalar-Field Cosmology

Models of cosmological scalar fields often feature "attractor solutions" to which the system evolves for a wide range of initial conditions. There is some tension between this well-known fact and another well-known fact: Liouville's theorem forbids true attractor behavior in a Hamiltonian system. In universes with vanishing spatial curvature, the field variables ϕ and $\dot{\phi}$ specify the system completely, defining an effective phase space. We investigate whether one can define a unique conserved measure on this effective phase space, showing that it exists for $m^2\phi^2$ potentials and deriving conditions for its existence in more general theories. We show that apparent attractors are places where this conserved measure diverges in the ϕ-$\dot{\phi}$ variables and suggest a physical understanding of attractor behavior that is compatible with Liouville's theorem.

This chapter is from Ref. [12], G. N. Remmen and S. M. Carroll, "Attractor Solutions in Scalar-Field Cosmology," Phys.Rev. ***D88*** *(2013) 083518,* arXiv:1309.2611 [gr-qc].

13.1 Introduction

Two of the favorite moves in the repertoire of the modern theoretical cosmologist are (1) positing one or more scalar fields whose energy density exerts an important influence on the evolution of the universe and (2) claiming (or at least aspiring to be able to claim) that certain conditions or behaviors qualify as "natural." These tendencies meet in the notion of cosmological attractors: dynamical conditions under which evolving scalar fields approach a certain

kind of behavior without finely-tuned initial conditions [322–333], whether in inflationary cosmology or late-time quintessence models. In dynamical systems theory, attractor behavior describes situations where a collection of phase-space points evolve into a certain region and never leave. This is incompatible with Liouville's theorem, which states that the volume of a region of phase space is invariant under time evolution. Hamiltonian systems, of which scalar-field cosmologies (Einstein's equation plus a dynamical scalar field, restricted to homogeneous configurations) are examples, obey Liouville's theorem, and therefore cannot support true attractor behavior.

So what is going on? In this chapter, we reconcile the appearance of attractor solutions in scalar-field cosmologies with their apparent mathematical impossibility by making two points. First, we point out the fact (well-known, although rarely stated explicitly) that the combined gravity/scalar-field equations exhibit an apparently accidental simplification in the case of flat universes. This simplification allows us to express the complete evolution in terms of an effective two-dimensional "phase space" with coordinates ϕ and $\dot{\phi}$, even though the true phase space is four-dimensional (since the scale factor and its conjugate momentum are independent variables). Of course, ϕ and $\dot{\phi}$ aren't canonical coordinates on phase space, so the measure $\mathrm{d}\dot{\phi} \wedge \mathrm{d}\phi$ isn't very physically meaningful.

Our second point is that it is seemingly possible to define a conserved measure on the ϕ-$\dot{\phi}$ effective phase space, although this measure looks very different from $\mathrm{d}\dot{\phi} \wedge \mathrm{d}\phi$. We cannot rigorously prove its existence in general, but we can show that it corresponds to a Lagrangian on effective phase space if it does exist; in the simple example of a canonical scalar field with a quadratic potential, we show that a unique measure on effective phase space exists and derive some of its properties. By construction, there can be no "attractor"

solutions with respect to this measure. Nevertheless, we suggest there is a relevant sense in which attractor solutions are physically meaningful, if certain functions of the phase-space variables are directly observable. Finally, we comment on the connection between this classical analysis and the boundary induced on phase space by the Planck scale.

13.2 Phase Space, Measures, and Attractors

We start by reviewing scalar-field cosmology in phase space, following Gibbons, Hawking, and Stewart (GHS) [64]. There are subtleties due to the fact that GR is a constrained system. In this section, we also discuss the intuitive idea of an attractor and contrast it with Hamiltonian behavior.

Because a phase space Γ of dimension $2n$ is a symplectic manifold, there is a closed two-form defined on Γ,

$$\omega = \sum_{i=1}^{n} \mathrm{d}p_i \wedge \mathrm{d}q^i. \tag{13.2.1}$$

This symplectic form defines the Liouville measure,

$$\Omega = \frac{(-1)^{n(n-1)/2}}{n!}\omega^n. \tag{13.2.2}$$

Liouville's theorem from classical mechanics states that this measure is conserved along the Hamiltonian flow vector $X_{\mathcal{H}}$. That is, given trajectories that initially cover some region $S \subset \Gamma$ and that evolve under $X_{\mathcal{H}}$ to cover region S', we have

$$\int_S \Omega = \int_{S'} \Omega. \tag{13.2.3}$$

Equivalently, the Lie derivative of Ω vanishes along $X_{\mathcal{H}}$,

$$\pounds_{X_{\mathcal{H}}}\Omega = 0. \tag{13.2.4}$$

Because the metric component g_{00} is not a propagating degree of freedom in the Einstein-Hilbert action, general relativity is a constrained system, in

which the Hamiltonian $\mathcal{H}$ is set to a boundary-condition-dependent constant along physical trajectories. That is, trajectories are confined to a hypersurface in Γ of dimension $2n-1$ for which $\mathcal{H}=\mathcal{H}_\star$; we will call this the Hamiltonian constraint surface,

$$C=\Gamma/\left\{\mathcal{H}=\mathcal{H}_\star\right\}. \tag{13.2.5}$$

The Hamiltonian flow vector, describing Hamiltonian evolution of trajectories in C, is

$$X_{\mathcal{H}}=\frac{\partial\mathcal{H}}{\partial p_i}\frac{\partial}{\partial q^i}-\frac{\partial\mathcal{H}}{\partial q^i}\frac{\partial}{\partial p_i}, \tag{13.2.6}$$

where (q^i,p_i) are the canonical coordinates and their conjugate momenta. The space of trajectories (as opposed to states) can be defined by taking the quotient

$$M=C/X_{\mathcal{H}}. \tag{13.2.7}$$

Previously, Gibbons, Hawking, and Stewart [64] contructed the unique measure on M for FRW universes. The GHS measure is unique in that it is positive, independent of parameterization, and respects the symmetries of the problem without introducing additional structures. It is obtained from the symplectic form ω by identifying the nth coordinate of phase space Γ as time t, so that

$$\omega=\tilde{\omega}+\mathrm{d}\mathcal{H}\wedge\mathrm{d}t=\sum_{i=1}^{n-1}\mathrm{d}p_i\wedge\mathrm{d}q^i+\mathrm{d}\mathcal{H}\wedge\mathrm{d}t. \tag{13.2.8}$$

The corresponding measure, a $(2n-2)$-form, is

$$\Theta=\frac{(-1)^{(n-1)(n-2)/2}}{(n-1)!}\tilde{\omega}^{n-1}. \tag{13.2.9}$$

The metric describing an FRW universe is

$$\mathrm{d}s^2=-N^2\mathrm{d}t^2+a^2\left(t\right)\left(\frac{\mathrm{d}r^2}{1-\kappa r^2}+r^2\mathrm{d}\Omega^2\right), \tag{13.2.10}$$

where $a\left(t\right)$ is the scale factor, normalized to unity at some time t_0, and N is the lapse function. The curvature parameter $\kappa\in\mathbb{R}$ has dimensions of $[\text{length}]^{-2}$.

We may also define $k = \kappa R_0^2 \in \{0, \pm 1\}$, where R_0 is the radius of curvature of the universe when $a(t_0) = 1$.

Studying the dynamics of the FRW scale factor coupled to some matter source is known as the minisuperspace approximation. The minisuperspace Lagrangian for gravity plus a scalar field with potential $V(\phi)$ is

$$\mathcal{L} = 3M_{\rm Pl}^2 \left(Na\kappa - \frac{a\dot{a}^2}{N} \right) + a^3 \left[\frac{\dot{\phi}^2}{2N} - NV(\phi) \right]. \tag{13.2.11}$$

The canonical momenta, defined as $p_i = \partial \mathcal{L} / \partial \dot{q}^i$, are

$$p_N = 0, \; p_a = -6N^{-1} M_{\rm Pl}^2 a\dot{a}, \text{ and } \; p_\phi = N^{-1} a^3 \dot{\phi}. \tag{13.2.12}$$

Note that N is a Lagrange multiplier: it is non-dynamical and will not be a part of the phase space. Performing a Legendre transformation, the Hamiltonian, in units where $M_{\rm Pl} = \sqrt{\hbar c / 8\pi G} = 1$, is

$$\begin{aligned} \mathcal{H} &= N \left[-\frac{p_a^2}{12a} + \frac{p_\phi^2}{2a^3} + a^3 V(\phi) - 3a\kappa \right] \\ &= -3a^3 N \left\{ \left(\frac{\dot{a}}{a} \right)^2 + \frac{\kappa}{a^2} - \frac{1}{3} \left[\frac{1}{2}\dot{\phi}^2 + V(\phi) \right] \right\}. \end{aligned} \tag{13.2.13}$$

The equation of motion for N sets it equal to an arbitrary constant, which we choose to be unity henceforth. Varying the action with respect to N gives the Hamiltonian constraint for FRW universes, $\mathcal{H}_\star = 0$, which is equivalent to the Friedmann equation,

$$H^2 = \frac{1}{3} \left[\frac{1}{2}\dot{\phi}^2 + V(\phi) \right] - \frac{\kappa}{a^2}, \tag{13.2.14}$$

where the Hubble parameter is $H \equiv \dot{a}/a$. Thus, Γ is four-dimensional, with (ϕ, p_ϕ, a, p_a) being a possible parameterization. The Hamiltonian constraint surface C, once a value of κ is chosen, is three-dimensional. The space of trajectories M is two-dimensional. The GHS measure can be written as the Liouville measure with the Hamiltonian constraint:

$$\Theta = \left(\mathrm{d}p_a \wedge \mathrm{d}a + \mathrm{d}p_\phi \wedge \mathrm{d}\phi \right)|_{\mathcal{H}=0}. \tag{13.2.15}$$

A true attractor in phase space can be thought of as a region toward which phase-space trajectories converge when plotted in canonical coordinates. More formally, an attractor is defined [334] as a region $A \subset \Gamma$ with the following properties:

1. A is compact;

2. Given a trajectory $\{\mathcal{P}(t, x_0)\} \subset \Gamma$ beginning at $\mathcal{P}(t_0, x_0) = x_0 \in A$, $\mathcal{P}(t, x_0) \in A$ for all $t > t_0$;

3. There exists a basin of attraction, a neighborhood B of A such that for all $x_B \in B$ and for any neighborhood N of A, there exists t_N such that $\mathcal{P}(t, x_B) \in N$ for all $t > t_N$;

4. Properties 2. and 3. are not satisfied by any $A' \subsetneq A$.

There are other, related, definitions of attractors in the mathematical literature [335]; in particular, a definition in terms of Lyapunov stability is possible (cf. Sec. 13.6, below).

An immediate consequence of Liouville's theorem is that no true attractor can exist in the phase space of a system described by a Hamiltonian [336]; see also Ref. [118], Sec. 22.6. Intuitively, if a bundle of trajectories converges along a particular axis in phase space in a given coordinate system, it must compensatingly spread out along other axes, to conserve the total phase-space measure. Though we may wish to describe such behavior as an "attractor," it is always possible to remove this apparent convergence by a canonical change of coordinates: in essence, there is no coordinate-independent notion of an attractor in the full four-dimensional phase space describing scalar-field cosmology in an FRW universe [337].

Despite the fact that it does not rigorously exist, however, the intuitive idea of an attractor appears in the literature on scalar-field cosmology, though

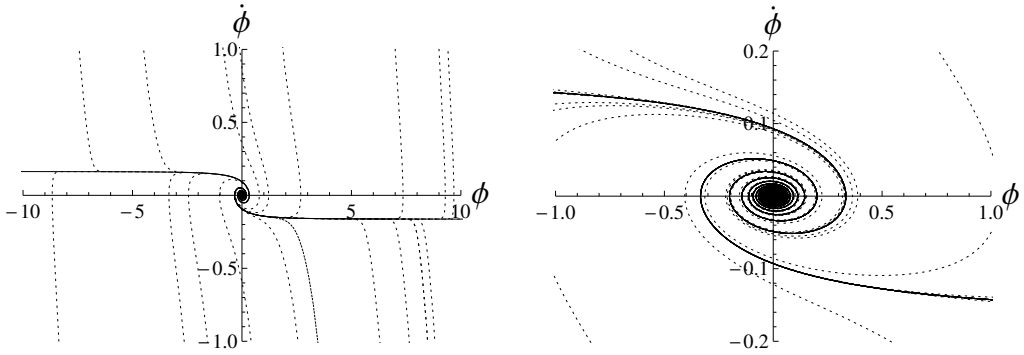

Figure 13.1. Apparent attractor solutions for an $m^2\phi^2$ potential, with equation of motion $\ddot{\phi} + \sqrt{3/2}\sqrt{m^2\phi^2 + \dot{\phi}^2}\dot{\phi} + m^2\phi = 0$. Solid: the apparent attractors; dotted: numerical solutions for random initial conditions. Plots are in ϕ-$\dot{\phi}$ space, in units where $M_{\rm Pl} = 1$; the scalar mass is chosen to be $m = 0.2M_{\rm Pl}$. At large field values, the solutions are approximated by the lines $\dot{\phi} = \pm\sqrt{2/3}m$, while for small field values, all solutions converge on the origin.

a definition of what is meant by an "attractor" is often left implicit. This often occurs as a result of plotting trajectories in some non-canonical phase-space variables, most commonly ϕ versus $\dot{\phi}$ [322, 323, 331]. However, as one can see in Figs. 13.1 and 13.2, apparent attractor behavior in $(\phi, \dot{\phi})$ coordinates need not correspond to attractor behavior when plotted in (ϕ, p_ϕ). Furthermore, recall that the full phase space Γ is four-dimensional, not two-dimensional: a and p_a are suppressed in Figs. 13.1 and 13.2, and initially nearby trajectories would generally spread in these variables. In other papers, the notion of an "attractor" is used in a manifestly coordinate-dependent manner, with respect to some physical observables that either become smaller with time [332] or for which differences between initially different trajectories vanish rapidly in some particular coordinates [327–329].

It is easy to see why such behavior is described as attractor-like: one simply looks at the plots, perhaps implicitly assuming a "graph paper measure" $\mathrm{d}\dot{\phi}\wedge\mathrm{d}\phi$. Though this assumption seems natural, it is a coordinate-dependent artifact, as ϕ and $\dot{\phi}$ are not canonically conjugate. It is our aim in this work to make all of these notions more rigorous, examining both the issue of the measure

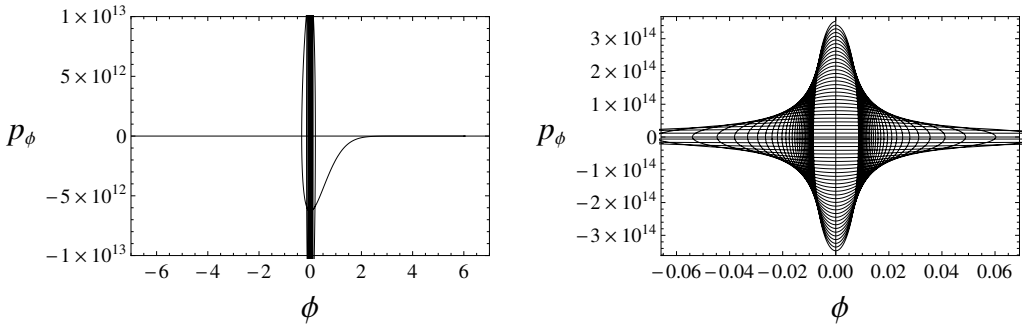

Figure 13.2. Numerical solution for evolution of an FRW universe with an $m^2\phi^2$ potential, with initial conditions $(\phi, \dot{\phi}) = (6, 0.25)$ at $a\,(t = 0) = 1$, plotted in (ϕ, p_ϕ) coordinates, where $p_\phi = a^3\dot{\phi}$ is the canonical momentum conjugate to ϕ. Units are chosen such that $M_{\rm Pl} = 1$, with the scalar mass $m = 0.2M_{\rm Pl}$. The apparent attractor behavior seen in Fig. 13.1 disappears in these coordinates.

on the space of field variables and the definition of apparent attractor-like behavior. Our results should help to create a common, more mathematically valid, and less *ad hoc* language for comparing results between different models of scalar-field cosmology.

13.3 Effective Phase Space for a Single Scalar Field

In this section, we identify a sense in which ϕ and $\dot{\phi}$, though not canonically conjugate, are special coordinates for universes with zero spatial curvature. That is, we will show that the full four-dimensional phase space Γ is larger than necessary to fully capture the dynamics of scalar-field cosmology for flat universes; ϕ-$\dot{\phi}$ space can be regarded as an effective phase space, in a sense that will be made precise. We proceed first by defining the notion of a vector field invariant map, making as general and coordinate-independent definitions as possible. In essence, given a map between two manifolds and a vector field on the first manifold, the map is vector field invariant if it provides a way of uniquely specifying a vector field on the second manifold. We will find that the map from Γ to ϕ-$\dot{\phi}$ space for flat universes is vector field invariant with respect to the Hamiltonian flow vector.

13.3.1 Vector Field Invariant Maps

Before investigating whether there is a sense in which the non-canonical coordinates $(\phi, \dot{\phi})$ constitute a parameterization with any special mathematical properties, we first require some definitions and notation. Given two manifolds M and N, a mapping $\psi : M \to N$, and $f \in \mathcal{F}(N)$, where $\mathcal{F}(N)$ is the space of smooth real-valued functions with domain N, the pullback $\psi^\star : \mathcal{F}(N) \to \mathcal{F}(M)$ of f by ψ is defined by:

$$\psi^* f = (f \circ \psi) : M \to \mathbb{R}. \tag{13.3.1}$$

We can think of f as specifying a coordinate on N and the pullback as specifying a coordinate on M. Now, at a given point $p \in M$, we may regard a vector $X(p)$ as a function $X_p : \mathcal{F}(M) \to \mathbb{R}$. If we think of $g \in \mathcal{F}(M)$ as specifying a coordinate (also called g) on M, then $X_p(g)$ gives the value of the g-component of the vector at p. A vector field X on M is the assignment of a vector X_p to each point $p \in M$ in a continuous and smooth fashion. Given the map $\psi : M \to N$ and a function $f : N \to \mathbb{R}$, the pushforward of X at $\psi(p) \in N$ is

$$(\psi_* X)_{\psi(p)}(f) = X_p(\psi^* f). \tag{13.3.2}$$

We note that the pushforward ψ_* is a map from the tangent space at p, T_pM, into $T_{\psi(p)}N$ and that $\psi_* X : \mathcal{F}(N) \to \mathbb{R}$. In this sense, we can write $\psi_*(X) = X \circ \psi^*$. For further reference, see Appendix C of Ref. [338] and Appendix A of Ref. [271].

We may now define "vector field invariance," a way of formalizing the idea that a many-to-one map creates a unique vector field. Suppose we have a map $\psi : M \to N$ and vector field X on M. For each point $q \in N$, write the preimage in M as $\psi^{-1}(q) = \{p \in M \mid \psi(p) = q\}$. Then say that the map ψ *is vector field invariant with respect to* X if for any function $f \in \mathcal{F}(N)$ and for all $q \in N$, $X_p(\psi^* f) = X_{p'}(\psi^* f)$ for all p, $p' \in \psi^{-1}(q)$. If a map ψ is vector

field invariant with respect to X, we may write $X_p(\psi^* f) = \tilde{X}_q(f)$ without ambiguity for $p \in \psi^{-1}(q)$. Then we have a unique vector field $\tilde{X}$ on N. We can say that $\tilde{X}$ is the vector field *induced by X on N under the (vector field invariant) map ψ*.

The images of integral curves that are distinct under a vector field invariant map do not intersect. If we have a vector field invariant map $\psi : M \to N$, where M has vector field X, then the images of integral curves of X are integral curves of $\tilde{X}$ in N. Therefore, by uniqueness, given two integral curves in M not mapped onto each other in N by ψ, their images in N cannot intersect. If M is the phase space for some Hamiltonian system and $\psi : M \to N$ is vector field invariant with respect to the Hamiltonian flow vector, one can therefore think of N as an *effective phase space*.

13.3.2 A Map for FRW Universes

We will now show that, for scalar-field cosmology in a flat universe, the choice of ϕ and $\dot{\phi}$ as coordinates allows one to eliminate a and $\dot{a}$ and thus reduce the dynamical phase space to two dimensions. Consider a map χ from the Hamiltonian constraint three-manifold C to a two-manifold K, where $\chi^{-1}(q)$ is the set of all points in C with equal values of ϕ and $\dot{\phi}$. That is, K is isomorphic to ϕ-$\dot{\phi}$ space. We will show that in a flat universe ($\kappa = 0$) with a scalar field described by a potential $V(\phi)$ and a canonical kinetic term, the map χ is vector field invariant with respect to the Hamiltonian flow vector $X_{\mathcal{H}}$.

It is sufficient to exhibit one such map χ, as all other maps such that the preimage of $q \in K$ is the set of all points in C with equal values of ϕ and $\dot{\phi}$ can be obtained from χ via a bijection. Without loss of generality, we may therefore specify coordinates $(\phi, \dot{\phi}, a, H)$ on the full phase space Γ, which are inherited by C, so that C is parameterized by four coordinates related by the Hamiltonian constraint. Note, however, that none of our conclusions are

dependent on choosing a and H as the other two coordinates. That is, the notion of vector field invariance of $\chi : C \to K$ is a statement only about ϕ and $\dot{\phi}$, independent of the other coordinates. With the Hamiltonian (13.2.13) and flow vector (13.2.6), we have

$$\begin{aligned} X_{\mathcal{H}}^{(\phi)} &= \frac{p_\phi}{a^3}, \\ X_{\mathcal{H}}^{(p_\phi)} &= -a^3 V'(\phi), \\ X_{\mathcal{H}}^{(a)} &= -\frac{p_a}{6a}, \\ X_{\mathcal{H}}^{(p_a)} &= -\frac{p_a^2}{12a^2} + \frac{3p_\phi^2}{2a^4} - 3a^2 V(\phi) + 3\kappa. \end{aligned} \tag{13.3.3}$$

Using the expressions for p_a and p_ϕ in Eq. (13.2.12) to rescale and eliminate a and $\dot{a}$ in favor of ϕ and $\dot{\phi}$, we have

$$\begin{aligned} X_{\mathcal{H}}^{(\phi)} &= \dot{\phi}, \\ X_{\mathcal{H}}^{(\dot{\phi})} &= \frac{1}{a^3} X_{\mathcal{H}}^{(p_\phi)} = -V'(\phi), \\ X_{\mathcal{H}}^{(H)} &= -\frac{1}{6a^2} X_{\mathcal{H}}^{(p_a)} = \frac{1}{2}H^2 - \frac{1}{4}\dot{\phi}^2 + \frac{1}{2}V(\phi) - \frac{\kappa}{2a^2}. \end{aligned} \tag{13.3.4}$$

Therefore, for $\kappa = 0$, the ϕ-, $\dot{\phi}$-, and H-components of the vector field are independent of a. Further, from the Friedmann equation (13.2.14), H can be written as a function of ϕ and $\dot{\phi}$ for $\kappa = 0$. Thus, for a flat universe, the ϕ-, $\dot{\phi}$-, and H-components of the Hamiltonian vector field $X_{\mathcal{H}}$ can be written in terms of ϕ and $\dot{\phi}$ alone. This is the slightly more careful version of our previous statement that ϕ and $\dot{\phi}$ allow a and $\dot{a}$ to be eliminated from the dynamics. Now, consider the map $\chi : C \to K$ defined by $\chi(a, \phi, \dot{\phi}, H) = (\phi, \dot{\phi})$. Under such a map, the condition for vector field invariance for a given vector field X is simply the condition that the ϕ-, $\dot{\phi}$-, and H-components of X can be written in terms of only ϕ and $\dot{\phi}$. Hence, we conclude that the map χ is vector field invariant with resepect to the Hamiltonian vector field $X_{\mathcal{H}}$ for a flat universe.

We have shown, for a universe of zero spatial curvature, that there is a sense in which $(\phi, \dot{\phi})$ become effective phase-space coordinates. This formalizes

the intuitive idea that the equations of motion can be written purely in terms of these variables. There exists a vector field invariant map with respect to the Hamiltonian flow vector, from the full three-dimensional constraint surface C to a two-dimensional manifold K: we find that K is isomorphic to ϕ-$\dot{\phi}$ space. This is a nontrivial property—it is not in general true, given a three-dimensional surface with a Hamiltonian vector field, that a vector field invariant mapping to a two-dimensional manifold exists. The criterion $\kappa = 0$ is necessary for $(\phi, \dot{\phi})$ to be an effective phase space; indeed, trajectories can cross in ϕ-$\dot{\phi}$ space if $\kappa \neq 0$. Furthermore, the projection of Γ onto two *canonical* coordinates *does not* constitute construction of an effective phase space; this fact can be illustrated dramatically by considering orbits in (ϕ, p_ϕ) for an $m^2\phi^2$ potential (see Fig. 13.2). In this sense, ϕ and $\dot{\phi}$ are special coordinates with which to parameterize the phase space of scalar-field cosmology.

13.3.3 The Geometrical Picture

One can develop more intuition about the notion of vector field invariance by considering the geometry of the Hamiltonian constraint submanifold embedded in the full phase space, for a specific model with $V(\phi) = m^2\phi^2/2$. The four-dimensional phase space Γ is foliated into three-dimensional Hamiltonian submanifolds C, each with a unique value of κ, with the Friedmann equation (13.2.14) giving the constraint.

Consider a Hamiltonian submanifold C for some choice of curvature κ. Restricted to a particular value of the scale factor a, the Hamiltonian submanifold becomes a two-dimensional surface C_a immersed within a three-dimensional space Γ_a. We can think of C as being the disjoint union of the C_a. Formally speaking, C is formed by the fibration of the family of manifolds C_a over the positive real line $\mathbb{R}_+$ parameterized by a: in general, this produces a non-factorizable three-manifold within Γ, since the C_a are often different in size

and shape for different values of a. As one can see in Fig. 13.3, all the C_a are the same if $\kappa = 0$. This is due to the fact that for the choice $\kappa = 0$, the Hamiltonian constraint (13.2.14) is independent of a in $(a, H, \phi, \dot{\phi})$ coordinates:

$$H^2 = \frac{1}{3}\left[\frac{1}{2}\dot{\phi}^2 + V(\phi)\right]. \tag{13.3.5}$$

More precisely, parameterizing $C_a = \Gamma_a \cap C$ with the other three coordinates on Γ (excluding a), we find that, for $\kappa = 0$, C_a contains the same set of points, a cone in Γ_a, regardless of the choice of a. Hence, one can pick a two-manifold $C_{a_\star}$ for any choice of scale factor $a_\star$ and find that C is a product space: $C = C_{a_\star} \times \mathbb{R}_+$. This is the sense in which a ceases to be a dynamical variable for flat universes.

As previously, let $K \cong \mathbb{R}^2$ denote ϕ-$\dot{\phi}$ space and consider the vector field invariant map $\chi : C \to K$ defined by $\chi(a, H, \phi, \dot{\phi}) = (\phi, \dot{\phi})$, where C is the Hamiltonian submanifold for a flat universe. More generally, we could let K be any space isomorphic to ϕ-$\dot{\phi}$ space and let χ be any function for which the preimage of a point in K is the set of all points in C with a particular value of ϕ and $\dot{\phi}$. It was previously shown that χ is vector field invariant with respect to the Hamiltonian flow vector $X_{\mathcal{H}}$. From Fig. 13.3 we can see why this is true. The Hamiltonian flow vector describes a vector field on C; on each manifold C_a, the H-, ϕ-, and $\dot{\phi}$-components of $X_{\mathcal{H}}$ give a vector field, which we can imagine describing flow tangent to each of the slices shown in Fig. 13.3. The projection of the vector field from a slice C_a down into the horizontal plane in Fig. 13.3 corresponds to the pushforward of the vector field from C_a to K. If this vector field in K is the same no matter which slice C_a we chose, then χ is vector field invariant with respect to $X_{\mathcal{H}}$. This is manifestly true for the flat universe case, because C factors as shown above. It is also clear from Fig. 13.3 that this is not true for $\kappa \neq 0$: the manifold C_a changes dramatically as a is varied, so the vector field that we push forward to K will be different for different a.

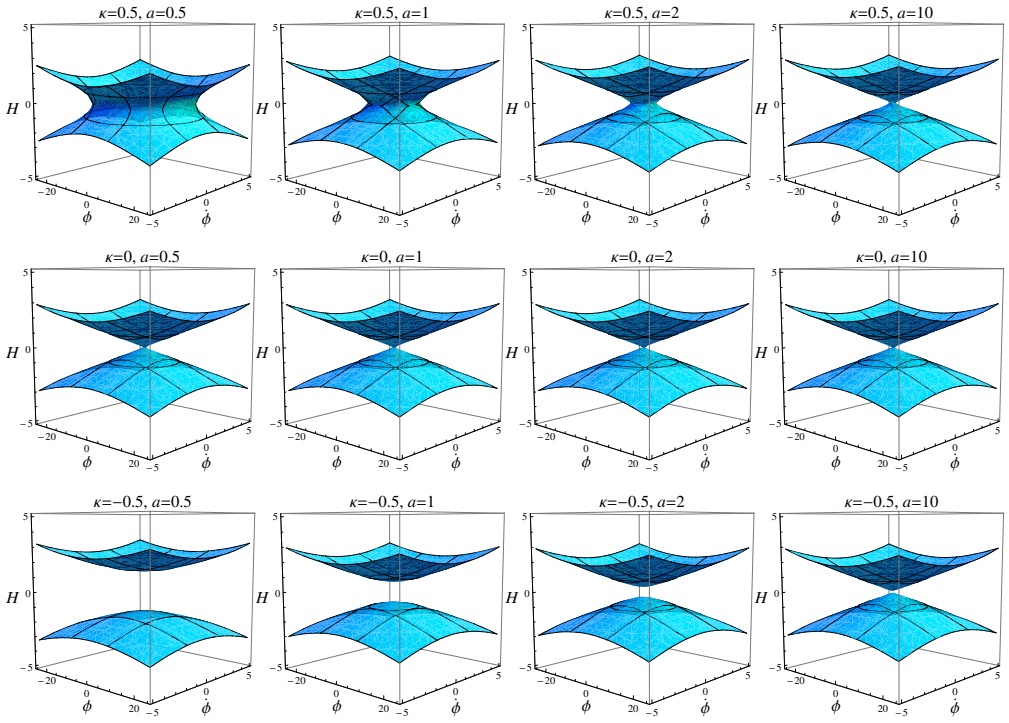

Figure 13.3. Plots of the Hamiltonian three-surface C for $V(\phi) = m^2\phi^2/2$, in $(a, H, \phi, \dot{\phi})$ coordinates, at slices of various values of the scale factor a. Top row: $\kappa = 0.5$, middle row: $\kappa = 0$, bottom row: $\kappa = -0.5$. Units used are $M_{\rm Pl} = 1$ and $m = 0.2$.

At this point we may ask again: is this property of ϕ and $\dot{\phi}$ really distinctive? That is, does a different choice of coordinates on Γ, say (a, p_a, ϕ, p_ϕ) give the same result: a map of the form ψ from (a, p_a, ϕ, p_ϕ) to (ϕ, p_ϕ) that, provided $\kappa = 0$, is vector field invariant with respect to $X_{\mathcal{H}}$? We saw in Fig. 13.2 that this is not the case, but it is useful to consider why vector field invariance fails for ϕ-p_ϕ space from the geometrical point of view. As we see in Fig. 13.4, even in the $\kappa = 0$ case, the partition of the manifold C into C_a yields an inequivalent set of points in Γ_a for different values of a when parameterized by (p_a, ϕ, p_ϕ), so that C is merely a fibration of the C_a over $\mathbb{R}_+$. Another way of saying this is that in (a, p_a, ϕ, p_ϕ) coordinates, C is non-factorizable even in the $\kappa = 0$ case. Hence, drawing components of the Hamiltonian flow vector field on the partition of C, we see that a projection into the ϕ-p_ϕ plane will give a vector

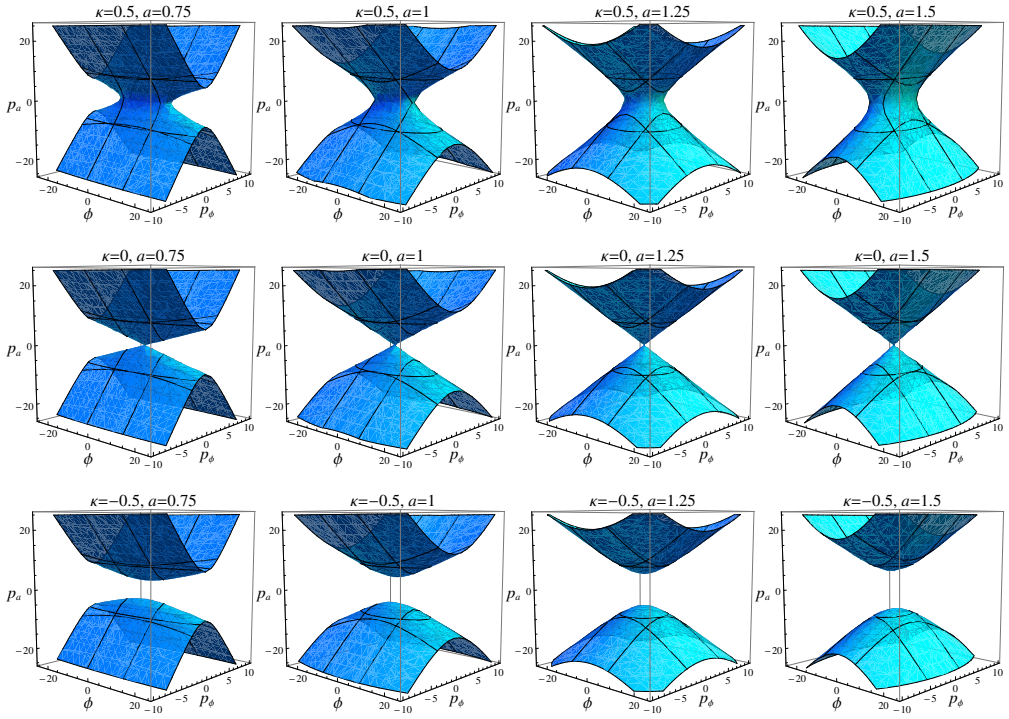

Figure 13.4. Plots of the Hamiltonian constraint manifold C for $V(\phi) = m^2\phi^2/2$, in (a, p_a, ϕ, p_ϕ) coordinates, in two-dimensional slices C_a at various values of the scale factor a. Top row: $\kappa = 0.5$, middle row: $\kappa = 0$, bottom row: $\kappa = -0.5$. Units used are $M_{\rm Pl} = 1$ and $m = 0.2$.

field that is different at different values of a: that is, $\psi : (a, p_a, \phi, p_\phi) \to (\phi, p_\phi)$ is *not* vector field invariant with respect to $X_{\mathcal{H}}$. In this way, we have shown that the property of vector field invariance that χ possesses is nontrivial and not a generic property of any map from C onto a two-dimensional manifold: ϕ and $\dot{\phi}$ are coordinates with a special property.

13.4 Constructing a Measure on Effective Phase Space

We now have in hand an effective phase space K for flat scalar-field cosmology, namely, ϕ-$\dot{\phi}$ space. Its properties are defined generally through the formalism of a vector field invariant map and it contains all of the dynamical variables describing the evolution of a flat FRW universe dominated by a scalar field. However, while K captures the entire dynamics of the system (every point

is part of a unique trajectory), it is not naturally a symplectic manifold, the coordinates $(\phi, \dot{\phi})$ are not canonically conjugate, and there is no reason to expect the naive measure $\mathrm{d}\dot{\phi} \wedge \mathrm{d}\phi$ to be conserved. We now ask whether these features can be corrected, by finding a measure on this effective phase space that actually is conserved.

While the Liouville measure (13.2.15) is appropriate for the full phase space Γ, we are interested now in finding a measure on the effective phase space. Taking a constructive approach, we first examine the constraint imposed by conservation of the measure under Hamiltonian evolution of trajectories, calling such a measure a "conserved measure." We then examine the question of whether the effective phase space itself has a Lagrangian description, that is, whether the equation of motion in terms of ϕ and $\dot{\phi}$ alone can be derived from a Lagrangian $\mathcal{L}_K$ defined on K. If such a Lagrangian exists, it allows us to define a conjugate momentum $\pi_\phi \equiv \partial \mathcal{L}_K / \partial \dot{\phi}$. The measure $\mathrm{d}\pi_\phi \wedge \mathrm{d}\phi$ on K is then automatically conserved under the Hamiltonian flow. We show that the converse is also true; if there is a conserved measure, there is a corresponding Lagrangian description. Finally, for the special case of an $m^2\phi^2$ potential, we examine the behavior of the measure at early and late times and prove that the measure on K exists.

13.4.1 Conservation under Hamiltonian Flow

As shown in Ref. [339], the GHS measure (13.2.15) diverges for flat universes ($\kappa = 0$); see also Ref. [340]. Specifically, as Ω_k, the fraction of the critical energy density parameterized by curvature, approaches zero, $\Theta \propto |\Omega_k|^{-5/2}$. In this sense, as Carroll and Tam [339] note, the flatness problem in cosmology is illusory, a consequence of implicitly assuming a flat measure on the space of FRW solutions; all universes but a set of measure zero are spatially flat, according to the GHS measure. This divergence was briefly noted by Gibbons, Hawking,

and Stewart [64]. The GHS measure is only well defined for Hamiltonian systems with an odd number of constraints (i.e., the Hamiltonian submanifold corresponds to a single constraint). However, it is important to note that the specification of a flat universe does not increase the number of constraints, since this just amounts to selecting a particular value of κ in Eq. (13.2.14). See discussion in Sec. 13.3.3 for details of the phase-space topology.

Given the observed (near-)flatness of our own universe [65], it is well motivated to consider the question of the measure on the subspace of Γ corresponding to flat universes. Because of its divergent behavior for $\kappa = 0$, the GHS measure cannot help us in this case. Earlier attempts to regularize the measure, for example by considering an ϵ-neighborhood around the zero-curvature Hamiltonian constraint surface [339] or by identifying universes with similar curvatures [340] have not proven satisfactory;[13.1] see also Refs. [341, 342]. A different approach seems to be required. As we have seen, the scale factor a becomes non-dynamical for $\kappa = 0$ and the scalar coordinates ϕ and $\dot{\phi}$ constitute an effective phase space, by virtue of the vector field invariant mapping discussed in Sec. 13.3. Though the GHS and Liouville measures give us no information in this subspace, we can use the principles and reasoning of the full phase-space argument to motivate the treatment of the measure question on effective phase space. As noted in Sec. 13.2, it is conventional to implicitly assume a flat measure $\mathrm{d}\dot{\phi} \wedge \mathrm{d}\phi$ in effective phase space when making statements about attractors, number of e-foldings, etc. Considering the measure question in effective phase space allows us to assess the validity of this assumption.

For simplicity of notation, let the vector field $\tilde{X}_{\mathcal{H}}$ induced from the Hamiltonian evolution vector $X_{\mathcal{H}}$ under the mapping $\chi : C \to K \cong (\phi, \dot{\phi})$ be written in K as $\mathbf{v}$. Define $x = \phi$ and $y = \dot{\phi}$. The measure on a two-dimensional space

[13.1] We thank Alan Guth for conversations on this point.

is a two-form $\boldsymbol{\sigma}$, which we can always write as

$$\boldsymbol{\sigma} = f(x,y)\mathrm{d}x \wedge \mathrm{d}y \tag{13.4.1}$$

for some function $f(x,y)$. We seek a measure that is conserved with evolution along $\mathbf{v}$:

$$\pounds_{\mathbf{v}}\boldsymbol{\sigma} = \pounds_{\mathbf{v}}\left[f\left(x,y\right)\mathrm{d}x \wedge \mathrm{d}y\right] = 0. \tag{13.4.2}$$

We can compactly express the condition (13.4.2) as the vector equation

$$\nabla \cdot (f\mathbf{v}) = 0. \tag{13.4.3}$$

Note that this is equivalent to one of the Euler equations of fluid dynamics for a steady flow, $\partial\rho/\partial t = 0$, where ρ is the density of the fluid and $\mathbf{v}$ its velocity field:

$$\frac{\partial \rho}{\partial t} + \nabla \cdot (\rho\mathbf{v}) = 0. \tag{13.4.4}$$

This is simply the statement of conservation of mass. Hence, our conserved two-form can be thought of as the density of fluid in a steady-flow system. The probability of a given bundle of trajectories is conserved under Hamiltonian evolution, just as the mass of a parcel of fluid is conserved as it flows.

For a single scalar with a canonical kinetic term, the vector field $\mathbf{v}$ can be found from $X_{\mathcal{H}}$ as follows. Setting $X_{\mathcal{H}}^{(p_\phi)}$ given in Eq. (13.3.3) equal to $\partial_t p_\phi$ [recalling from Eq. (13.2.12) that $p_\phi = a^3\dot{\phi}$], we have the Klein-Gordon equation

$$\ddot{\phi} + 3H\dot{\phi} + V'\left(\phi\right) = 0. \tag{13.4.5}$$

With H as given by the Friedmann equation (the Hamiltonian constraint) (13.3.5), we have

$$\ddot{\phi} = -\sqrt{3}\dot{\phi}\sqrt{\frac{1}{2}\dot{\phi}^2 + V\left(\phi\right)} - V'\left(\phi\right) = v^{\dot{\phi}}. \tag{13.4.6}$$

The vector field in ϕ-$\dot{\phi}$ space is therefore

$$\mathbf{v} = \left(y,\ -\sqrt{3}y\sqrt{V(x) + \frac{1}{2}y^2} - V'(x) \right). \tag{13.4.7}$$

13.4.2 Existence of a Lagrangian

We now have an equation of motion (13.4.6) for ϕ, obtained from the Friedmann and Klein-Gordon equations and defined by a potential $V(\phi)$. We are looking for a Lagrangian on the effective phase space $K \cong (x, y)$ from which an equivalent equation of motion can be derived. One reason for considering a Lagrangian description is that the direct approach, i.e., finding a closed-form solution to the Euler equation (13.4.3) for the vector field (13.4.7), is highly nontrivial for a typical potential.

The existence of a Lagrangian given an equation of motion is a famous question known as the inverse problem of the calculus of variations, which was finally solved by Douglas in 1941 [343]. See Ref. [344] for further reference. Suppose we have an equation of motion in a single variable

$$\ddot{x} = F(x, \dot{x}). \tag{13.4.8}$$

Then Douglas' theorem states that there exists a Lagrangian for which the Euler-Lagrange equation gives the correct equation of motion (13.4.8) if and only if there exists a function f satisfying the Helmholtz condition

$$\frac{\mathrm{d}f}{\mathrm{d}t} + \frac{\partial F}{\partial \dot{x}} f = 0, \tag{13.4.9}$$

or equivalently, with $y = \dot{x}$,

$$\frac{\partial f}{\partial t} + \frac{\partial}{\partial x}(\dot{x} f) + \frac{\partial}{\partial y}(F f) = 0. \tag{13.4.10}$$

For the problem at hand, defining $\phi = x$ and $\dot{\phi} = y$ as before, Eq. (13.4.6) can be written

$$F(x, y) = -\sqrt{3}y\sqrt{\frac{1}{2}y^2 + V(x)} - V'(x) = \ddot{x}. \tag{13.4.11}$$

Noting from Eq. (13.4.7) that $\mathbf{v} = (y,\ F)$, we are able to write the Helmholtz condition in the form

$$\frac{\partial f}{\partial t} + \frac{\partial}{\partial x}\left(v^x f\right) + \frac{\partial}{\partial y}\left(v^y f\right) = \frac{\partial f}{\partial t} + \nabla \cdot (f\mathbf{v}) = 0. \tag{13.4.12}$$

This is precisely the Euler equation for fluid flow (13.4.4), with f taking the place of the density. If there is a measure $f\ \mathrm{d}\dot{\phi} \wedge \mathrm{d}\phi$ on ϕ-$\dot{\phi}$ space conserved along the Hamiltonian flow vector, then $\nabla \cdot (f\mathbf{v}) = 0$. Thus, we have proven the following:

> There exists a Hamiltonian-flow conserved measure on ϕ-$\dot{\phi}$ space if and only if the equation of motion $\ddot{\phi} + \sqrt{3}\dot{\phi}\sqrt{\dot{\phi}^2/2 + V(\phi)} + V'(\phi) = 0$ possesses a Lagrangian description in effective phase space. More specifically, there exists a time-independent measure on ϕ-$\dot{\phi}$ space if and only if the Helmholtz condition is satisfied by a time-independent function $f(\phi, \dot{\phi})$.

In other words, a Lagrangian description of the equation of motion (13.4.11) [cf. Eq. (13.4.6)] exists if and only if the Helmholtz condition (13.4.9) is satisfied. In turn, the Helmholtz condition is satisfied if and only if there is a function f satisfying the Euler equation (13.4.4) for fluid flow with the Hamiltonian vector field (13.4.7). Whether or not there is such a function is difficult to establish in general, although we will give an argument in the case of $m^2\phi^2$ potentials.

13.4.3 The Conjugate Momentum and the Measure

If the Helmholtz condition (13.4.9) is satisfied, then the equation of motion can be written in the form

$$A\left(t, \phi, \dot{\phi}\right)\ddot{\phi} + B\left(t, \phi, \dot{\phi}\right) = 0, \tag{13.4.13}$$

where

$$\frac{\partial B}{\partial \dot{\phi}} = \left(\frac{\partial}{\partial t} + \dot{\phi}\frac{\partial}{\partial \phi}\right) A. \tag{13.4.14}$$

Explicitly, given f satisfying Eq. (13.4.12) and with $A = f(t, \phi, \dot{\phi})$ and $B = -f(t, \phi, \dot{\phi})\, F(\phi, \dot{\phi})$, where F is given in Eq. (13.4.11), it can be shown that Eq. (13.4.14) is satisfied and that Eq. (13.4.13) corresponds to the correct equation of motion (13.4.6). Then as shown in Ref. [344], the Lagrangian can be constructed explicitly:

$$\mathcal{L}_K\left(t, \phi, \dot{\phi}\right) = G\left(t, \phi, \dot{\phi}\right) + C\left(t, \phi\right), \tag{13.4.15}$$

where

$$\begin{aligned} G\big(t,\phi,\dot{\phi}\big) &= \dot{\phi}\int_0^1 \mathrm{d}\tau' \Big[\dot{\phi}\int_0^1 \mathrm{d}\tau A\big(t,\phi,\tau\dot{\phi}\big)\Big]\big(t,\phi,\tau'\dot{\phi}\big)\,, \\ C(t,\phi) &= \phi\int_0^1 \mathrm{d}\tau\, W\left(t,\tau\phi\right), \quad \text{and} \\ W(t,\phi) &= -B - \frac{\partial G}{\partial \phi} + \frac{\partial^2 G}{\partial\dot{\phi}\partial t} + \frac{\partial^2 G}{\partial\phi\partial\dot{\phi}}\dot{\phi}. \end{aligned} \tag{13.4.16}$$

We can then extract the momentum π_ϕ conjugate to ϕ in ϕ-$\dot{\phi}$ space via

$$\pi_\phi = \frac{\partial \mathcal{L}_K}{\partial \dot{\phi}} = \frac{\partial G}{\partial \dot{\phi}}. \tag{13.4.17}$$

Then the Liouville measure on ϕ-$\dot{\phi}$ space is

$$\mathrm{d}\pi_\phi \wedge \mathrm{d}\phi = \frac{\partial^2 G}{\partial \dot{\phi}^2}\mathrm{d}\dot{\phi} \wedge \mathrm{d}\phi. \tag{13.4.18}$$

With $A = f$, it can be shown from Eq. (13.4.16) that $\partial_{\dot{\phi}}^2 G = f$ and hence

$$\mathrm{d}\pi_\phi \wedge \mathrm{d}\phi = f\ \mathrm{d}\dot{\phi} \wedge \mathrm{d}\phi. \tag{13.4.19}$$

In the previous section, we demonstrated that finding a Lagrangian producing the equation of motion on ϕ-$\dot{\phi}$ space was the same problem as constructing a Hamiltonian-flow conserved measure on that space. The result we have proven in this section states something stronger:

> The natural Liouville measure $\mathrm{d}\pi_\phi \wedge \mathrm{d}\phi$ on effective phase space that one obtains from the effective Lagrangian, if it exists, is the same measure that one finds by explicitly constructing a nontrivial two-form $f\,\mathrm{d}\dot{\phi} \wedge \mathrm{d}\phi$ conserved under Hamiltonian evolution.

We note that these results are applicable to any single-field $V(\phi)$ model with canonical kinetic term, or with slight generalization, to any dynamical problem in a single variable.

13.5 The Measure on Effective Phase Space for Quadratic Potentials

It is illustrative to explicitly investigate the behavior of the measure on effective phase space for a specific model,

$$V(\phi) = \frac{1}{2}m^2\phi^2. \tag{13.5.1}$$

Ideally, one would like to obtain a closed-form expression for the measure; however, solving the partial differential equation explicitly proves to be prohibitive. We obtain expressions for the behavior of the measure in two limits, $H \gg m$ and $H \ll m$, which can be viewed as corresponding to early and late times. Finally, we use the Cauchy-Kowalevski theorem to prove that a unique measure obeying the constraint (13.4.3) exists, up to overall normalization.

13.5.1 Constraining the Behavior of the Measure

It is convenient at this point to reparameterize the vector field in terms of polar coordinates. Again setting $x = \phi$ and $y = \dot{\phi}$, let

$$r = \sqrt{y^2 + m^2x^2} = \sqrt{6}H \tag{13.5.2}$$

and

$$\tan\theta = \frac{y}{mx}. \tag{13.5.3}$$

Then the vector field (13.4.7) can be written as

$$\mathbf{v} = -\sqrt{\frac{3}{2}}r^2\sin^2\theta\,\hat{\mathbf{r}} - \left(mr + \sqrt{\frac{3}{2}}r^2\sin\theta\cos\theta\right)\hat{\boldsymbol{\theta}}, \tag{13.5.4}$$

where $\hat{\mathbf{r}} = \hat{\mathbf{x}}m^{-1}\cos\theta + \hat{\mathbf{y}}\sin\theta$ and $\hat{\boldsymbol{\theta}} = -\hat{\mathbf{x}}m^{-1}\sin\theta + \hat{\mathbf{y}}\cos\theta$ are unit vectors under the appropriate scaling of axes. The constraint (13.4.3) for a time-

independent measure f may be expressed as an explicit partial differential equation

$$-\frac{1}{r}\sqrt{\frac{3}{2}}\sin^2\theta\partial_r\left(r^3 f\right) - m\partial_\theta f - \sqrt{\frac{3}{2}}r\partial_\theta\left(f\sin\theta\cos\theta\right) = 0. \tag{13.5.5}$$

Late universe limit

The late-universe, $H \to 0$ limit corresponds to $r \to 0$. Suppose first that f does not diverge in this limit. Then as $\mathbf{v}\cdot\hat{\mathbf{r}} = -\sqrt{3/2}r^2\sin^2\theta < 0$ for all r, θ, it follows that for any circular disk R of radius $r_\star$ centered at the origin,

$$\int_R \nabla\cdot(f\mathbf{v})\mathrm{d}A = \oint_{\partial R}(f\mathbf{v})\cdot\hat{\mathbf{n}}\mathrm{d}\ell = -\sqrt{\frac{3}{2}}\int_0^{2\pi} f\left(r_\star,\theta\right) r_\star^3\sin^2\theta\,\mathrm{d}\theta. \tag{13.5.6}$$

Since the expression on the right-hand side is negative, $\nabla\cdot(f\mathbf{v})$ is not identically zero. If, however, f diverges as $r \to 0$, we must include another boundary term: effectively, the disk becomes an annulus, with the point at $r = 0$ removed. In general, this does not allow us to show that $\nabla\cdot(f\mathbf{v}) \not\equiv 0$. Thus, we conclude that f must diverge as $r \to 0$.

Near the origin, where r is small, physically corresponds to small Hubble parameter in units of the scalar field mass, $H \ll m$. In this limit, we may take the leading order in r in Eq. (13.5.5), as this will dwarf all other terms:

$$\partial_\theta\left(mf\right) \xrightarrow[r\to 0]{} 0. \tag{13.5.7}$$

This means that f is well behaved in its angular coordinate near the origin: we do not have any ambiguity in defining f as $r \to 0$ as would occur for, e.g., $f \to \sin\theta$. Near the origin, we have $f \to p\left(r\right)$, where $p\left(r\right)$ is the leading (i.e., most divergent) part of f. In general,

$$f\left(r,\theta\right) = p\left(r\right) + q\left(r,\theta\right), \tag{13.5.8}$$

where q is subleading as $r \to 0$.

Thus, Eq. (13.5.5) implies that for small r,

$$\sin^2\theta\left(r^3p' + 3r^2p\right) + \sqrt{\frac{2}{3}}mr\partial_\theta q + r^2p\left(\cos^2\theta - \sin^2\theta\right) = 0, \tag{13.5.9}$$

as all other terms, e.g., $\partial_r(r^3q)$, are of lesser order in r. A solution also has the requirement that $q(r,\theta)$ be periodic in θ. We obtain a solution to Eq. (13.5.9) if

$$r^3p' + 3r^2p = 0 \tag{13.5.10}$$

and

$$\sqrt{\frac{2}{3}}mr\partial_\theta q + r^2p\left(\cos^2\theta - \sin^2\theta\right) = 0, \tag{13.5.11}$$

which imply

$$p = \frac{C}{r^3} \tag{13.5.12}$$

and

$$q = -\sqrt{\frac{3}{2}}\frac{C}{mr^2}\sin\theta\cos\theta. \tag{13.5.13}$$

Thus, there exists a solution to Eq. (13.5.5) such that as $r \to 0$,

$$f \to C\left(\frac{1}{r^3} - \sqrt{\frac{3}{2}}\frac{\sin\theta\cos\theta}{mr^2}\right). \tag{13.5.14}$$

As the small r form (13.5.14) of the measure is divergent, with degree greater than 2, it is not normalizable over a region containing the point $\phi = \dot{\phi} = 0$. However, as we shall see in Sec. 13.5.2, excising the origin and considering the measure on the punctured plane allows us to prove that the measure exists and is well defined for $(\phi, \dot{\phi}) \neq (0,0)$.

A question that remains is whether this solution is unique, i.e., whether a nontrivial solution to Eq. (13.5.5) must have the behavior (13.5.14) near the origin. Writing any near-origin solution as $f(r,\theta) = p(r) + q(r,\theta)$ as above and demanding that $q(r,\theta)$ be periodic in θ implies that $\partial_\theta q$ must also be periodic

and have zero mean, i.e., $T^{-1}\int_0^T(\partial_\theta q)\,\mathrm{d}\theta \xrightarrow[T\to\infty]{} 0$. But from Eq. (13.5.9) we have

$$\partial_\theta q = -\sqrt{\frac{3}{2}}\frac{1}{m}\Big[rp\left(\cos^2\theta - \sin^2\theta\right) + \sin^2\theta\left(r^2p' + 3rp\right)\Big]. \tag{13.5.15}$$

At fixed r, this expression has zero mean only if $r^2p' + 3rp = 0$. Hence, the solution in Eqs. (13.5.12) and (13.5.13) is unique. That is, any nontrivial solution to Eq. (13.5.5) must have the form (13.5.14) as $r \to 0$.

Early universe limit

In the large r limit, which corresponds to $H \gg m$, we take the vector field (13.5.4) at leading order in r:

$$\mathbf{v} \approx -\sqrt{\frac{3}{2}}r^2\sin^2\theta\,\hat{\mathbf{r}} - \sqrt{\frac{3}{2}}r^2\sin\theta\cos\theta\,\hat{\boldsymbol{\theta}}. \tag{13.5.16}$$

Hence, the Euler constraint (13.4.3) [explicitly, Eq. (13.5.5)] requires, for large r, that the measure satisfy

$$\partial_\theta f = -r\tan\theta\partial_r f - (2\tan\theta + \cot\theta)\,f. \tag{13.5.17}$$

For f to be periodic in θ with period 2π for fixed r, we must have $\partial_\theta f$ periodic in θ with zero mean (i.e., $\partial_\theta f$ oscillates about zero). This requirement is satisfied by the odd functions $\tan\theta$ and $2\tan\theta + \cot\theta$, so $\partial_r f$ must be periodic in θ with period 2π. We therefore take $f(r,\theta)$ separable as an ansatz:

$$f(r,\theta) = R(r)\,\Theta(\theta). \tag{13.5.18}$$

Hence,

$$0 = r\frac{\partial_r R}{R} + 3 + \frac{\partial_\theta\left(\Theta\sin\theta\cos\theta\right)}{\Theta\sin^2\theta}, \tag{13.5.19}$$

which has solutions

$$R = Cr^{\gamma-3} \tag{13.5.20}$$

and

$$\Theta = C \csc\theta \cos^{\gamma-1}\theta, \tag{13.5.21}$$

where $\gamma \in \mathbb{R}$ is arbitrary. Demanding that f be positive everywhere, we have the leading order behavior

$$f \to C r^{\gamma-3} \left| \frac{\cos^{\gamma-1}\theta}{\sin\theta} \right| \quad \text{as} \quad r \to \infty. \tag{13.5.22}$$

The large r behavior for finite mass m is a weighted sum or integral of this family of solutions, determined by matching onto the measure for intermediate values of r. The coefficients of the sum must be found numerically for a particular value of m. Note that the $r \to \infty$ behavior of f given in Eq. (13.5.22) diverges near $y \to 0$. This corresponds to the apparent attractor solution plotted in Fig. 13.1: on large scales in effective phase space, the vector field points toward the ϕ axis, toward the apparent attractors near $\dot{\phi} = \pm\sqrt{2/3}m$. Any trajectory that starts out at large r is driven toward one of these attractor curves, which on large scales (equivalently, for small m) are very near $\dot{\phi} = 0$. Hence, the behavior of this solution is physically sensible. Imposing the condition $\gamma < 1$ makes the radial integral over large r convergent; this restriction could be viewed as physically reasonable, as evolving any universe backward must result in $H \to \infty$, i.e., the big bang, in finite time.

The measure near the apparent attractor

Comparing the $r \to 0$ behavior (13.5.14) and the $r \to \infty$ behavior (13.5.22) of the measure, we see that in both limits, f diverges wherever trajectories converge in ϕ-$\dot{\phi}$ space: for the early universe (large r) this occurs along the apparent attractor solution, approximated by $\dot{\phi} \approx 0$, while for small r this occurs at the origin, which corresponds to reheating. Therefore, it is well motivated to suppose that any solution to the full constraint equation $\nabla \cdot (f\mathbf{v}) = 0$ diverges along the full apparent attractor (Fig. 13.1), i.e., the curves in ϕ-$\dot{\phi}$ space

satisfying

$$\frac{\mathrm{d}\dot{\phi}}{\mathrm{d}\phi} = -\frac{\sqrt{\frac{3}{2}}\sqrt{m^2\phi^2 + \dot{\phi}^2}\dot{\phi} + m^2\phi}{\dot{\phi}} \tag{13.5.23}$$

with initial condition $\dot{\phi} = \pm\sqrt{2/3}m$, $\phi \to \mp\infty$. It is interesting to note the following similarity between the expressions (13.5.22) and (13.5.14) for the measure in the early and late universe. Defining d as the distance to the apparent attractor in the ϕ-$\dot{\phi}$ plane and considering successive ringdown orbits, it is possible to show that $f \sim 1/rd$ during reheating. Similarly, the $\gamma = 1$ solution to Eq. (13.5.22) also corresponds to $1/rd$ in the early universe. It is not clear and appears unlikely that these attractor solutions can be written in analytical form, which would imply that there is no analytical expression for the Lagrangian or measure.

13.5.2 Existence of the Measure for $m^2\phi^2$ Potentials

A natural question to ask, given the constraint imposed on a time-independent measure on the effective phase space K, is whether a nontrivial solution exists, i.e., does there exist a probability distribution f satisfying $\nabla \cdot (f\mathbf{v}) = 0$ for $\mathbf{v}$ given in Eq. (13.4.7)? In general, the answer to this question is dependent on the potential $V(\phi)$, but in light of the results (13.5.14) and (13.5.22), we will see that we can answer the question in the affirmative for an $m^2\phi^2$ potential.

As in Eq. (13.5.5), we express the partial differential equation that f must satisfy in polar coordinates in effective phase space. Define a function

$$g(r,\theta) = r^3 \sin^2\theta \left[f(r,\theta) - f_\epsilon\right], \tag{13.5.24}$$

where $f_\epsilon \equiv f(r \to \epsilon)$ is a constant, for some small $\epsilon > 0$. This expression is well defined because, as shown in Eq. (13.5.7), we must have $f(r,\theta) \xrightarrow[r\to 0]{} f(r)$, independent of θ. We are thinking of the solution for g as a Cauchy problem, with initial data

$$g\left(r=\epsilon,\theta\right)=0 \tag{13.5.25}$$

and evolution in r rather than t.

The constraint equation (13.5.5) for f becomes, in terms of g,

$$\begin{aligned}\partial_r g &= \sin^2\theta\partial_r\left(r^3 f\right)\\ &= -\sqrt{\frac{2}{3}}mr\partial_\theta f - r^2\partial_\theta\left(f\sin\theta\cos\theta\right)\\ &= -\sqrt{\frac{2}{3}}\frac{m}{r^2}\partial_\theta\left(g\csc^2\theta\right) - \frac{1}{r}\partial_\theta\left(g\cot\theta\right).\end{aligned} \tag{13.5.26}$$

Defining

$$a\left(r,\theta\right) = -\sqrt{\frac{2}{3}}\frac{m}{r^2}\csc^2\theta - \frac{1}{r}\cot\theta \tag{13.5.27}$$

and

$$b\left(\theta,g\right) = \left(\frac{1}{r} + 2\sqrt{\frac{2}{3}}\frac{m}{r^2}\cot\theta\right)g\csc^2\theta, \tag{13.5.28}$$

the constraint becomes

$$\partial_r g = a\left(r,\theta\right)\partial_\theta g + b\left(\theta,g\right). \tag{13.5.29}$$

A measure on ϕ-$\dot{\phi}$ space exists if and only if there is a solution g to the evolution equation (13.5.29) with initial data (13.5.25).

The function a is analytic in the entire upper half-plane $\mathbb{R}^2_+$ ($0<\theta<\pi$, corresponding to $\dot{\phi}>0$). The function b is analytic in terms of r in $\mathbb{R}^2_+$ and is analytic in terms of g near $g=0$. Since the upper half-plane is open, the Cauchy-Kowalevski theorem [345] guarantees that there exists a unique analytic solution to the evolution equation (13.5.29) in $\mathbb{R}^2_+$. We note that the specification of the initial data (13.5.25), in setting the constant f_ϵ, amounts to simply selecting the constant C such that $f\to Cr^{-3}$ [cf. Eq. (13.5.14)] near the origin. This argument also holds separately for the lower half-plane $\mathbb{R}^2_-$ ($\pi<\theta<2\pi$, corresponding to $\dot{\phi}<0$). The divergence in a and b at $\dot{\phi}=0$ is a coordinate singularity if $\phi\neq 0$; the vector field is finite and trajectories are

smooth as they cross the ϕ axis. We can impose continuity of the measure to select the same normalization in the upper and lower half-planes. The Cauchy-Kowalevski theorem thus guarantees existence and uniqueness everywhere for $r > \epsilon$. Strictly speaking, for any particular, finite ϵ, there will be higher-order corrections to the form Cr^{-3} of the measure (i.e., terms less divergent than r^{-3}), as computed in Eq. (13.5.14). However, the overall normalization C remains well defined, since we can also use the Cauchy-Kowalevski theorem to guarantee uniqueness upon evolving the measure inward for $r < \epsilon$. We can now consider a family of such Cauchy problems, for different values of ϵ but all with the same value of C. This family of measures is uniformly convergent as $\epsilon \to 0$, converging to the form Cr^{-3} near the origin. Thus, the Cauchy-Kowalevski theorem guarantees the existence, and uniqueness up to normalization, of the measure on the entire punctured plane $\mathbb{R}^2 \backslash \{(\phi = 0, \dot{\phi} = 0)\}$. In summary, we have proven the following result:

> Up to normalization, there exists a unique conserved measure on the effective phase space $(\phi, \dot{\phi})$ for scalar-field cosmologies with $m^2\phi^2$ potentials, excluding the origin.

It is well founded to conjecture that the effective phase-space measure always exists for any reasonable potential $V(\phi)$. A requisite property of the potential is that, if the the measure diverges on a set contained in a neighborhood W (e.g., during reheating), there is a unique solution in an open neighborhood $U \backslash W$, cf. Eq. (13.5.14). This is equivalent to the statement that the measure on K does not behave chaotically as the boundary of U with W is varied.

13.6 The Physical Meaning of Attractors

We have seen that it is possible to define a conserved measure on K, the effective ϕ-$\dot{\phi}$ phase space of scalar-field cosmology in a flat universe. This was

shown for $V(\phi) = m^2\phi^2/2$, and seems likely to hold for other smooth potentials in single-field models. Because Liouville's theorem is obeyed with respect to such a measure (unlike the naive measure $\mathrm{d}\dot{\phi} \wedge \mathrm{d}\phi$), true attractor behavior is impossible. The apparent attractor behavior familiar in cosmology is an artifact of using the ϕ-$\dot{\phi}$ coordinates, which are not canonically conjugate. It is nevertheless worth asking whether there are other ways of thinking about attractors that are physically meaningful. In this section we suggest two possibilities.

The first possibility rests on the idea that ϕ and $\dot{\phi}$, while not canonically conjugate, are nevertheless the directly physically observable features of the scalar field. In this sense they define preferred coordinates in which to follow the evolution. If one accepts that notion, we can define an apparent attractor as a region in phase space for which Lyapunov exponents are highly negative along particular axes. With trajectories $\mathbf{x}(t)$ in phase space labeled by coordinates x^α indexed by α, we define the Lyapunov exponent [346] along each coordinate axis,

$$\lambda_\alpha = \lim_{t\to\infty} \lim_{\delta x^\alpha(0)\to 0} \frac{1}{t} \log \frac{|\delta x^\alpha(t)|}{|\delta x^\alpha(0)|}, \tag{13.6.1}$$

where $\delta x^\alpha(t)$ is the separation between two trajectories, in the α-coordinate, at time t. (Note that λ_α is a function of position in phase space.) Then

$$|\delta x^\alpha(t)| \sim e^{\lambda_\alpha t} |\delta x(0)|. \tag{13.6.2}$$

By Liouville's theorem, the sum of the Lyapunov exponents in canonically conjugate coordinates is zero and, in fact, a canonical transformation of the coordinates on phase space can be made such that all the Lyapunov exponents vanish [337]. However, in the case of the linear attractors for $m^2\phi^2$ potentials located near $\dot{\phi} = \pm\sqrt{2/3}m$, the Lyapunov exponent in the $\dot{\phi}$ direction is negative, forcing trajectories to appear (in ϕ-$\dot{\phi}$ space) to converge (cf. Fig

13.1). This definition of an apparent attractor is consistent with the common motivation for using ϕ-$\dot{\phi}$ space in the first place: the coordinates are physically intuitive and trajectories exhibit "attractor-like" behavior. In contrast, plotting in (ϕ, π_ϕ), where π_ϕ is the canonical momentum (13.4.17) associated with ϕ under the Lagrangian on effective phase space (Sec. 13.4.3), will cause the apparent attractor behavior to vanish. In (ϕ, π_ϕ) coordinates, bundles of trajectories will not shrink, but will instead contract along one axis while expanding along the other. While the Lyapunov exponent characterization is manifestly coordinate-dependent, it has the virtues of being intuitive and capturing the sense in which the word "attractor" is used in much of the literature, cf. Ref. [333].

Another point worth emphasizing is that our analysis here has been entirely classical. In real cosmological evolution, there will be a boundary in phase space past which a classical picture is inadequate; we expect that this would occur at least when any physical quantity (the Hubble constant, the radius of curvature, or the field energy) reached the Planck scale. If one had a true theory of cosmological initial conditions that implied a probability measure for trajectories near this boundary, that would presumably supersede the classical Liouville-type measures we have been focusing on in this chapter. In the absence of such a theory, of course, it makes more sense to use such well-defined classical measures rather than to place too much emphasis on the naive choice $\mathrm{d}\dot{\phi} \wedge \mathrm{d}\phi$.

13.7 Conclusions

It is common in literature on inflation (as well as quintessence models) to make use of the idea of attractor solutions. However, this notion is not well defined for a Hamiltonian system. In this chapter, we have attempted to clarify the

relationship between the dissipationless dynamics of scalar-field cosmology and apparent attractor behavior.

We showed that, in a universe with vanishing spatial curvature, there is a sense in which ϕ and $\dot{\phi}$ become effective phase-space variables. Namely, the map from the four-dimensional phase space to ϕ-$\dot{\phi}$ space is vector field invariant under the Hamiltonian flow vector. As a result, trajectories do not cross in ϕ-$\dot{\phi}$ space and, in this mapping, the phase space is effectively two-dimensional and we observe the appearance of attractor-like behavior. In making (coordinate-dependent) observations about "attractors," many authors plot in $(\phi, \dot{\phi})$ coordinates, despite their noncanonicity, and suppress the other two phase-space dimensions.

We explored the existence of a conserved measure on the effective phase space. The GHS measure, while possessing many useful properties, diverges for flat universes and so can give no information about the measure on the effective phase space of $(\phi, \dot{\phi})$. Such a measure can be constructed "from scratch" by finding a two-form $\boldsymbol{\sigma} = f(\phi, \dot{\phi}) \mathrm{d}\dot{\phi} \wedge \mathrm{d}\phi$ on ϕ-$\dot{\phi}$ space that is conserved under Hamiltonian flow (that is, whose Lie derivative along the Hamiltonian flow vector vanishes). Using Douglas' theorem and the Helmholtz condition, we proved that, for $V(\phi)$ inflation, such a measure constructed in this way exists if and only if there is a Lagrangian description of the system in the two-dimensional effective phase space. Furthermore, using this Lagrangian, one can define a momentum conjugate to ϕ in the effective phase space, $\pi_\phi = \partial \mathcal{L}_K / \partial \dot{\phi}$ (not to be confused with p_ϕ, the momentum conjugate to ϕ in the full four-dimensional phase space), and use this to define a measure $\mathrm{d}\pi_\phi \wedge \mathrm{d}\phi$. We proved that this measure is identical to the measure $f\, \mathrm{d}\dot{\phi} \wedge \mathrm{d}\phi$ that one constructs "from scratch"; demanding conservation under Hamiltonian flow is enough to specify the measure. For the specific model of inflation with a quadratic potential,

we found the behavior that the effective phase-space measure must possess in the late ($H \ll m$) and early ($H \gg m$) limits and used the Cauchy-Kowalevski theorem to prove that a unique analytic solution for the measure exists up to normalization, provided ϕ and $\dot{\phi}$ do not both vanish. It is reasonable to conjecture that a similar existence/uniqueness result should hold for a large class of potentials.

Finally, we discussed the meaning of apparent attractors. While the dynamics of scalar-field cosmology is conservative, evolution can nevertheless approach certain characteristics if we express it in terms of preferred variables. It can happen, for example, that the Lyapunov exponents can be negative along certain axes. By Liouville's theorem, these more general formulations of apparent attractors are necessarily coordinate-dependent.

The idea of attractor-like behavior is central to the intuitive idea of inflationary cosmology: the development of a smooth, flat FRW universe from a large set of initial conditions. Despite the fact that this behavior cannot occur in canonical phase-space variables, it is useful to consider how the notion of an apparent attractor can best be defined to capture this intuition. This helps clarify the idea of naturalness in cosmological evolution.

Chapter 14
How Many e-Folds Should We Expect from High-Scale Inflation?

We address the issue of how many e-folds we would naturally expect if inflation occurred at an energy scale of order $10^{16}\,\mathrm{GeV}$. We use the canonical measure on trajectories in classical phase space, specialized to the case of flat universes with a single scalar field. While there is no exact analytic expression for the measure, we are able to derive conditions that determine its behavior. For a quadratic potential $V(\phi) = m^2\phi^2/2$ with $m = 2 \times 10^{13}\,\mathrm{GeV}$ and cutoff at $M_{\mathrm{Pl}} = 2.4 \times 10^{18}\,\mathrm{GeV}$, we find an expectation value of 2×10^{10} e-folds on the set of Friedmann-Robertson-Walker trajectories. For cosine inflation $V(\phi) = \Lambda^4[1 - \cos(\phi/f)]$ with $f = 1.5 \times 10^{19}\,\mathrm{GeV}$, we find that the expected total number of e-folds is 50, which would just satisfy the observed requirements of our own universe; if f is larger, more than 50 e-folds are generically attained. We conclude that one should expect a large amount of inflation in large-field models and more limited inflation in small-field (hilltop) scenarios.

This chapter is from Ref. [13], G. N. Remmen and S. M. Carroll, "How Many e-Folds Should We Expect from High-Scale Inflation?," Phys. Rev. ***D90*** *(2014) 063517,* arXiv:1405.5538 [hep-th]. *Note that, while the specific observations of the BICEP2 experiment [347] have been largely attributed to galactic dust rather than cosmological tensor perturbations [348] in the time since Ref. [13] was published, the results of this chapter remain completely applicable to a large class of inflationary models as a means of quantifying tuning for different theories of inflation given cosmological observations.*

14.1 Introduction

The possible detection of tensor perturbations in the cosmic microwave background (CMB) by the BICEP2 experiment [347] suggests that inflation occurred at a high energy scale [349]: $E_{\rm I} = 2 \times 10^{16}$ GeV, just two orders of magnitude below the reduced Planck scale $M_{\rm Pl} = 1/\sqrt{8\pi G} = 2.4 \times 10^{18}$ GeV. Knowing this parameter with some confidence allows both for much more focused inflationary model-building and for quantitative exploration of some of the conceptual issues underlying the inflationary paradigm. In this chapter, we address one of the latter: given an inflaton potential that is able to reproduce the measured cosmological parameters, how much inflation is likely to have occurred? In the present chapter, we answer this question, finding that the expected number of e-folds of inflation depends dramatically on the general type of inflaton potential chosen.

The amount of inflation that occurs is measured by the number of e-folds,

$$N = \int_{a_{\rm i}}^{a_{\rm f}} {\rm d} \log a = \int_{t_{\rm i}}^{t_{\rm f}} H \, {\rm d}t. \tag{14.1.1}$$

Here, $a_{\rm i}$ and $a_{\rm f}$ are the values of the scale factor at the beginning and end of inflation, while $t_{\rm i}$ and $t_{\rm f}$ are the corresponding proper times. We can define the period during which inflation is occurring as that for which the universe is accelerating, $\ddot{a} > 0$. In conventional inflationary models, it is necessary to achieve at least 50 e-folds to successfully address the horizon problem. It is generally accepted that this requirement can be met by a wide variety of potentials.

We would like to know not only whether a certain potential can possibly produce sufficient amounts of inflation, but whether such an outcome is actually likely. Presumably, a complete theory of cosmological initial conditions in the context of quantum gravity would provide a unique answer to this question,

but we don't have such a theory at present. What we do have are classical models of inflaton dynamics coupled to general relativity. Any classical theory comes with a natural measure on phase space, the Liouville measure. Gibbons, Hawking, and Stewart (GHS) showed how to use this measure to define a canonical measure on cosmological *trajectories* (rather than individual points in phase space) [64]. In this measure, we can calculate the fraction of universes with given properties, such as "more than 50 e-folds of inflation." Given the current state of the art, this is the best we can do to decide whether such solutions are likely or not.

The GHS measure comes with a technical problem when applied to (homogeneous, isotropic) Friedmann-Robertson-Walker (FRW) cosmologies: it diverges as the spatial curvature approaches zero, assigning almost all measure to flat universes. Different proposals have been advanced for dealing with this divergence, including removing the region of infinite measure by hand [340]. As noted in Refs. [12, 339], the divergence for flat universes is an indication that, in the canonical measure, almost all cosmological spacetimes are flat. For this reason, and also given the physical relevance of spatially-flat solutions [65], it is on these that we concentrate our efforts. In a previous paper [12], we developed a formalism for defining the Hamiltonian-conserved measure on the effective two-dimensional phase space for a canonical scalar field with a potential in a flat FRW cosmology. Although we did not prove the uniqueness of this measure in arbitrary theories, we could establish it for quadratic potentials and expect it to hold for well-behaved potentials more generally.

In this chapter, we employ the formalism developed in Ref. [12] to study high-scale inflation. We focus on two representative models: quadratic inflation and cosine ("natural") inflation. We find dramatically different quantitative results for the two cases. In quadratic inflation, given that the potential is

chosen to fit observed cosmological parameters, we find that large amounts of inflation are favored by the canonical measure—billions of e-folds of inflation — provided we extrapolate the quadratic potential up to the Planck scale $H = M_{\rm Pl}$ and allow the inflaton field ϕ to run over a super-Planckian range $\sim 10^5 M_{\rm Pl}$. Moreover, we find that almost all trajectories experience well more than 50 e-folds. For cosine potentials, by contrast, the expected amount of inflation under the canonical measure is relatively small: if the symmetry-breaking parameter f is set to the reduced Planck scale, $M_{\rm Pl} = 2.4 \times 10^{18}$ GeV, we expect of order one e-fold, with the probability of attaining as many as 50 e-folds being exponentially small. These numbers depend sensitively on f; once it is above 10^{19} GeV, as favored by the BICEP2 result [347, 350], the probability of getting more than 50 e-folds rises above 50%.

This last result is interesting, since cosine potentials feature "hilltops" from which trajectories with arbitrarily large numbers of e-folds can originate. Our analysis demonstrates that, while such lingering solutions are allowed, they contribute a relatively small amount to the measure on the space of trajectories. We conjecture that this behavior reflects a more general difference between potentials that rise up to the Planck scale, in which we expect large amounts of inflation, and models with potential maxima below the Planck scale, where the expected number of e-folds will be comparatively small.

Any analysis of this form necessarily comes with caveats. As noted, we are using a classical measure, whereas a particular theory of initial conditions (e.g., a proposal for the wave function of the universe) will presumably make its own predictions. More seriously, our analysis applies only to universes that are assumed to be homogeneous from the start. Once perturbations are included, it is clear that most universes should be wildly inhomogeneous; the existence of the sufficiently smooth initial conditions necessary for inflation to

begin is highly non-generic [339, 351]. Given the evidence that inflation did happen, we consider the expected number of e-folds according to the canonical measure to be a useful diagnostic of which models are robust and which are more delicate. An ultimate justification for why inflation occurs in the first place awaits further insight.

This chapter is organized as follows. In Sec. 14.2 we first review the formalism of Refs. [12, 64] for finding the canonical measure on phase space, as well as the sense in which phase space becomes effectively only two-dimensional for flat FRW cosmologies. The connection between the measure on effective phase space and the measure on the space of possible trajectories of evolution of a FRW universe is presented. Next, in Sec. 14.3 we derive some general properties of the measure for arbitrary slow-roll and hilltop potentials. Finally, we examine representative models of each class, quadratic inflation and cosine inflation, in Secs. 14.4 and 14.5, making statistical calculations on the ensemble of all FRW universes and finding the expected number of e-folds of inflation attained.

14.2 The Probability Distribution on the Set of Universes

14.2.1 The Hamiltonian-Conserved Measure

We are interested in the theory of a homogeneous scalar field in an expanding FRW universe. The action is

$$S = \int \mathrm{d}^4x\,\sqrt{-g}\left[\frac{M_{\rm Pl}^2}{2}R - \frac{1}{2}g^{\mu\nu}\partial_\mu\phi\partial_\nu\phi - V(\phi)\right]. \tag{14.2.1}$$

The metric can be written

$$\mathrm{d}s^2 = -N^2(t)\mathrm{d}t^2 + a^2(t)\left(\frac{\mathrm{d}r^2}{1-\kappa r^2} + r^2\mathrm{d}\Omega^2\right), \tag{14.2.2}$$

where N is the lapse function and the curvature parameter κ is an arbitrary real parameter with mass dimension 2. The number κ is fixed for a given FRW universe and we can write $k = \kappa R_0^2 \in \{-1, 0, 1\}$, where R_0 is the radius of curvature of the universe at unit scale factor. Taking $\phi(t)$ to depend only on time, the Hamiltonian is

$$\begin{aligned}\mathcal{H} &= -3a^3 N M_{\mathrm{Pl}}^2 \left\{ \frac{\dot{a}^2}{a^2} + \frac{\kappa}{a^2} - \frac{1}{3M_{\mathrm{Pl}}^2} \left[\frac{1}{2}\dot{\phi}^2 + V(\phi) \right] \right\} \\ &= N \left[-\frac{p_a^2}{12 a M_{\mathrm{Pl}}^2} + \frac{p_\phi^2}{2a^3} + a^3 V(\phi) - 3a\kappa M_{\mathrm{Pl}}^2 \right],\end{aligned} \tag{14.2.3}$$

where p_a and p_ϕ are the momenta conjugate to the scale factor and scalar field, respectively. The scalar equation of motion is

$$\ddot{\phi} + 3H\dot{\phi} + V'(\phi) = 0, \tag{14.2.4}$$

where $V'(\phi) = \mathrm{d}V/\mathrm{d}\phi$. The Hamiltonian constraint, which comes from varying with respect to N, is the Friedmann equation,

$$H^2 = \frac{1}{3M_{\mathrm{Pl}}^2} \left[\frac{1}{2}\dot{\phi}^2 + V(\phi) \right] - \frac{\kappa}{a^2}, \tag{14.2.5}$$

where $H = \dot{a}/a$ is the Hubble parameter.

Any classical theory comes with a preferred choice of measure on phase space: the Liouville measure, which is preserved under time evolution. In cosmology our interest is less in a measure on individual points in phase space and more in a measure on trajectories through time or specific cosmological evolutions. Gibbons, Hawking, and Stewart [64] showed how to construct such a measure for a scalar field coupled to general relativity. The phase space is naively four-dimensional, with coordinates given by a and ϕ and their conjugate momenta. But the Hamiltonian constraint, implemented by the Friedmann equation, cuts this down to three dimensions. The space of trajectories (equivalent under the equations of motion to the space of initial conditions) is one lower, leaving us with a two-dimensional space. GHS were

able to construct a unique measure on this space that is positive and invariant under time evolution (for further discussion see Refs. [12, 339, 340]).

As Ref. [339] shows, the GHS measure [64] has an interesting property: on a transverse surface in phase space defined by fixed Hubble parameter, the measure diverges for small curvature κ as $|\Omega_k|^{-5/2}$. This behavior has the good feature that it implies that the collection of non-flat FRW universes is a set of measure zero under the GHS measure; that is, the flatness problem in cosmology is solved by the GHS measure, since almost all trajectories are flat. However, from the point of view of understanding the set of flat FRW universes itself, this behavior poses a technical challenge. It is difficult to regularize the divergence in the GHS measure to construct a well-defined measure within the space of flat universes.

In our previous paper, we showed how to find a measure on the space of flat universes by constructing it by hand, subject to the requirement that it be conserved under time evolution [12]. We note from Eqs. (14.2.4) and (14.2.5) that the scale factor a disappears from the equations of motion when $\kappa = 0$. The effective phase space is therefore only two-dimensional; specifying the two quantities ϕ and $\dot{\phi}$ completely determines the solution (although they are not conjugate variables). The set of trajectories in effective phase space is therefore one-dimensional. In Ref. [12] we formalized the notion of an effective phase space via the property of vector field invariance between two manifolds. We argued that there exists a unique measure on this space that is conserved under Hamiltonian flow, in analogy with the conventional Liouville measure, which one can use to construct a measure on the space of flat universes.

The time evolution given by Eq. (14.2.4) can be characterized by a vector field $\mathbf{v}$ on ϕ-$\dot{\phi}$ space, with components

$$\mathbf{v} = \left(\dot{\phi}, -V'(\phi) - 3H\dot{\phi}\right). \tag{14.2.6}$$

The Hubble parameter (and thus the scale factor, up to an irrelevant scaling) is then fixed by Eq. (14.2.5). We seek a two-form

$$\boldsymbol{\sigma} = \sigma(\phi, \dot{\phi})\, \mathrm{d}\dot{\phi} \wedge \mathrm{d}\phi \tag{14.2.7}$$

that is conserved under evolution,

$$\pounds_{\mathbf{v}} \boldsymbol{\sigma} = 0. \tag{14.2.8}$$

Using the definition of the Lie derivative and rearranging, we can equivalently write in component form

$$\partial_\mu (\sigma v^\mu) = 0, \tag{14.2.9}$$

where $\partial_\mu \equiv \partial/\partial x^\mu$ and $x^\mu = (\phi, \dot{\phi})$. A two-form $\boldsymbol{\sigma}$ for which σ satisfies the Hamiltonian-conservation constraint (14.2.9)—the same as the Euler equation for stationary fluid flow—is the natural measure on the effective phase space, exactly in analogy with the Liouville measure. We will call the function σ, which forms the probability distribution on effective phase space in a given coordinate system, the *measure density*.

At this point, it is natural to ask whether there is a Lagrangian description $\mathcal{L}_\Phi$ of the trajectories on the effective phase space Φ. Using Douglas's theorem and the Helmholtz conditions, we showed [12] that there exists a time-independent Lagrangian description of the equation of motion (14.2.4) on effective phase space if and only if there exists a Hamiltonian-conserved measure: in fact, finding the Lagrangian gives a measure satisfying Eq. (14.2.9) and vice versa. Further, defining $\pi_\phi = \partial \mathcal{L}_\Phi / \partial \dot{\phi}$ as the conjugate momentum on Φ, one finds that the Liouville measure $\mathrm{d}\pi_\phi \wedge \mathrm{d}\phi$ on effective phase space under $\mathcal{L}_\Phi$ is just equal to $\sigma \mathrm{d}\dot{\phi} \wedge \mathrm{d}\phi$, obtained merely by demanding conservation under Hamiltonian evolution.[14.1] For the specific example of $m^2\phi^2$ inflation, we

[14.1] The corresponding Hamiltonian on effective phase space, $\mathcal{H}_\Phi = \pi_\phi \dot{\phi} - \mathcal{L}_\Phi$, is of course not subject to any additional Hamiltonian constraint as in the full phase space; that is, the Friedmann equation is merely a redefinition of coordinates on Φ and does not constrain $\mathcal{H}_\Phi$. With this definition, the measure can be written as $\mathrm{d}\mathcal{H}_\Phi \wedge \mathrm{d}t$.

proved that such a measure exists and is unique; such an existence/uniqueness result likely holds for any reasonably well-behaved potential $V(\phi)$.

14.2.2 The Space of Trajectories

Given the appropriate measure on our effective phase space, one can use this to determine the natural measure on the space of trajectories. In general, given some arbitrary measure density on a two-dimensional manifold and a one-parameter collection of curves that cover the manifold, there is not a well-defined probability distribution on the set of curves. However, the Hamiltonian-conserved measure density on effective phase space is not an arbitrary function vis-à-vis the family of trajectories. Following Refs. [12, 64], we can construct a measure on the space of trajectories in terms of a one-dimensional measure on any curve transverse to those trajectories, by demanding that the physical result be independent of our choice of transverse curve.

We begin by choosing some curve in the ϕ-$\dot{\phi}$ effective phase space on which to evaluate the measure density $\sigma(\phi, \dot{\phi})$. For simplicity, we'll imagine choosing $H =$constant surfaces, but any other slicing transverse to the trajectories that evolves monotonically in time would work just as well.[14.2] We can reparameterize ϕ-$\dot{\phi}$ space in terms of H and another coordinate, which we will call θ. For the bundle of trajectories that, on the H_1 surface, is centered at θ_1 and spans $\mathrm{d}\theta_1$, we write the measure as $P(\theta_1)|_{H_1}\,\mathrm{d}\theta_1$. Suppose this bundle of trajectories evolves to $H = H_2$, on which surface it is centered at θ_2 and spans $\mathrm{d}\theta_2$. We could equivalently write its probability measure as $P(\theta_2)|_{H_2}\,\mathrm{d}\theta_2$. Of course, the functional forms of $P(\theta_1)|_{H_1}$ and $P(\theta_2)|_{H_2}$ can be very different. However, this is the same bundle of trajectories, so for the measure on the space of trajectories to be well defined, we require

[14.2] Note that, regardless of the potential, the scalar equation assures that H evolves monotonically in time, with $\dot{H} = -\dot{\phi}^2/2M_{\mathrm{Pl}}^2$.

$$P\left(\theta_1\right)|_{H_1} \mathrm{d}\theta_1 = P\left(\theta_2\right)|_{H_2} \mathrm{d}\theta_2. \tag{14.2.10}$$

Now, we note that, given a parcel on effective phase space covering the region $\mathrm{d}\theta_1 \mathrm{d}H_1$ that evolves to $\mathrm{d}\theta_2 \mathrm{d}H_2$, we have

$$\sigma\left(H_1, \theta_1\right) \mathrm{d}\theta_1 \mathrm{d}H_1 = \sigma\left(H_2, \theta_2\right) \mathrm{d}\theta_2 \mathrm{d}H_2. \tag{14.2.11}$$

This is just the statement of Liouville's theorem for effective phase space, i.e., the requirement that σ satisfy (14.2.9). Hence, the correct way to compute $P(\theta)|_H$, the probability distribution on the space of trajectories, parameterized by the coordinate θ with which the trajectory intersects the H surface, is

$$P\left(\theta\right)|_H \propto \sigma\left(H, \theta\right) \mathrm{d}H. \tag{14.2.12}$$

We can divide through by $\mathrm{d}t$, since t evolves uniformly for all trajectories. We therefore have

$$P\left(\theta\right)|_H = \frac{\sigma\left(H, \theta\right) |\dot{H}|}{\int \sigma\left(H, \theta'\right) |\dot{H}| \mathrm{d}\theta'}. \tag{14.2.13}$$

Note that we suppressed the arguments (H, θ) of $\dot{H}$.

Eq. (14.2.13) is the important expression for this chapter. The measure on the space of trajectories is constructed by finding a conserved measure density σ on the effective phase space and evaluating $|\dot{H}|$ times this measure along a surface of constant H.

As a consistency check, we can derive Eq. (14.2.13) in a slightly different way. If we had written the effective phase space measure in the coordinates $(t,\, \theta)$ as $\tilde{\sigma}(t, \theta) \mathrm{d}\theta \wedge \mathrm{d}t = -\sigma(H, \theta) \mathrm{d}\theta \wedge \mathrm{d}H$ (with the minus sign compensating for the fact that H decreases with t, so that σ and $\tilde{\sigma}$ are positive) we could have equivalently defined the measure on the space of trajectories by explicitly performing the integration over t:

$$P\left(\theta_0\right)|_{H_0} \propto \int_0^\infty \tilde{\sigma}\left(t, \theta\left(t\right)\right) \mathrm{d}t, \tag{14.2.14}$$

where the path $(t, \theta(t))$ is chosen such that $\theta(t_0) = \theta_0$ and $H(t_0) = H_0$ for some t_0. Since t evolves uniformly for all trajectories, we have

$$P(\theta_0)|_{H_0} \propto \tilde{\sigma}(t_0, \theta_0)|_{H(t_0)=H_0} = \sigma(H_0, \theta_0)\,|\dot{H}|, \tag{14.2.15}$$

in agreement with Eq. (14.2.13).

We are now equipped to make quantitative statements about probabilities of different FRW trajectories for universes with zero curvature and compare these predictions for different models of inflation.

14.3 The Effective Phase Space Measure for Generic Potentials

Before examining specific models of inflation, it will first be informative to examine the behavior of the effective phase space measure $\boldsymbol{\sigma}$, without assuming an explicit functional form of the potential, in two representative classes of inflation: slow roll down a potential and quasi-de Sitter inflation near a local maximum in a potential, i.e., a hilltop. The cases are distinct because the fixed point in effective phase space corresponding to a stationary field at a potential maximum is a distinguished trajectory by itself and must be treated carefully.

For this analysis it will be useful to define dimensionless coordinates

$$x = \frac{\phi}{M_{\rm Pl}} \quad \text{and} \quad y = \frac{\dot{\phi}}{M_{\rm Pl}^2}, \tag{14.3.1}$$

which form a vector

$$\mathbf{x} = (x, y). \tag{14.3.2}$$

We then define a dimensionless speed in effective phase space

$$\tilde{\mathbf{v}} \equiv \frac{\dot{\mathbf{x}}}{M_{\rm Pl}} = \left(y, -\tilde{V}'(x) - \sqrt{3}y\sqrt{y^2/2 + \tilde{V}(x)}\right), \tag{14.3.3}$$

defining $\tilde{V}(x) \equiv V(\phi(x))/M_{\rm Pl}^4$ as a dimensionless potential and notation $\tilde{V}'(x) \equiv \mathrm{d}\tilde{V}/\mathrm{d}x$.

It will also be useful to define a norm for vectors and covectors using a flat fiducial metric:

$$|\mathbf{c}| \equiv [(c^1)^2 + (c^2)^2]^{1/2}. \tag{14.3.4}$$

The definition of the norm is simply a mathematical convenience; the fiducial metric should not be regarded as a physical metric on effective phase space. It will be convenient in the analysis below, where we derive conditions on the behavior of the measure density, although these conditions would hold even without using the norm notation. Of course, the physical content of the results is independent of the choice of metric, though the expressions themselves would look different for various choices of norm. As usual, placing bars around scalar quantities, e.g., $|\partial_\mu \tilde{v}^\mu|$, simply denotes absolute value.

Define the first potential slow-roll parameter

$$\epsilon_V \equiv \frac{M_{\rm Pl}^2}{2}\left[\frac{V'(\phi)}{V(\phi)}\right]^2 = \frac{1}{2}\left[\frac{\tilde{V}'(x)}{\tilde{V}(x)}\right]^2 \tag{14.3.5}$$

and the first Hubble slow-roll parameter:

$$\varepsilon \equiv -\frac{\dot{H}}{H^2} = \frac{\dot{\phi}^2}{2H^2 M_{\rm Pl}^2} = 3\frac{y^2}{y^2 + 2\tilde{V}(x)}. \tag{14.3.6}$$

Substituting Eq. (14.3.6) into Eq. (14.3.3) and rearranging, one finds

$$\frac{|\tilde{\mathbf{v}}|^2}{\tilde{V}(x)^2} = \frac{4\varepsilon^2 y^{-2} + 18\varepsilon}{(3-\varepsilon)^2} + 2\epsilon_V + 6s\frac{\sqrt{2\epsilon_V}\sqrt{2\varepsilon}}{3-\varepsilon}. \tag{14.3.7}$$

Here, $s \equiv \mathrm{sgn}[y\tilde{V}'(x)] = \pm 1$ indicates whether the potential is increasing ($s = +1$) or decreasing ($s = -1$) in the direction along which the field is evolving; we will generally have $s = -1$ during inflation. Furthermore, after simplifying with Eq. (14.2.5), we have

$$\frac{\partial_\mu \tilde{v}^\mu}{|y|} = -\frac{3}{\sqrt{2\varepsilon}} - \sqrt{\frac{\varepsilon}{2}}, \tag{14.3.8}$$

where ∂_μ denotes partial differentiation with respect to the dimensionless coordinates x^μ in Eq. (14.3.1). Note that Eqs. (14.3.7) and (14.3.8) are exact

expressions: no slow-roll approximation has yet been made.

14.3.1 Slow Roll down a Potential

As we discussed in Ref. [12], slow-roll behavior corresponds to apparent attractors in effective phase space—places where the conserved measure grows large. In this subsection we consider monotonic slow-roll behavior, characterized by two conditions imposed on Eqs. (14.2.4) and (14.2.5):

$$\begin{aligned} \dot{\phi}^2 \ll |V(\phi)| \;\text{ so }\quad & H^2 \simeq \frac{1}{3M_{\rm Pl}^2} V(\phi) \\ |\ddot{\phi}| \ll |H\dot{\phi}|, |V'(\phi)| \;\text{ so }\quad & 3H\dot{\phi} \simeq -V'(\phi)\,. \end{aligned} \tag{14.3.9}$$

Then $\varepsilon \approx \epsilon_V \equiv \epsilon \ll 1$ and we have from Eq. (14.3.7):

$$\frac{|\tilde{\mathbf{v}}|^2}{\tilde{V}(x)^2} \simeq \frac{4\epsilon^2}{9y^2} + 4(1+s)\epsilon. \tag{14.3.10}$$

Further, imposing $H^2 \ll M_{\rm Pl}^2$, so $\varepsilon \gg y^2$,

$$|\tilde{\mathbf{v}}| \simeq |y| \simeq \sqrt{\frac{2}{3}\epsilon \tilde{V}(x)}. \tag{14.3.11}$$

Similarly, in the slow-roll regime,

$$\partial_\mu \tilde{v}^\mu \simeq -\frac{3|y|}{\sqrt{2\epsilon}} \simeq -\sqrt{3\tilde{V}(x)}. \tag{14.3.12}$$

Note that the second slow-roll condition in Eq. (14.3.9) does not necessarily apply near a hilltop, as $|H\dot{\phi}| \gg |\ddot{\phi}|$ can fail. This is an important distinction; as we will see, behavior of the measure density near a hilltop in effective phase space is very different from what we find for trajectories that are uniformly slowly rolling down a potential. Our slow-roll conditions in this subsection are most compatible with potentials with $V''(\phi) > 0$, such as monomial models, in which hilltop behavior is manifestly absent.

Now, we examine what implications our analysis has for the form of the measure $\boldsymbol{\sigma} = \sigma(x,y)\mathrm{d}y \wedge \mathrm{d}x$ on effective phase space. With the requirement (14.2.9) that the measure be conserved under Hamiltonian evolution, Eqs. (14.3.11) and (14.3.12) imply that

$$\frac{(\partial_\mu \log\sigma)\tilde{v}^\mu}{|\tilde{\mathbf{v}}|} = \frac{|\partial_\mu \tilde{v}^\mu|}{|\tilde{\mathbf{v}}|} \simeq \frac{3}{\sqrt{2\epsilon}} \tag{14.3.13}$$

near the slow-roll regime and for $H^2 \ll M_{\rm Pl}^2$. Note that the left side of Eq. (14.3.13) is just the gradient of $\log\sigma$ along a slow-roll curve; thus, the closer to slow roll we approach and the farther along a slow-roll trajectory we progress, the larger σ becomes. In particular, for a slow-roll trajectory that evolves from $\mathbf{x}_1$ to $\mathbf{x}_2$ in effective phase space, we have

$$\frac{\sigma(\mathbf{x}_2)}{\sigma(\mathbf{x}_1)} \simeq \exp\left[3\int_{\rm C} \frac{{\rm d}\ell}{\sqrt{2\epsilon(\mathbf{x})}}\right], \tag{14.3.14}$$

where C is the segment of the slow-roll curve in the plane between $\mathbf{x}_1$ and $\mathbf{x}_2$ and ${\rm d}\ell$ is the line element along this curve, defined with respect to the fiducial metric. Hence, we generically expect that any region in effective phase space satisfying our slow-roll conditions (14.3.9) will have large measure density σ on effective phase space, relative to nearby regions. Of course, the measure density on effective phase space can be large in regions that fail the slow-roll conditions, such as during reheating, in which apparent attractor solutions traverse long paths in compact regions of effective phase space and trajectories appear to converge. All of these statements are made with regard to the phase space measure, not the measure on the space of trajectories; the factor of $|\dot{H}|$ in Eq. (14.2.13) makes this an important distinction.

14.3.2 Inflation on a Hilltop Potential

In a model of inflation governed by a potential with a hilltop (a local maximum), there are two types of classical solutions that differ qualitatively from the usual picture of the inflaton field rolling down the potential and reheating: 1) fixed point trajectories, i.e., exactly de Sitter solutions, which start with $\dot{\phi} = 0$ at the top of the potential and inflate forever; and 2) roll-up trajectories, which start from somewhere on the slope of the potential and asymptotically

approach the fixed point. The *fixed point* is the location $(\phi, \dot{\phi} = 0)$, equivalently $(x, y = 0) \equiv \mathbf{x}_0$ in effective phase space, for which ϕ is at the hilltop of $V(\phi)$. We would like to elucidate the behavior of the phase space measure density $\sigma(x, y)$ in the region of effective phase space near the fixed point.

Near the fixed point, $\dot{\phi}^2 \ll V(\phi)$, so Eq. (14.3.8) implies $|\partial_\mu \tilde{v}^\mu| \to (3\tilde{V}(x))^{1/2}$, as in Eq. (14.3.12); moreover, $\tilde{\mathbf{v}} \to 0$. With Eq. (14.2.9) requiring conservation of the measure under Hamiltonian evolution, the Cauchy-Schwarz inequality implies:

$$|\partial_\mu \tilde{v}^\mu| \sigma \leq |\boldsymbol{\partial} \sigma| |\tilde{\mathbf{v}}|, \tag{14.3.15}$$

where we use the vector notation $\boldsymbol{\partial}$ for ∂_μ. Since $|\partial_\mu \tilde{v}^\mu|$ is finite and $|\tilde{\mathbf{v}}| \to 0$, requiring that σ be smooth implies

$$\sigma(\mathbf{x}) \to 0 \ \text{ as } \ \mathbf{x} \to \mathbf{x}_0. \tag{14.3.16}$$

Even if we relaxed the assumption of regularity, we can still show that σ is small near the fixed point. We observe, given a smooth, slowly-varying potential $V(\phi)$, that any fixed point in effective phase space will be at the terminus of an apparent attractor, that is, a region where both slow-roll conditions (14.3.9) are met. As trajectories flow from near the fixed point along the apparent attractor, the first condition is always met, while the second becomes an increasingly good approximation. Hence, our conclusion from the slow-roll regime becomes applicable and so Eq. (14.3.14) implies that the effective phase space measure near the fixed point is exponentially suppressed compared with the measure further down the slow-roll apparent attractor. Other than the fixed point trajectory itself—which is irrelevant to inflation, since the field does not evolve—there is, relative to the slow-roll regime, very little measure near the hilltop. Recall that slicing effective phase space into sets of constant H to parameterize the space of trajectories incurs an additional factor of $|\dot{H}|$ to

convert the phase space measure density into the probability distribution on the trajectories; this suppresses the measure assigned to the roll-up trajectories even more. However, we have shown here that roll-up trajectories are suppressed in the canonical measure on effective phase space, even without the help of this additional factor. Since the measure is conserved under Hamiltonian evolution, any roll-up trajectory, i.e., the FRW evolution that comes arbitrarily close to de Sitter, is a set of measure zero.

14.4 Quadratic Inflation

14.4.1 Preliminaries

As a representative example of slow-roll inflation with $V''(\phi) > 0$, we consider monomial inflation with a quadratic potential,

$$V(\phi) = \frac{1}{2}m^2\phi^2. \tag{14.4.1}$$

If the recent BICEP2 discovery of B-mode polarization [347] is the result of primordial gravitational waves, then this simple model is in good agreement with the observed tensor perturbations. A canonical model in the inflationary literature [352, 353] and one of theoretical interest [354], the set of quadratic and related potentials is an important area of current investigation [355, 356], given the status of observations [65, 347].

It will eventually be useful to redefine our dimensionless coordinates $\mathbf{x}$ differently from those in Eq. (14.3.1):

$$x = \frac{\phi}{\sqrt{6}M_{\rm Pl}} \quad \text{and} \quad y = \frac{\dot{\phi}}{\sqrt{6}mM_{\rm Pl}}. \tag{14.4.2}$$

We define polar coordinates (z, θ):

$$z \equiv \sqrt{x^2 + y^2} = \frac{H}{m} \tag{14.4.3}$$

and

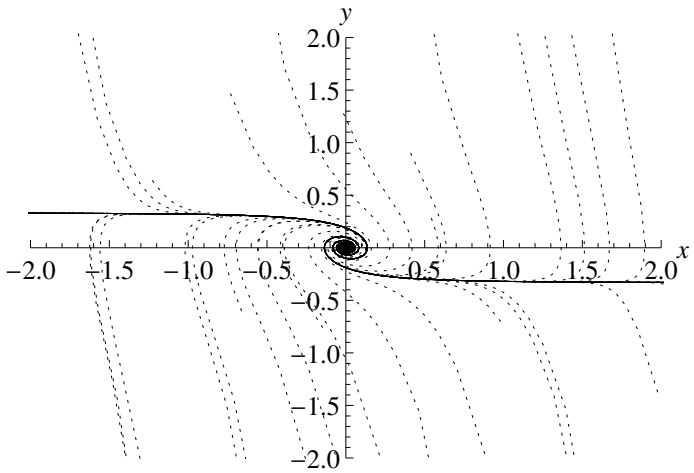

Figure 14.1. Trajectories in effective phase space for quadratic inflation. The field value and velocity are parameterized by the variables x and y, defined in Eq. (14.4.2). The dark nearly-horizontal lines indicate the apparent attractors, where the conserved measure grows large. For clarity we used the (unrealistic) value of $m = 0.2M_{\rm Pl}$ to make this plot.

$$\tan\theta = \frac{y}{x} = \frac{\dot\phi}{m\phi}, \tag{14.4.4}$$

so $\dot\phi = \sqrt{6}mM_{\rm Pl}z\sin\theta$ and $\phi = \sqrt{6}M_{\rm Pl}z\cos\theta$.

Using Eq. (14.2.4), we can plot trajectories in the ϕ-$\dot\phi$ plane and see explicitly the effective phase space behavior, as shown in Fig. 14.1. In particular, note the apparent attractor solutions that appear at $y = \pm 1/3$, corresponding to $\dot\phi = \pm\sqrt{2/3}mM_{\rm Pl}$. Of course, in a strict phase-space sense, these "attractors" are illusory [12]. In the Liouville measure, phase-space density is conserved. The apparent attractor behavior actually indicates that the measure density grows large in that region.

The apparent attractor solution at $\dot\phi = \pm\sqrt{2/3}mM_{\rm Pl}$ intersects the $H =$ constant ellipse at

$$|\sin\theta| = \frac{m}{3H}. \tag{14.4.5}$$

In the early universe ($H \gg m$), we therefore have $\theta \simeq 0$ or π on the apparent attractor.

14.4.2 Counting e-Folds

For the quadratic potential (14.4.1), one has the potential slow-roll parameter from Eq. (14.3.5):

$$\epsilon_V = 2\left(\frac{M_{\rm Pl}}{\phi}\right)^2. \tag{14.4.6}$$

Inflation (and counting of e-folds N) ends when $\epsilon_V = 1$, which occurs at $\phi_{\rm f} = \sqrt{2}M_{\rm Pl}$.

For slow roll, $H^2 \simeq V/3M_{\rm Pl}^2$, the scalar equation (14.2.4) becomes $3H\dot{\phi} \simeq -V'$ and so $H{\rm d}t \simeq \pm{\rm d}\phi/\sqrt{2\epsilon_V}M_{\rm Pl}$. Thus, when the field value is ϕ, the number of e-folds remaining before the end of inflation is

$$N\left(\phi\right) = \left|\int_{|\phi|}^{\sqrt{2}M_{\rm Pl}} \frac{1}{\sqrt{2\epsilon_V}}\frac{{\rm d}\phi'}{M_{\rm Pl}}\right| = \frac{1}{4}\left(\frac{\phi}{M_{\rm Pl}}\right)^2 - \frac{1}{2}, \tag{14.4.7}$$

which is accurate as long as the slow-roll conditions (14.3.9) are satisfied. While exact number of e-folds, defined in Eq. (14.1.1) using the full expression for $H{\rm d}t$ given in Eq. (14.2.5), would have corrections near the end of inflation where the slow-roll conditions begin to break down, we shall see that this will not appreciably affect the total number of e-folds that we ultimately compute. Consider a trajectory that starts at angle θ on the surface where $H = M_{\rm Pl}$. We will call this the Planck surface; of course, one could choose a different ultraviolet cutoff $\Lambda_{\rm UV} \gg m$ for the effective field theory, on which to start evaluating trajectories at time $t = 0$. In that case, one could simply replace $M_{\rm Pl}$ by $\Lambda_{\rm UV}$ as appropriate in all of our e-fold counting. For simplicity we will choose $\Lambda_{\rm UV} = M_{\rm Pl}$. The initial field value for ϕ is then $\sqrt{6}(\cos\theta)M_{\rm Pl}^2/m$.

In the $H \gg m$ region of ϕ-$\dot{\phi}$ space, trajectories snap quickly to the apparent attractor, with $\dot{\phi}$ changing much faster than $m\phi$. That is, using the scalar equation, we have

$$\frac{\dot{\mathbf{x}}}{m} = \left(y, -x - 3y\sqrt{x^2+y^2}\right). \tag{14.4.8}$$

Thus, when $z = \sqrt{x^2 + y^2} \gg 1$, we have $\dot{y} \gg \dot{x}$, as claimed. Hence, $x(t = 0)$, to a very good approximation, is equal to x at the time when the trajectory starts the slow-roll process. Therefore, we can write the *total* number of e-folds that this trajectory (parameterized by θ on the Planck surface) undergoes as

$$N_{\rm tot} = \frac{3}{2}\left(\frac{M_{\rm Pl}}{m}\right)^2 \cos^2\theta - \frac{1}{2} \simeq \frac{3}{2}\left(\frac{M_{\rm Pl}}{m}\right)^2 \cos^2\theta. \tag{14.4.9}$$

Maximal inflation occurs when $\theta \simeq 0$ or π, i.e., the trajectory starts out near the apparent attractor at the Planck scale, which gives

$$N_{\rm max} = \frac{3}{2}\left(\frac{M_{\rm Pl}}{m}\right)^2 \tag{14.4.10}$$

e-folds of inflation. Comparing the analytical prediction (14.4.9) with numerical simulation, we find very good agreement.

14.4.3 How Many e-Folds Should We Expect in Quadratic Inflation?

We know from Eq. (14.4.9) how to predict the total number of e-folds of inflation a trajectory will undergo based on a particular parameterization of the family of trajectories, namely, by the angular coordinate θ with which the trajectory intersects a surface of particular energy density, in this case $H = M_{\rm Pl}$. We would now like to ask the question of how many e-folds we should expect, using the prescription for finding the appropriate measure (14.2.13) on the space of trajectories, as described in Sec. 14.2.2.

First, we need to find the measure $\boldsymbol{\sigma}$ on effective phase space for quadratic inflation. In the (z, θ) coordinates defined in Eqs. (14.4.3) and (14.4.4), we can write the velocity (14.2.6) of trajectories in effective phase space as

$$\mathbf{v} = \dot{\mathbf{x}} = -3mz^2 \sin^2\theta \hat{\mathbf{z}} - m\left(z + 3z^2 \sin\theta\cos\theta\right)\hat{\boldsymbol{\theta}}, \tag{14.4.11}$$

where

$$\begin{pmatrix} \hat{\mathbf{z}} \\ \hat{\boldsymbol{\theta}} \end{pmatrix} = \begin{pmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{pmatrix} \begin{pmatrix} \hat{\mathbf{x}} \\ \hat{\mathbf{y}} \end{pmatrix}. \tag{14.4.12}$$

In the early universe, when $H \gg m$ (such as on the Planck ellipse $H = M_{\rm Pl}$), we have $z \gg 1$. Thus, $\mathbf{v}$ becomes approximately

$$\mathbf{v} \simeq -3mz^2 \sin^2\theta \hat{\mathbf{z}} - 3mz^2 \sin\theta\cos\theta\hat{\boldsymbol{\theta}}, \tag{14.4.13}$$

and so the requirement (14.2.9) for $\boldsymbol{\sigma}$ to be conserved under Hamiltonian evolution becomes

$$\partial_\theta\sigma = -z\tan\theta\partial_z\sigma - (2\tan\theta + \cot\theta)\,\sigma. \tag{14.4.14}$$

The general solutions for σ take the form [12]

$$\sigma = \sum_\gamma C_\gamma z^{\gamma-3}\left|\frac{\cos^{\gamma-1}\theta}{\sin\theta}\right|, \tag{14.4.15}$$

for $z \gg 1$, where γ, $C_\gamma \in \mathbb{R}$. We note that σ diverges along the $\sin\theta = 0$ axis, corresponding to the buildup of trajectories along the apparent attractor; in an exact numerical solution, the distribution σ would become large on the apparent attractor solution, as is clear from Fig. 14.1, using the fluid flow analogy. For the potential (14.4.1), we proved in Ref. [12] that the measure $\boldsymbol{\sigma}$ has a unique solution; hence, many possible solutions in Eq. (14.4.15) are spurious or unphysical.

As we can see from flow of the vector field shown in Fig. 14.1, we should require that σ be finite everywhere except on the apparent attractor solution; imposing this condition requires $\gamma \geq 1$. Further, at fixed θ, trajectories become more squeezed together as z decreases, since more and more time evolution is compressed into a smaller and smaller range of H. Hence, we should require σ to be a non-increasing function of z at fixed θ, so $\gamma \leq 3$. Imposing the further requirement that σ be infinitely differentiable everywhere except the apparent attractor solution selects $\gamma = 3$ as the physical solution, so we end up with

$$\sigma(H = M_{\rm Pl}, \theta) \propto \left|\frac{\cos^2\theta}{\sin\theta}\right|. \tag{14.4.16}$$

In Sec. 14.2.2 we demonstrated how to obtain the measure on the space of

trajectories from $\boldsymbol{\sigma}$, *cf.* Ref. [64]. Specifically, the probability distribution on the space of trajectories, parameterized by θ on some surface of constant H, is, up to normalization, given by $\sigma(H,\theta)\,|\dot{H}|$. Using Eq. (14.2.4), we have in the (z,θ) coordinates:

$$\dot{H} = -\frac{\dot{\phi}^2}{2M_{\rm Pl}^2} = -3m^2 z^2 \sin^2\theta. \tag{14.4.17}$$

Thus, the probability distribution on the space of trajectories on the Planck surface (where $z = M_{\rm Pl}/m = \text{constant}$), parameterized by the coordinate θ, is

$$P(\theta)|_{H=M_{\rm Pl}} = \frac{3}{4}\left|\cos^2\theta\sin\theta\right|. \tag{14.4.18}$$

The overall normalization has been fixed by requiring $\int \mathrm{d}\theta\, P(\theta) = 1$.

Finally, we can now compute the expected total number of e-folds of inflation, using the canonical measure (14.4.18) and our e-fold counting (14.4.9):

$$\begin{aligned}\langle N_{\rm tot}\rangle &= \int_0^{2\pi} N(\theta)\; P(\theta)|_{H=M_{\rm Pl}}\,\mathrm{d}\theta \\ &= \frac{9}{8}\left(\frac{M_{\rm Pl}}{m}\right)^2 \int_0^{2\pi} \cos^4\theta\,|\sin\theta|\,\mathrm{d}\theta \\ &= \frac{9}{10}\left(\frac{M_{\rm Pl}}{m}\right)^2 \\ &= \frac{3}{5}N_{\rm max}.\end{aligned} \tag{14.4.19}$$

Now, assuming a quadratic potential (14.4.1), the amplitude of observed CMB scalar perturbations is

$$\Delta_{\rm s}^2(k_{\rm CMB}) = \frac{1}{6\pi^2}\left(\frac{m}{M_{\rm Pl}}\right)^2 N_{\rm CMB}^2, \tag{14.4.20}$$

where $N_{\rm CMB} \approx 50$ is the number of e-folds between horizon exit of CMB scales and the end of inflation. Using the Planck observations [65] for the amplitude of scalar perturbations, we have $m = 7\times10^{-6}M_{\rm Pl} = 2\times10^{13}\,\text{GeV}$, which implies that for quadratic inflation we expect

$$\langle N_{\rm tot}\rangle = 2\times10^{10}. \tag{14.4.21}$$

That is, typical universes under the canonical measure (14.4.18) with the

inflaton mass we obtain by positing a quadratic potential (14.4.1) and requiring consistency with Planck [65] undergo much more than the required number of e-folds needed to solve the horizon problem; hence, one can view our observed universe as natural in the theory of quadratic inflation, with regard to the canonical measure (with the caveats about inhomogeneities noted in the Introduction).

Looking at the conclusion another way, we note that the probability for $N_{\rm tot}$ to be greater than some particular value N_0 is just the probability that

$$\cos^2\theta > (2/3)(m/M_{\rm Pl})^2 N_0 \equiv \cos^2\zeta. \tag{14.4.22}$$

Thus,

$$\begin{aligned}\Pr\left(N_{\rm tot} > N_0\right) &= 4\times\frac{3}{4}\int_0^{\zeta}\cos^2\theta\sin\theta{\rm d}\theta \\ &= 1-\left(\frac{2}{3}\right)^{3/2}\left(\frac{m}{M_{\rm Pl}}\right)^3 N_0^{3/2} \\ &= 1-\left(\frac{N_0}{N_{\rm max}}\right)^{3/2},\end{aligned} \tag{14.4.23}$$

where $N_{\rm max}$ is defined in Eq. (14.4.10). That is, if $m = 2\times10^{13}\,{\rm GeV}$, the probability of having fewer than 50 e-folds of inflation is of order 10^{-13}. Differentiating Eq. (14.4.23), the measure on the space of trajectories can be written in terms of the number $N_{\rm tot}$ of e-folds ultimately achieved, between zero and $N_{\rm max}$:

$$P\left(N_{\rm tot}\right){\rm d}N_{\rm tot} = \frac{3}{2N_{\rm max}^{3/2}}N_{\rm tot}^{1/2}{\rm d}N_{\rm tot}. \tag{14.4.24}$$

Universes that undergo 50 or more e-folds of inflation, like our own, are overwhelmingly generic from the perspective of the canonical measure for high-scale quadratic inflation.

The specific number $\langle N_{\rm tot}\rangle = 2\times10^{10}$ is suggestive, but it shouldn't be taken too literally. In quadratic inflation, the field has a value $\phi\sim10M_{\rm Pl}$ at the epoch when currently observable large-scale perturbations are being generated;

our calculation fearlessly extrapolates the functional form of the potential to values of order $10^5 M_{\rm Pl}$, where there is little reason for it to be trusted. Nevertheless, we expect that our result has a robust physical interpretation for more general potentials: in large-field inflation, when the potential increases to the Planck limit, it is natural to achieve a large amount of inflation. There are certainly some trajectories that spend little or no time on the apparent attractor, remaining dominated by kinetic energy all the way up to Planck densities. Our results suggest that, in large-field inflation, such trajectories are extremely unlikely, as generic evolution quickly snaps to the apparent attractor, yielding many e-folds of inflation.

14.5 Cosine (Natural) Inflation

14.5.1 Preliminaries

We now turn to the model of cosine or "natural" inflation [357, 358], in which the inflaton ϕ could be a pseudo-Nambu-Goldstone boson $\theta = \phi/f$ with a global shift symmetry broken at scale f. The global symmetry is explicitly broken at scale Λ, giving the boson a mass, via a potential

$$V(\phi) = \Lambda^4 \left[1 - \cos\left(\phi/f\right)\right]. \tag{14.5.1}$$

Cosine inflation is representative of the general class of hilltop inflation models: the inflaton potential has a region where $V''(\phi) < 0$. Qualitatively, this model has similarities and differences with monomial inflation models. Like monomial inflation, it can exhibit slow-roll behavior. Unlike monomial models, however, hilltop models have trajectories in which the inflaton stays near the top of the potential for a parametrically long time and pure de Sitter space is allowed if the field sits exactly at the potential maximum. Such trajectories would seem to allow hilltop models to achieve a very large number of e-foldings without the

concomitant large excursion in field values endemic to monomial models and potentially troublesome from the effective field theory perspective. Ultraviolet completions of cosine inflation models have been investigated [359–361], which improve the applicability of effective field theoretic reasoning. Cosine inflation models are of significant current interest [350] and generically have regions of parameter space that can achieve agreement with observations from BICEP2 [347] and Planck [65]. In cosine inflation, the field can without loss of generality be restricted to the interval between $\pm\pi f$, with the periodic identification $\phi \sim \phi + 2\pi f$ as an equivalence class.

As for quadratic inflation, we will find dimensionless coordinates useful (different from Eqs. (14.3.1) and (14.4.2)):

$$x = \sqrt{\frac{2}{3}}\frac{f}{M_{\rm Pl}}\sin\left(\phi/2f\right) \quad \text{and} \quad y = \frac{f\dot{\phi}}{\sqrt{6}\Lambda^2 M_{\rm Pl}}. \tag{14.5.2}$$

Because of the restricted range of the field, x is isomorphic to ϕ, so our discussion about vector field invariance from Ref. [12] applies and (x, y) forms an effective phase space. Note that $x \in [-\sqrt{2/3}f/M_{\rm Pl}, \sqrt{2/3}f/M_{\rm Pl}]$, with the identification $x \sim x + 2\sqrt{2/3}f/M_{\rm Pl}$. As before, define polar coordinates

$$z \equiv \sqrt{x^2 + y^2} = \frac{f}{\Lambda^2}H \tag{14.5.3}$$

and

$$\tan\theta = \frac{y}{x} = \frac{\dot{\phi}}{2\Lambda^2 \sin\left(\phi/2f\right)}. \tag{14.5.4}$$

Because x can only take values between $\pm\sqrt{2/3}f/M_{\rm Pl}$, the Planck surface $H = M_{\rm Pl}$ subtends a finite set of angles $[\theta_0, \pi - \theta_0] \cup [\pi + \theta_0, 2\pi - \theta_0]$, where

$$\cos\theta_0 = \frac{\sqrt{\frac{2}{3}}\frac{f}{M_{\rm Pl}}}{\frac{f}{\Lambda^2}M_{\rm Pl}} = \sqrt{\frac{2}{3}}\frac{\Lambda^2}{M_{\rm Pl}^2} \ll 1, \tag{14.5.5}$$

i.e., θ_0 is close to $\pi/2$ or $3\pi/2$.

In (x, y) coordinates, the velocity vector $\mathbf{v} = \dot{\mathbf{x}}$, using Eq. (14.2.4), is

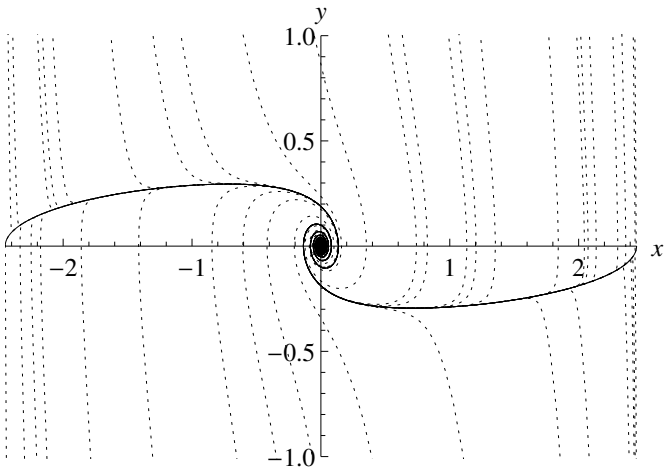

Figure 14.2. Trajectories in effective phase space for cosine inflation. The field value and velocity are parameterized by the variables x and y, defined in Eq. (14.5.2). The dark spirals indicate the apparent attractors, where the conserved measure grows large. For this plot we used $f = 3M_{\rm Pl}$, $\Lambda = 0.1M_{\rm Pl}$.

$$\frac{f\mathbf{v}}{\Lambda^2} = y\sqrt{1 - \frac{3}{2}\frac{M_{\rm Pl}^2}{f^2}x^2}\;\hat{\mathbf{x}} - \left(3y\sqrt{x^2+y^2} + x\sqrt{1 - \frac{3}{2}\frac{M_{\rm Pl}^2}{f^2}x^2}\right)\hat{\mathbf{y}}, \qquad (14.5.6)$$

or equivalently, in polar coordinates (z, θ),

$$\frac{f\mathbf{v}}{\Lambda^2} = -3z^2\sin^2\theta\;\hat{\mathbf{z}} - \left(3z^2\sin\theta\cos\theta + z\sqrt{1 - \frac{3}{2}\frac{M_{\rm Pl}^2}{f^2}z^2\cos^2\theta}\right)\hat{\boldsymbol{\theta}}, \qquad (14.5.7)$$

where $\hat{\mathbf{x}}$, $\hat{\mathbf{y}}$, $\hat{\mathbf{z}}$, and $\hat{\boldsymbol{\theta}}$ are related as in Eq. (14.4.12). Plotting integral curves of this vector field, one can visualize trajectories in effective phase space, shown in Fig. 14.2. As for quadratic inflation, there is an apparent attractor, but for ϕ near $\pm f$, lingering behavior near the hilltop is also possible.

14.5.2 Counting e-Folds

With the potential slow-roll parameter ϵ_V defined as in Eq. (14.3.5), for the cosine inflation potential (14.5.1) one has

$$\epsilon_V = \frac{M_{\rm Pl}^2}{2f^2}\frac{\sin^2(\phi/f)}{[1-\cos(\phi/f)]^2} = \frac{1-b^2x^2}{3x^2}, \qquad (14.5.8)$$

for convenience defining a constant

$$b \equiv \sqrt{3/2} M_{\rm Pl}/f. \tag{14.5.9}$$

Inflation—and counting of e-folds—ends when $\epsilon_V = 1$, which occurs at $|x| = (3+b^2)^{-1/2}$.

For slow roll and assuming $\ddot{\phi}$ is small compared to other terms in the scalar equation (14.2.4), we have $H^2 \simeq V/3M_{\rm Pl}^2$ and $3H\dot{\phi} \simeq -V'$, so

$$H{\rm d}t \simeq \pm \frac{{\rm d}\phi}{\sqrt{2\epsilon_V} M_{\rm Pl}} = \pm \frac{3|x|{\rm d}x}{1-b^2x^2}, \tag{14.5.10}$$

after using Eqs. (14.5.2) and (14.5.8).[14.3] Thus, when the field value is x, the number of e-folds remaining before the end of inflation is

$$\begin{aligned} N(x) &= \left| \int_{|x|}^{\left(3+b^2\right)^{-1/2}} \frac{3x'{\rm d}x'}{1-b^2(x')^2} \right| \\ &= \frac{3}{2b^2} \log \left[\frac{1}{(1-b^2x^2)\left(1+\frac{1}{3}b^2\right)} \right]. \end{aligned} \tag{14.5.11}$$

We would like to parameterize the number of e-folds a trajectory undergoes based upon its coordinate θ on the Planck surface, *not* its coordinate x when it enters the slow-roll regime. From the vector field in Eqs. (14.5.6) and (14.5.7), we see that when $z = \sqrt{x^2+y^2} \gg 1$ and $y \gg x$ (which is true on the Planck surface) we have $\dot{y} \gg \dot{x}$. Therefore, as for quadratic inflation, we are able to approximate $x\,(\text{Planck surface}) \simeq x\,(\text{enter slow roll})$ for a given trajectory.[14.4] The total number of e-folds attained by a trajectory that starts out at angle θ on the Planck surface is then

[14.3] Though in general a hilltop trajectory can violate the condition that $3H\dot{\phi} \simeq -V'$, one can show that, for the potential (14.5.1), the total number of e-folds we compute is accurate even without this assumption. See footnote 14.5.

[14.4] Note that this approximation leads us to assign nonzero measure to the set of trajectories that come arbitrarily close to the fixed point. It therefore assigns nonzero measure where the roll-up trajectory intersects the Planck surface, which we argued in Sec. 14.3 is not strictly correct. However, if anything, this assumption should overestimate the expected total number of e-folds.

$$N_{\rm tot}\left(\theta\right) = \frac{3}{2b^2}\log\left[\frac{1}{\left(1-\frac{\cos^2\theta}{\cos^2\theta_0}\right)\left(1+\frac{1}{3}b^2\right)}\right], \tag{14.5.12}$$

where θ_0 is defined in Eq. (14.5.5). Note that when x approaches $1/b$, i.e., when θ approaches θ_0, $N_{\rm tot}$ diverges, as we would expect. In our approximation that $x(\text{Planck surface}) \simeq x(\text{enter slow roll})$, $x(\text{Planck surface}) = 1/b$ is identified as the roll-up trajectory discussed in Sec. 14.3.2.

14.5.3 How Many e-Folds Should We Expect in Cosine Inflation?

From Eq. (14.5.12), we know, given a trajectory that intersects the Planck surface with angular coordinate θ, how many e-folds that trajectory will ultimately undergo. As in Sec. 14.4.3, we now turn to the question of how many e-folds we should expect under the canonical measure (14.2.13) on the space of trajectories. As shown in Sec. 14.2.2, we must first find the Hamiltonian-conserved measure—a measure whose density satisfies the condition (14.2.9)—on effective phase space.

In our z coordinates (14.5.2), the $H = M_{\rm Pl}$ surface corresponds to

$$z = \frac{M_{\rm Pl}f}{\Lambda^2} \gg 1. \tag{14.5.13}$$

As we have previously noted, had we used a different ultraviolet cutoff $\Lambda_{\rm UV}$, other than $M_{\rm Pl}$, all of the results that follow would be the same, with $M_{\rm Pl}$ replaced by $\Lambda_{\rm UV}$, so our conclusions would not qualitatively change. Taking the large-z limit of Eq. (14.5.7), we have

$$\mathbf{v} \simeq -3\frac{\Lambda^2}{f}z^2\sin^2\theta\hat{\mathbf{z}} - 3\frac{\Lambda^2}{f}z^2\sin\theta\cos\theta\hat{\boldsymbol{\theta}}, \tag{14.5.14}$$

which is identical to Eq. (14.4.13) up to a multiplicative factor. That is, we have turned the large-H behavior of cosine inflation into the large-H behavior of quadratic inflation, through the judicious choice of coordinates (14.5.2). The effective phase space measure density therefore takes the same general form

(14.4.15) and on physical grounds we can restrict to the $\gamma = 3$ case for the reasons discussed in Sec. 14.4.3, so that the measure density becomes as in Eq. (14.4.16).

There is nevertheless an important difference between the quadratic and cosine inflation scenarios, since ϕ is restricted to a small window in the latter. This implies that the normalization of Eq. (14.4.16) will be different. As we have noted, this restriction translates into a restriction of values of θ on the Planck surface to a small range near $\pi/2$ and near $3\pi/2$, with width given in Eq. (14.5.5). Hence, σ does not diverge on the $H = M_{\rm Pl}$ surface for cosine inflation. In the z coordinates, we have as in Eq. (14.4.17)

$$\dot{H} = -3\frac{\Lambda^4}{f^2} z^2 \sin^2\theta. \tag{14.5.15}$$

The probability distribution over the space of trajectories, parameterized by the angle θ on the $H = M_{\rm Pl}$ surface, is therefore

$$P\left(\theta\right)\big|_{H=M_{\rm Pl}} = \frac{3}{4\cos^3\theta_0}\left|\cos^2\theta\sin\theta\right|, \tag{14.5.16}$$

with θ_0 defined in Eq. (14.5.5) as before. The normalization once again comes from demanding that the total probability equal unity.

Having found the canonical measure (14.5.16) on the space of trajectories, we can now use our e-fold counting (14.5.12) to compute the expectation value for the total number of e-folds attained by a FRW universe in the cosine inflation model:

$$\begin{aligned}
\langle N_{\rm tot}\rangle &= 2\int_{\theta_0}^{\pi-\theta_0} N_{\rm tot}\left(\theta\right) P\left(\theta\right)\big|_{H=M_{\rm Pl}}\,{\rm d}\theta \\
&= \frac{9}{2b^2}\int_0^1 \log\left[\frac{1}{(1-u^2)\left(1+\frac{1}{3}b^2\right)}\right] u^2{\rm d}u \\
&= \frac{f^2}{3M_{\rm Pl}^2}\left[8-6\log 2-3\log\left(1+\frac{M_{\rm Pl}^2}{2f^2}\right)\right] \\
&\simeq \left(\frac{8}{3}-2\log 2\right)\frac{f^2}{M_{\rm Pl}^2},
\end{aligned} \tag{14.5.17}$$

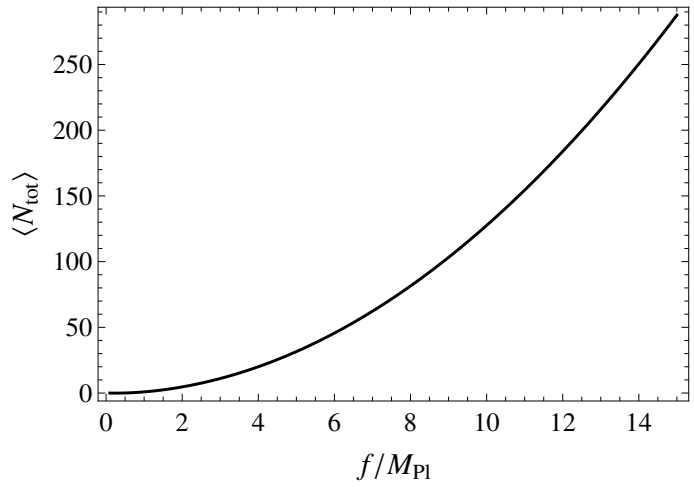

Figure 14.3. Expected number of e-folds $\langle N_{\rm tot}\rangle$, as computed in Eq. (14.5.17) using the canonical measure on the space of trajectories, for cosine inflation with potential $V(\phi) = \Lambda^4[1 - \cos(\phi/f)]$.

which is plotted in Fig. 14.3; the constant b was defined in Eq. (14.5.9).[14.5] For example, setting $f = 1.2 \times 10^{19}\,\text{GeV}$ (the unreduced Planck mass) gives

$$\langle N_{\rm tot}\rangle = 32. \tag{14.5.18}$$

This implies an insufficient amount of inflation to address the horizon problem, but clearly $\langle N_{\rm tot}\rangle$ can be increased by a small boost in f.

Interestingly, $\langle N_{\rm tot}\rangle$ is independent of Λ, only depending on f. This can be understood as follows: for small $\phi \ll f$, cosine inflation is equivalent to quadratic inflation, with mass Λ^2/f taking the place of m. Then the expected number of e-folds should be of order $N_{\rm max}$ (14.4.10), multiplied by a factor of $\cos^2\theta_0$, as given in Eq. (14.5.5), to account for the limited allowed range of ϕ, *cf.* Eq. (14.4.9); this reasoning would lead one to expect $\langle N_{\rm tot}\rangle \sim f^2/M_{\rm Pl}^2$, which is indeed what we find. We find that $f > 6.3\,M_{\rm Pl} = 1.5 \times 10^{19}$ GeV is needed in order to have $\langle N_{\rm total}\rangle > 50$.

[14.5] An exact expression for $H\mathrm{d}t$ (14.5.10) would have ε in place of ϵ_V. Taking into account relaxation of the slow-roll conditions near the hilltop, one can show that, if $f \ll M_{\rm Pl}$, then the total e-fold count we estimate should be increased by a factor of at most $M_{\rm Pl}/\sqrt{6}f$. However, this would still lead to less than one e-fold of inflation expected under the canonical measure in the $f \lesssim M_{\rm Pl}$ case. Moreover, one can show that, even near the hilltop, the approximation $\epsilon_V \simeq \varepsilon$ is very accurate in the $f \gtrsim M_{\rm Pl}$ case.

Recently, in light of results from Refs. [65, 347], much attention has been devoted to cosine inflation. By varying $f/M_{\rm Pl}$, a one-parameter family of predictions is obtained that is able to achieve agreement with either the Planck or BICEP2 results [350]. In particular, $f \sim 5 - 10 \times M_{\rm Pl}$ was found to be in better agreement[14.6] with the Planck observations [65], while larger f (which brings the predictions closer to those of quadratic inflation) is in better agreement with BICEP2.

What we have found is that smaller values of f are, in the sense of the canonical measure, highly unlikely to give a universe consistent with the observed uniformity of the CMB. In particular, if $f \leq M_{\rm Pl} = 2.4 \times 10^{18}\,\text{GeV}$, we have less than one e-fold of inflation. More quantitatively, we can compute the probability of attaining a given number N_0 of e-folds as a function of $f/M_{\rm Pl}$. From our expression (14.5.12) for $N_{\rm tot}$ as a function of θ on the Planck surface, we find that this is just the probability that

$$\cos^2\theta > \cos^2\theta_0\left[1 - \frac{3}{3+b^2}\exp\left(-\frac{2b^2}{3}N_0\right)\right] \equiv \cos^2\delta. \tag{14.5.19}$$

That is, evaluating the integral $4 \times (3\sec^3\theta_0/4)\int_{\theta_0}^{\delta}\cos^2\theta\sin\theta\mathrm{d}\theta = \Pr(N_{\rm tot} > N_0)$, we find

$$\Pr\left(N_{\rm tot} > N_0\right) = 1 - \left[1 - \left(1 + \frac{M_{\rm Pl}^2}{2f^2}\right)^{-1}\exp\left(-\frac{M_{\rm Pl}^2}{f^2}N_0\right)\right]^{3/2}. \tag{14.5.20}$$

The result is plotted in Fig. 14.4 for $N_0 = 50$. We find that, if $f \leq 2M_{\rm Pl} = 4.9 \times 10^{18}\,\text{GeV}$, the probability under the canonical measure of attaining 50 or more e-folds of inflation is less than 10^{-5}. While the details of Eq. (14.5.20) break down if $f \lesssim M_{\rm Pl}$ due to corrections to the slow-roll approximation near the hilltop, the expected total number of e-folds (14.5.17) remains valid and the probability of attaining more than 50 e-folds for $f \lesssim M_{\rm Pl}$ remains infinitesimal.

[14.6] Note that this range of f could also be written as $f \sim 1 - 2 \times m_{\rm Pl}$, where $m_{\rm Pl} = 1/\sqrt{G} = 1.2 \times 10^{19}\,\text{GeV}$ is the unreduced Planck mass.

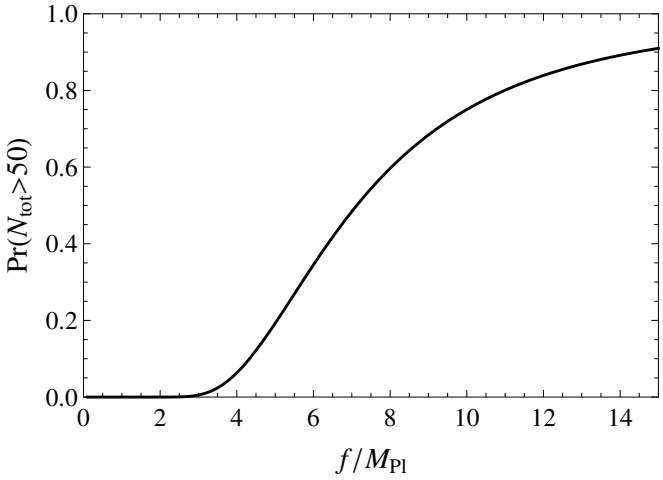

Figure 14.4. The probability of obtaining 50 or more e-folds of inflation as a function of f for cosine inflation with potential $V(\phi) = \Lambda^4[1 - \cos(\phi/f)]$, as computed using the canonical measure on the space of trajectories and starting on the $H = M_{\rm Pl}$ surface.

That is, for $f \lesssim M_{\rm Pl}$, the overwhelming majority of universes will under-inflate. On the other hand, if $f = 100 M_{\rm Pl} = 2.4 \times 10^{20}\,{\rm GeV}$, we find that the probability of a universe attaining at least 50 e-folds of inflation is approximately 0.99964. Hence, the probability of a FRW universe undergoing sufficient inflation to explain the observed uniformity of the CMB is sensitively dependent on $f/M_{\rm Pl}$ in cosine inflation, with larger values of $f \gtrsim \mathcal{O}\,({\rm few}) \times M_{\rm Pl} \sim 10^{19}\,{\rm GeV}$ much preferred.

If f is too small in cosine inflation, then our universe is finely tuned from the perspective of the canonical measure. Hence, models of cosine inflation with f on the order of $10^{18}\,{\rm GeV}$ or less do not solve the cosmological fine-tuning problems that are the original purpose of inflationary theory. For cosine inflation to truly be natural in the cosmological sense, f must be above $10^{19}\,{\rm GeV}$. On the other hand, this result helps motivate the possibility that our universe did experience just the right amount of inflation but not too much, suggesting that there may be observable relics of the pre-inflationary universe that might be observable on very large angular scales. Interestingly,

in the large-f limit that is favored by the canonical measure, the observational predictions of cosine inflation merge with those of quadratic inflation.

As with the quadratic case, we expect our results for cosine inflation to be indicative of a more general lesson for hilltop (small-field) models. Unlike the large-field case, where the potential rises all the way to the Planck density, in cosine inflation the maximum is well below that scale. There are trajectories that linger for an arbitrarily large number of e-folds in the slow-roll regime near the top of the hill, but there are also trajectories that exhibit a kinetic-dominated phase of evolution prior to a finite period of slow roll. Our result shows that it is the latter category that are most likely, as quantified by the conserved measure on effective phase space.

14.6 Conclusions

The recent BICEP2 discovery, if verified, suggests that high-scale cosmic inflation is the correct theory of the very early universe. With characteristic energy of order $10^{16}\,\mathrm{GeV}$, observational signatures of inflation open the door to physics on the threshold of the Planck scale. Many models of inflation are currently being investigated for their ability to fit precision CMB observations. The current success of relatively simple models of inflation, driven by a single scalar field with a potential and a canonical kinetic term, is impressive. Given the large set of possible inflaton potentials, it is of vital importance to develop useful theoretical tools that enable observations to discriminate among competing models.

The theory of cosmic inflation was originally posited to solve problems of fine-tuning of initial conditions, such as the uniformity of the CMB temperature, lack of observed monopoles, and smallness of curvature. Given the current wealth of precise cosmological measurements, it is well motivated to apply the

same question of genericness to various proposed models of inflation. That is, given a particular model, does it generically produce the observed properties of our universe? In particular, does it typically produce the requisite number of e-folds $(40 - 60)$ to account for the uniformity of the CMB? Inherent in such questions is the idea of a measure: a probability distribution on the set of all possible FRW universes. Following GHS [64], in Ref. [12] we developed a formalism for constructing such a measure on the subset of flat universes (on which the GHS measure diverges).

In the present chapter, we investigated the behavior of the effective phase space measure for two general classes of potentials important for single-field inflation: slow roll down a potential and lingering behavior near a potential hilltop. In the former case, we showed that the effective phase space measure generically becomes large, while in the latter case it generically becomes small. That is, trajectories that linger arbitrarily long near quasi-de Sitter space at a potential hilltop are disfavored by the canonical measure, while trajectories that slow roll and eventually reheat are favored. We next quantitatively examined the statistical conclusions offered by the canonical measure for two representative inflaton potentials: quadratic inflation and cosine inflation. Interestingly, the statistical expectation for the amount of inflation experienced in these two cases differed dramatically. For quadratic inflation, we found that, given an inflaton mass consistent with the observed amplitude of scalar perturbations [65], nearly all trajectories undergo 50 e-folds of inflation. In fact, generic trajectories experience billions of e-folds. On the other hand, for cosine inflation with symmetry-breaking parameter f, typical trajectories under-inflate unless $f \gtrsim 10^{19}$ GeV. Above this scale, 50 e-folds are generically attainable and the observational cosmology predictions of cosine inflation merge with those of the quadratic potential.

From our demonstration with these two examples, we illustrated the utility of the canonical measure in elucidating differences in physical predictions among models of inflation. While a given potential may have *some* trajectory—some possible history of a FRW universe —that undergoes enough inflation to correspond to our observed universe, that does not mean that this trajectory is generic. Indeed, in some models, such as cosine potentials with $f \lesssim M_{\rm Pl}$, the vast majority of trajectories, as weighted by the canonical measure, do not undergo sufficient inflation, despite the existence of a small subset of finely-tuned trajectories that do. The canonical measure allows one to quantify the amount of tuning required in a given model to reproduce our universe.[14.7] The degree of tuning required on the space of trajectories to produce at least 50 e-folds of inflation (or whatever other observed quantity one is computing) should correspond inversely with the degree of credence given a particular model, modulo theoretical bias. That is, given two potentials, one generically attaining many e-folds and another in which only a small subset (as computed in the canonical measure) of trajectories attain 50 e-folds, the former model should be favored: one could say that such a model is more "natural," in the sense that it requires less fine-tuning to match observations. This approach is an interesting parallel to current discussions in particle physics regarding naturalness of the electroweak scale and the amount of tuning required in various models, such as supersymmetry.

The contrapositive of this line of thinking is also illuminating. If future cosmological observations point to a particular inflaton potential for which our universe is not generic under the canonical measure, that would shed light on even higher-scale physics. Such a circumstance would tell us that our universe is tuned—on a non-generic trajectory—from the point of view of the

[14.7] All of these statements are made under the assumption of homogeneity, ignoring perturbations. Given a universe in which inflation occurs at all, this is a very good approximation.

classical measure. This would indicate the importance of intrinsically quantum gravitational processes or some ultimate theory of initial conditions.

As we enter an era of precision inflationary cosmology, models of inflation will be subjected to increasingly refined measurement. In the effort to determine which models best reflect reality, the notion of naturalness, in the sense of genericness under the canonical measure on the space of trajectories, can be very useful. The methods developed in this chapter provide for quantitative probabilistic comparison among models of inflation, providing a new means of shedding light on the earliest moments of our universe.

Bibliography

[1] C. Cheung and G. N. Remmen, "Naturalness and the Weak Gravity Conjecture," *Phys. Rev. Lett.* **113** (2014) 051601, arXiv:1402.2287 [hep-ph].

[2] C. Cheung and G. N. Remmen, "Infrared Consistency and the Weak Gravity Conjecture," *JHEP* **1412** (2014) 087, arXiv:1407.7865 [hep-th].

[3] B. Bellazzini, C. Cheung, and G. N. Remmen, "Quantum Gravity Constraints from Unitarity and Analyticity," *Phys. Rev.* **D93** (2016) 064076, arXiv:1509.00851 [hep-th].

[4] C. Cheung and G. N. Remmen, "Positive Signs in Massive Gravity," *JHEP* **04** (2016) 002, arXiv:1601.04068 [hep-th].

[5] C. Cheung and G. N. Remmen, "Positivity of Curvature-Squared Corrections in Gravity," *Phys. Rev. Lett.* **118** (2017) 051601, arXiv:1608.02942 [hep-th].

[6] C. Cheung and G. N. Remmen, "Twofold Symmetries of the Pure Gravity Action," *JHEP* **01** (2017) 104, arXiv:1612.03927 [hep-th].

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

[8] N. Bao, J. Pollack, and G. N. Remmen, "Wormhole and Entanglement (Non-)Detection in the ER=EPR Correspondence," *JHEP* **11** (2015) 126, arXiv:1509.05426 [hep-th].

[9] G. N. Remmen, N. Bao, and J. Pollack, "Entanglement Conservation, ER=EPR, and a New Classical Area Theorem for Wormholes," *JHEP* **07** (2016) 048, arXiv:1604.08217 [hep-th].

[10] N. Bao, C. Cao, S. M. Carroll, A. Chatwin-Davies, N. Hunter-Jones, J. Pollack, and G. N. Remmen, "Consistency Conditions for an AdS Multiscale Entanglement Renormalization Ansatz Correspondence," *Phys. Rev.* **D91** (2015) 125036, arXiv:1504.06632 [hep-th].

[11] S. M. Carroll and G. N. Remmen, "What is the Entropy in Entropic Gravity?," *Phys. Rev.* **D93** (2016) 124052, arXiv:1601.07558 [hep-th].

[12] G. N. Remmen and S. M. Carroll, "Attractor Solutions in Scalar-Field Cosmology," *Phys.Rev.* **D88** (2013) 083518, arXiv:1309.2611 [gr-qc].

[13] G. N. Remmen and S. M. Carroll, "How Many e-Folds Should We Expect from High-Scale Inflation?," *Phys. Rev.* **D90** (2014) 063517, arXiv:1405.5538 [hep-th].

[14] **CMS** Collaboration, S. Chatrchyan et al., "Observation of a New Boson at a Mass of 125 GeV with the CMS Experiment at the LHC," *Phys.Lett.* **B716** (2012) 30, arXiv:1207.7235 [hep-ex].

[15] **ATLAS** Collaboration, G. Aad et al., "Observation of a New Particle in the Search for the Standard Model Higgs Boson with the ATLAS Detector at the LHC," *Phys.Lett.* **B716** (2012) 1, arXiv:1207.7214 [hep-ex].

[16] S. Weinberg, *The Quantum Theory of Fields. Vol. 1: Foundations.* Cambridge University Press, 2005.

[17] H. Elvang and Y.-t. Huang, "Scattering Amplitudes," arXiv:1308.1697 [hep-th].

[18] L. J. Dixon, "A Brief Introduction to Modern Amplitude Methods," in *Proceedings, 2012 European School of High-Energy Physics (ESHEP 2012): La Pommeraye, Anjou, France, June 6-19, 2012*, p. 31. 2014. arXiv:1310.5353 [hep-ph].

[19] K. G. Wilson and J. Kogut, "The Renormalization Group and the ϵ Expansion," *Phys. Rept.* **12** (1974) 75.

[20] M. Schwartz, *Quantum Field Theory and the Standard Model.* Cambridge University Press, 2013.

[21] J. F. Donoghue, "Introduction to the Effective Field Theory Description of Gravity," in *Advanced School on Effective Theories*, F. Cornet and M. J. Herrero, eds. 1995. arXiv:gr-qc/9512024 [gr-qc].

[22] J. F. Donoghue, "General Relativity as an Effective Field Theory: The Leading Quantum Corrections," *Phys. Rev.* **D50** (1994) 3874, arXiv:gr-qc/9405057 [gr-qc].

[23] M. Green, J. Schwarz, and E. Witten, *Superstring Theory.* Cambridge University Press, Cambridge, 1987.

[24] K. Becker, M. Becker, and J. H. Schwarz, *String Theory and M-Theory: A Modern Introduction.* Cambridge University Press, 2007.

[25] G. Veneziano, "Construction of a Crossing-Symmetric, Regge-Behaved Amplitude for Linearly Rising Trajectories," *Nuovo Cim.* **A57** (1968) 190.

[26] S. Caron-Huot, Z. Komargodski, A. Sever, and A. Zhiboedov, "Strings from Massive Higher Spins: The Asymptotic Uniqueness of the Veneziano Amplitude," arXiv:1607.04253 [hep-th].

[27] C. Vafa, "The String Landscape and the Swampland," arXiv:hep-th/0509212 [hep-th].

[28] N. Arkani-Hamed, S. Dimopoulos, and S. Kachru, "Predictive Landscapes and New Physics at a TeV," arXiv:hep-th/0501082 [hep-th].

[29] H. Ooguri and C. Vafa, "On the Geometry of the String Landscape and the Swampland," *Nucl. Phys.* **B766** (2007) 21, arXiv:hep-th/0605264 [hep-th].

[30] N. Arkani-Hamed, L. Motl, A. Nicolis, and C. Vafa, "The String Landscape, Black Holes and Gravity as the Weakest Force," *JHEP* **0706** (2007) 060, arXiv:hep-th/0601001 [hep-th].

[31] A. Adams, N. Arkani-Hamed, S. Dubovsky, A. Nicolis, and R. Rattazzi, "Causality, Analyticity and an IR Obstruction to UV Completion," *JHEP* **0610** (2006) 014, arXiv:hep-th/0602178 [hep-th].

[32] S. Weinberg, "Implications of Dynamical Symmetry Breaking: An Addendum," *Phys. Rev.* **D19** (1979) 1277.

[33] L. Susskind, "Dynamics of Spontaneous Symmetry Breaking in the Weinberg-Salam Theory," *Phys. Rev.* **D20** (1979) 2619.

[34] A. Jenkins and D. O'Connell, "The Story of $\mathcal{O}$: Positivity Constraints in Effective Field Theories," arXiv:hep-th/0609159 [hep-th].

[35] G. Dvali, A. Franca, and C. Gomez, "Road Signs for UV-Completion," arXiv:1204.6388 [hep-th].

[36] Y. Aharonov, A. Komar, and L. Susskind, "Superluminal Behavior, Causality, and Instability," *Phys. Rev.* **182** (1969) 1400.

[37] X. O. Camanho, J. D. Edelstein, J. Maldacena, and A. Zhiboedov, "Causality Constraints on Corrections to the Graviton Three-Point Coupling," *JHEP* **02** (2016) 020, arXiv:1407.5597 [hep-th].

[38] K. Hinterbichler, "Theoretical Aspects of Massive Gravity," *Rev.Mod.Phys.* **84** (2012) 671, arXiv:1105.3735 [hep-th].

[39] H. van Dam and M. J. G. Veltman, "Massive and Mass-less Yang-Mills and Gravitational Fields," *Nucl. Phys.* **B22** (1970) 397.

[40] V. I. Zakharov, "Linearized Gravitation Theory and the Graviton Mass," *JETP Lett.* **12** (1970) 312.

[41] D. G. Boulware and S. Deser, "Can Gravitation Have a Finite Range?," *Phys. Rev.* **D6** (1972) 3368.

[42] N. Arkani-Hamed, H. Georgi, and M. D. Schwartz, "Effective Field Theory for Massive Gravitons and Gravity in Theory Space," *Annals Phys.* **305** (2003) 96, arXiv:hep-th/0210184 [hep-th].

[43] C. de Rham and G. Gabadadze, "Generalization of the Fierz-Pauli Action," *Phys. Rev.* **D82** (2010) 044020, arXiv:1007.0443 [hep-th].

[44] C. de Rham, G. Gabadadze, and A. J. Tolley, "Resummation of Massive Gravity," *Phys. Rev. Lett.* **106** (2011) 231101, arXiv:1011.1232 [hep-th].

[45] H. Kawai, D. C. Lewellen, and S. H. H. Tye, "A Relation Between Tree Amplitudes of Closed and Open Strings," *Nucl. Phys.* **B269** (1986) 1.

[46] Z. Bern, J. J. M. Carrasco, and H. Johansson, "New Relations for Gauge-Theory Amplitudes," *Phys. Rev.* **D78** (2008) 085011, arXiv:0805.3993 [hep-ph].

[47] M. Veltman, "Quantum Theory of Gravitation," in *Les Houches 1975, Proceedings, Methods In Field Theory*, p. 265. Amsterdam, 1976.

[48] B. S. DeWitt, "Quantum Theory of Gravity. II. The Manifestly Covariant Theory," *Phys.Rev.* **162** (1967) 1195.

[49] B. S. DeWitt, "Quantum Theory of Gravity. III. Applications of the Covariant Theory," *Phys.Rev.* **162** (1967) 1239.

[50] L. Susskind, "The World as a Hologram," *J.Math.Phys.* **36** (1995) 6377, arXiv:hep-th/9409089 [hep-th].

[51] G. 't Hooft, "Dimensional Reduction in Quantum Gravity," arXiv:gr-qc/9310026 [gr-qc].

[52] J. M. Maldacena, "The Large-N Limit of Superconformal Field Theories and Supergravity," *Int.J.Theor.Phys.* **38** (1999) 1113, arXiv:hep-th/9711200 [hep-th].

[53] E. Witten, "Anti-de Sitter Space and Holography," *Adv. Theor. Math. Phys.* **2** (1998) 253, arXiv:hep-th/9802150 [hep-th].

[54] 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, arXiv:hep-th/9905111 [hep-th].

[55] A. Einstein and N. Rosen, "The Particle Problem in the General Theory of Relativity," *Phys.Rev.* **48** (1935) 73.

[56] J. M. Maldacena, "Eternal Black Holes in Anti-de Sitter," *JHEP* **0304** (2003) 021, arXiv:hep-th/0106112 [hep-th].

[57] J. Maldacena and L. Susskind, "Cool Horizons for Entangled Black Holes," *Fortsch.Phys.* **61** (2013) 781, arXiv:1306.0533 [hep-th].

[58] A. Einstein, B. Podolsky, and N. Rosen, "Can Quantum-Mechanical Description of Physical Reality be Considered Complete?," *Phys.Rev.* **47** (1935) 777.

[59] T. Jacobson, "Thermodynamics of Spacetime: The Einstein Equation of State," *Phys.Rev.Lett.* **75** (1995) 1260, arXiv:gr-qc/9504004 [gr-qc].

[60] T. Jacobson, "Entanglement Equilibrium and the Einstein Equation," *Phys. Rev. Lett.* **116** (2016) 201101, arXiv:1505.04753 [gr-qc].

[61] H. Casini, "Relative Entropy and the Bekenstein Bound," *Class.Quant.Grav.* **25** (2008) 205021, arXiv:0804.2182 [hep-th].

[62] A. H. Guth, "The Inflationary Universe: A Possible Solution to the Horizon and Flatness Problems," *Phys.Rev.* **D23** (1981) 347.

[63] A. D. Linde, "A New Inflationary Universe Scenario: A Possible Solution of the Horizon, Flatness, Homogeneity, Isotropy and Primordial Monopole Problems," *Phys.Lett.* **B108** (1982) 389.

[64] G. Gibbons, S. Hawking, and J. Stewart, "A Natural Measure on the Set of All Universes," *Nucl.Phys.* **B281** (1987) 736.

[65] **Planck** Collaboration, P. A. R. Ade et al., "Planck 2013 Results. XVI. Cosmological Parameters," *Astron. Astrophys.* **571** (2014) A16, arXiv:1303.5076 [astro-ph.CO].

[66] J.D. Bekenstein, "Nonexistence of Baryon Number for Static Black Holes," *Phys.Rev.* **D5** (1972) 1239.

[67] J.D. Bekenstein, "Transcendence of the Law of Baryon-Number Conservation in Black Hole Physics," *Phys.Rev.Lett.* **28** (1972) 452.

[68] S.W. Hawking and G.F.R. Ellis, *The Large Scale Structure of Space-Time.* Cambridge University Press, Cambridge, 1973.

[69] S.R. Coleman, J. Preskill, and F. Wilczek, "Quantum Hair on Black Holes," *Nucl.Phys.* **B378** (1992) 175, arXiv:hep-th/9201059 [hep-th].

[70] G.W. Gibbons, "Vacuum Polarization and the Spontaneous Loss of Charge by Black Holes," *Commun.Math.Phys.* **44** (1975) 245.

[71] R. Bousso, "The Holographic Principle," *Rev.Mod.Phys.* **74** (2002) 825, arXiv:hep-th/0203101 [hep-th].

[72] L. Susskind, "Trouble for Remnants," arXiv:hep-th/9501106 [hep-th].

[73] S.B. Giddings, "Black Holes and Massive Remnants," *Phys.Rev.* **D46** (1992) 1347, arXiv:hep-th/9203059 [hep-th].

[74] S.L. Adler and R.B. Pearson, "'No Hair' Theorems for the Abelian Higgs and Goldstone Models," *Phys.Rev.* **D18** (1978) 2798.

[75] J. Preskill, "Quantum Hair," *Phys.Scripta* **T36** (1991) 258.

[76] G. Dvali, "Black Holes and Large N Species Solution to the Hierarchy Problem," *Fortsch.Phys.* **58** (2010) 528, arXiv:0706.2050 [hep-th].

[77] S. Dimopoulos, S. Kachru, J. McGreevy, and J.G. Wacker, "N-flation," *JCAP* **0808** (2008) 003, arXiv:hep-th/0507205 [hep-th].

[78] T.A. Wagner, S. Schlamminger, J.H. Gundlach, and E.G. Adelberger, "Torsion-Balance Tests of the Weak Equivalence Principle," *Class.Quant.Grav.* **29** (2012) 184002, arXiv:1207.2442 [gr-qc].

[79] S. Schlamminger, K.-Y. Choi, T.A. Wagner, J.H. Gundlach, and E.G. Adelberger, "Test of the Equivalence Principle Using a Rotating Torsion Balance," *Phys.Rev.Lett.* **100** (2008) 041101, arXiv:0712.0607 [gr-qc].

[80] S.M. Ransom, I.H. Stairs, A.M. Archibald, J.W.T. Hessels, D.L. Kaplan, et al., "A Millisecond Pulsar in a Stellar Triple System," *Nature* **505** (2014) 520, arXiv:1401.0535 [astro-ph.SR].

[81] T.W. Murphy, Jr., E.G. Adelberger, J.B.R. Battat, C.D. Hoyle, N.H. Johnson, et al., "APOLLO: Millimeter Lunar Laser Ranging," *Class.Quant.Grav.* **29** (2012) 184005.

[82] J. Overduin, F. Everitt, P. Worden, and J. Mester, "STEP and Fundamental Physics," *Class.Quant.Grav.* **29** (2012) 184012, arXiv:1401.4784 [gr-qc].

[83] A.M. Nobili, M. Shao, R. Pegna, G. Zavattini, S.G. Turyshev, et al., "'Galileo Galilei' (GG): Space Test of the Weak Equivalence Principle to 10^{-17} and Laboratory Demonstrations," *Class.Quant.Grav.* **29** (2012) 184011.

[84] P. Touboul, G. Metris, V. Lebat, and A. Robert, "The MICROSCOPE Experiment, Ready for the In-Orbit Test of the Equivalence Principle," *Class.Quant.Grav.* **29** (2012) 184010.

[85] E. Giusarma, R. de Putter, S. Ho, and O. Mena, "Constraints on Neutrino Masses from Planck and Galaxy Clustering Data," *Phys.Rev.* **D88** (2013) 063515, arXiv:1306.5544 [astro-ph.CO].

[86] R. de Putter et al., "New Neutrino Mass Bounds from Sloan Digital Sky

Survey III Data Release 8 Photometric Luminous Galaxies," *Astrophys.J.* **761** (2012) 12, arXiv:1201.1909 [astro-ph.CO].

[87] C. Cheung and Y. Nomura, "Higgs Descendants," *Phys.Rev.* **D86** (2012) 015004, arXiv:1112.3043 [hep-ph].

[88] W. Heisenberg and H. Euler, "Consequences of Dirac's Theory of Positrons," *Z.Phys.* **98** (1936) 714, arXiv:physics/0605038 [physics].

[89] J. S. Schwinger, "On Gauge Invariance and Vacuum Polarization," *Phys.Rev.* **82** (1951) 664.

[90] V. Weisskopf, "The Electrodynamics of the Vacuum Based on the Quantum Theory of the Electron," *Kong.Dans.Vid.Selsk.Math-fys.Medd.* **14** (1936) 1.

[91] S. Deser, R. Jackiw, and G. 't Hooft, "Three-Dimensional Einstein Gravity: Dynamics of Flat Space," *Annals Phys.* **152** (1984) 220.

[92] A. Ritz and R. Delbourgo, "The Low Energy Effective Lagrangian for Photon Interactions in Any Dimension," *Int.J.Mod.Phys.* **A11** (1996) 253, arXiv:hep-th/9503160 [hep-th].

[93] I. T. Drummond and S. J. Hathrell, "QED Vacuum Polarization in a Background Gravitational Field and its Effect on the Velocity of Photons," *Phys.Rev.* **D22** (1980) 343.

[94] M. Chaichian and J. Fischer, "Higher Dimensional Space-Time and Unitarity Bound on the Scattering Amplitude," *Nucl.Phys.* **B303** (1988) 557.

[95] M. Chaichian, J. Fischer, and Y. Vernov, "Generalization of the Froissart-Martin Bounds to Scattering in a Space-Time of General Dimension," *Nucl.Phys.* **B383** (1992) 151.

[96] S. B. Giddings and M. Srednicki, "High-Energy Gravitational Scattering and Black Hole Resonances," *Phys.Rev.* **D77** (2008) 085025, arXiv:0711.5012 [hep-th].

[97] S. B. Giddings and R. A. Porto, "The Gravitational S-Matrix," *Phys. Rev.* **D81** (2010) 025002, arXiv:0908.0004 [hep-th].

[98] H. Alnes, F. Ravndal, and I. K. Wehus, "Black-Body Radiation in Extra Dimensions," *J.Phys.* **A40** (2007) 14309, arXiv:quant-ph/0506131 [quant-ph].

[99] T. R. Cardoso and A. S. de Castro, "The Blackbody Radiation in D-Dimensional Universes," *Rev.Bras.Ens.Fis.* **27** (2005) 559, arXiv:quant-ph/0510002 [quant-ph].

[100] A. A. Garcia and C. Campuzano, "All Static Circularly Symmetric Perfect Fluid Solutions of (2+1) Gravity," *Phys.Rev.* **D67** (2003) 064014, arXiv:gr-qc/0211014 [gr-qc].

[101] R. Tolman, *The Theory of the Relativity of Motion.* University of California Press, 1917.

[102] G. Benford, D. Book, and W. Newcomb, "The Tachyonic Antitelephone," *Phys.Rev.* **D2** (1970) 263.

[103] D. Bohm, *The Special Theory of Relativity.* Taylor & Francis, 2012.

[104] S. M. Carroll, E. Farhi, A. H. Guth, and K. D. Olum, "Energy-Momentum Restrictions on the Creation of Gott Time Machines," *Phys.Rev.* **D50** (1994) 6190, arXiv:gr-qc/9404065 [gr-qc].

[105] S. Deser and R. Jackiw, "Time Travel?," *Comments Nucl.Part.Phys.* **20** (1992) 337, arXiv:hep-th/9206094 [hep-th].

[106] S. Deser, R. Jackiw, and G. 't Hooft, "Physical Cosmic Strings Do Not Generate Closed Timelike Curves," *Phys.Rev.Lett.* **68** (1992) 267.

[107] G. V. Dunne, "Heisenberg-Euler Effective Lagrangians: Basics and Extensions," in *From Fields to Strings: Circumnavigating Theoretical Physics: Ian Kogan Memorial Collection, Vol. 1*, M. Shifman, A. Vainshtein, J. Wheater, and I. Kogan, eds., p. 445. World Scientific, 2005. arXiv:hep-th/0406216 [hep-th].

[108] F. Bastianelli, J. M. Davila, and C. Schubert, "Gravitational Corrections to the Euler-Heisenberg Lagrangian," *JHEP* **0903** (2009) 086, arXiv:0812.4849 [hep-th].

[109] F. Bastianelli, O. Corradini, J. Davila, and C. Schubert, "On the Low-Energy Limit of One-Loop Photon–Graviton Amplitudes," *Phys.Lett.* **B716** (2012) 345, arXiv:1202.4502 [hep-th].

[110] S. Deser and P. van Nieuwenhuizen, "One-Loop Divergences of Quantized Einstein-Maxwell Fields," *Phys.Rev.* **D10** (1974) 401.

[111] B. Bellazzini, L. Martucci, and R. Torre, "Symmetries, Sum Rules and Constraints on Effective Field Theories," *JHEP* **1409** (2014) 100, arXiv:1405.2960 [hep-th].

[112] S. Dimopoulos and G. L. Landsberg, "Black Holes at the Large Hadron Collider," *Phys. Rev. Lett.* **87** (2001) 161602, arXiv:hep-ph/0106295 [hep-ph].

[113] J. Distler, B. Grinstein, R. A. Porto, and I. Z. Rothstein, "Falsifying Models of New Physics via WW Scattering," *Phys. Rev. Lett.* **98** (2007) 041601, arXiv:hep-ph/0604255 [hep-ph].

[114] R. C. Brower, J. Polchinski, M. J. Strassler, and C.-I. Tan, "The Pomeron and Gauge/String Duality," *JHEP* **0712** (2007) 005, arXiv:hep-th/0603115 [hep-th].

[115] C. Cheung, D. O'Connell, and B. Wecht, "BCFW Recursion Relations and String Theory," *JHEP* **1009** (2010) 052, arXiv:1002.4674 [hep-th].

[116] S. Weinberg, "Feynman Rules for Any Spin," *Phys.Rev.* **133** (1964) B1318.

[117] S. Weinberg, "Feynman Rules for Any Spin. III," *Phys.Rev.* **181** (1969) 1893.

[118] C. Misner, K. Thorne, and J. Wheeler, *Gravitation*. W. H. Freeman, 1973.

[119] R. Daniels and G. Shore, "'Faster than Light' Photons and Charged Black Holes," *Nucl.Phys.* **B425** (1994) 634, arXiv:hep-th/9310114 [hep-th].

[120] F. Ravndal, "Radiative Corrections to the Stefan-Boltzmann Law," arXiv:hep-ph/9709220 [hep-ph].

[121] T. J. Hollowood and G. M. Shore, "The Causal Structure of QED in Curved Spacetime: Analyticity and the Refractive Index," *JHEP* **0812** (2008) 091, arXiv:0806.1019 [hep-th].

[122] J. Hartle and S. Hawking, "Path Integral Derivation of Black Hole Radiance," *Phys.Rev.* **D13** (1976) 2188.

[123] M. Visser, "Gravitational Vacuum Polarization. I. Energy Conditions in the Hartle-Hawking Vacuum," *Phys.Rev.* **D54** (1996) 5103, arXiv:gr-qc/9604007 [gr-qc].

[124] L. Susskind, L. Thorlacius, and J. Uglum, "The Stretched Horizon and Black Hole Complementarity," *Phys.Rev.* **D48** (1993) 3743, arXiv:hep-th/9306069 [hep-th].

[125] P. Hayden and J. Preskill, "Black Holes as Mirrors: Quantum Information in Random Subsystems," *JHEP* **0709** (2007) 120, arXiv:0708.4025 [hep-th].

[126] M. Kruskal, "Maximal Extension of Schwarzschild Metric," *Phys.Rev.* **119** (1960) 1743.

[127] R. W. Fuller and J. A. Wheeler, "Causality and Multiply Connected Space-Time," *Phys.Rev.* **128** (1962) 919.

[128] K. S. Thorne, "Closed Timelike Curves," in *General Relativity and Gravitation 1992*, R. Gleiser, C. Kozameh, and O. Moreschi, eds., p. 295. Taylor & Francis, 1993.

[129] M. Morris, K. Thorne, and U. Yurtsever, "Wormholes, Time Machines, and the Weak Energy Condition," *Phys.Rev.Lett.* **61** (1988) 1446.

[130] R. Sundrum, "Gravitational Lorentz Violation and Superluminality via AdS/CFT Duality," *Phys.Rev.* **D77** (2008) 086002, arXiv:0708.1871 [hep-th].

[131] M. Van Raamsdonk, "Building up Spacetime with Quantum Entanglement," *Gen.Rel.Grav.* **42** (2010) 2323, arXiv:1005.3035 [hep-th].

[132] S. Ryu and T. Takayanagi, "Holographic Derivation of Entanglement Entropy from AdS/CFT," *Phys.Rev.Lett.* **96** (2006) 181602, arXiv:hep-th/0603001 [hep-th].

[133] S. Leichenauer, "Disrupting Entanglement of Black Holes," *Phys. Rev.* **D90** (2014) 046009, arXiv:1405.7365 [hep-th].

[134] S. Dubovsky and S. Sibiryakov, "Spontaneous Breaking of Lorentz Invariance, Black Holes and Perpetuum Mobile of the 2nd Kind," *Phys.Lett.* **B638** (2006) 509, arXiv:hep-th/0603158 [hep-th].

[135] C. Fronsdal, "Massless Fields with Integer Spin," *Phys.Rev.* **D18** (1978) 3624.

[136] L. Singh and C. Hagen, "Lagrangian Formulation for Arbitrary Spin. 1. The Boson Case," *Phys.Rev.* **D9** (1974) 898.

[137] A. V. Manohar and V. Mateu, "Dispersion Relation Bounds for $\pi\pi$ Scattering," *Phys. Rev.* **D77** (2008) 094019, arXiv:0801.3222 [hep-ph].

[138] T. N. Pham and T. N. Truong, "Evaluation of the Derivative Quartic Terms

of the Meson Chiral Lagrangian from Forward Dispersion Relations," *Phys. Rev.* **D31** (1985) 3027.

[139] B. Zwiebach, "Curvature Squared Terms and String Theories," *Phys. Lett.* **B156** (1985) 315.

[140] B. Zumino, "Gravity Theories in More than Four Dimensions," *Phys. Rept.* **137** (1986) 109.

[141] S. Fulling, R. C. King, B. Wybourne, and C. Cummins, "Normal Forms for Tensor Polynomials. 1: The Riemann Tensor," *Class.Quant.Grav.* **9** (1992) 1151.

[142] H. Georgi, "On-Shell Effective Field Theory," *Nucl. Phys.* **B361** (1991) 339.

[143] S. Deser and D. Seminara, "Tree Amplitudes and Two-Loop Counterterms in $D = 11$ Supergravity," *Phys. Rev.* **D62** (2000) 084010, arXiv:hep-th/0002241 [hep-th].

[144] A. Gruzinov and M. Kleban, "Causality Constrains Higher Curvature Corrections to Gravity," *Class. Quant. Grav.* **24** (2007) 3521, arXiv:hep-th/0612015 [hep-th].

[145] I. Low, R. Rattazzi, and A. Vichi, "Theoretical Constraints on the Higgs Effective Couplings," *JHEP* **04** (2010) 126, arXiv:0907.5413 [hep-ph].

[146] M. de Roo, H. Suelmann, and A. Wiedemann, "Supersymmetric R^4-Actions in Ten Dimensions," *Phys. Lett.* **B280** (1992) 39.

[147] A. A. Tseytlin, "Heterotic – Type I Superstring Duality and Low-Energy Effective Actions," *Nucl. Phys.* **B467** (1996) 383, arXiv:hep-th/9512081 [hep-th].

[148] I. Jack, D. R. T. Jones, and N. Mohammedi, "The Four-Loop Metric β-Function for the Bosonic σ-Model," *Phys. Lett.* **B220** (1989) 171.

[149] I. Jack, D. R. T. Jones, and N. Mohammedi, "A Four-Loop Calculation of the Metric β-Function for the Bosonic σ-Model and the String Effective Action," *Nucl. Phys.* **B322** (1989) 431.

[150] D. J. Gross and E. Witten, "Superstring Modifications of Einstein's Equations," *Nucl. Phys.* **B277** (1986) 1.

[151] D. J. Gross and J. H. Sloan, "The Quartic Effective Action for the Heterotic String," *Nucl. Phys.* **B291** (1987) 41.

[152] Y. Kikuchi, C. Marzban, and Y. J. Ng, "Heterotic String Modifications of Einstein's and Yang-Mills' Actions," *Phys. Lett.* **B176** (1986) 57.

[153] Z. Bern, C. Cheung, H.-H. Chi, S. Davies, L. Dixon, and J. Nohle, "Evanescent Effects Can Alter Ultraviolet Divergences in Quantum Gravity without Physical Consequences," *Phys. Rev. Lett.* **115** (2015) 211301, arXiv:1507.06118 [hep-th].

[154] K. Melnikov and V. G. Serbo, "Processes with the t-Channel Singularity in the Physical Region: Finite Beam Sizes Make Cross-Sections Finite," *Nucl. Phys.* **B483** (1997) 67, arXiv:hep-ph/9601290 [hep-ph].

[155] A. Martin, "Extension of the Axiomatic Analyticity Domain of Scattering Amplitudes by Unitarity. 1.," *Nuovo Cim.* **A42** (1965) 930.

[156] A. Nicolis, R. Rattazzi, and E. Trincherini, "Energy's and Amplitudes' Positivity," *JHEP* **05** (2010) 095, arXiv:0912.4258 [hep-th]. [Erratum: JHEP11,128(2011)].

[157] S. Sarkar and A. C. Wall, "Second Law Violations in Lovelock Gravity for Black Hole Mergers," *Phys. Rev.* **D83** (2011) 124048, arXiv:1011.4988 [gr-qc].

[158] H. Elvang, D. Z. Freedman, L.-Y. Hung, M. Kiermaier, R. C. Myers, and S. Theisen, "On Renormalization Group Flows and the a-Theorem in 6d," *JHEP* **10** (2012) 011, arXiv:1205.3994 [hep-th].

[159] D. Baumann, D. Green, H. Lee, and R. A. Porto, "Signs of Analyticity in Single-Field Inflation," *Phys. Rev.* **D93** (2016) 023523, arXiv:1502.07304 [hep-th].

[160] D. Croon, V. Sanz, and J. Setford, "Goldstone Inflation," *JHEP* **10** (2015) 020, arXiv:1503.08097 [hep-ph].

[161] B. Heidenreich, M. Reece, and T. Rudelius, "Weak Gravity Strongly Constrains Large-Field Axion Inflation," *JHEP* **12** (2015) 108, arXiv:1506.03447 [hep-th].

[162] A. de la Fuente, P. Saraswat, and R. Sundrum, "Natural Inflation and Quantum Gravity," *Phys. Rev. Lett.* **114** (2015) 151303, arXiv:1412.3457 [hep-th].

[163] J. Brown, W. Cottrell, G. Shiu, and P. Soler, "Fencing in the Swampland: Quantum Gravity Constraints on Large Field Inflation," *JHEP* **10** (2015) 023, arXiv:1503.04783 [hep-th].

[164] A. Nicolis, R. Rattazzi, and E. Trincherini, "The Galileon as a Local Modification of Gravity," *Phys. Rev.* **D79** (2009) 064036, arXiv:0811.2197 [hep-th].

[165] S. F. Hassan and R. A. Rosen, "Resolving the Ghost Problem in Nonlinear Massive Gravity," *Phys. Rev. Lett.* **108** (2012) 041101, arXiv:1106.3344 [hep-th].

[166] G. R. Dvali, G. Gabadadze, and M. Porrati, "4-D Gravity on a Brane in 5-D Minkowski Space," *Phys. Lett.* **B485** (2000) 208, arXiv:hep-th/0005016 [hep-th].

[167] M. A. Luty, M. Porrati, and R. Rattazzi, "Strong Interactions and Stability in the DGP Model," *JHEP* **09** (2003) 029, arXiv:hep-th/0303116 [hep-th].

[168] A. Aubert, "Strong Coupling in Massive Gravity by Direct Calculation," *Phys. Rev.* **D69** (2004) 087502, arXiv:hep-th/0312246 [hep-th].

[169] M. D. Schwartz, "Constructing Gravitational Dimensions," *Phys. Rev.* **D68** (2003) 024029, arXiv:hep-th/0303114 [hep-th].

[170] M. Froissart, "Asymptotic Behavior and Subtractions in the Mandelstam Representation," *Phys. Rev.* **123** (1961) 1053.

[171] A. Martin, "Unitarity and High-Energy Behavior of Scattering Amplitudes," *Phys. Rev.* **129** (1963) 1432.

[172] A. A. Ahmadi, A. Olshevsky, P. A. Parrilo, and J. N. Tsitsiklis, "NP-Hardness of Deciding Convexity of Quartic Polynomials and Related Problems," *Mathematical Programming* **137** (2013) 453, arXiv:1012.1908 [math.OC].

[173] A. A. Ahmadi, "On the Difficulty of Deciding Asymptotic Stability of Cubic Homogeneous Vector Fields," in *Proceedings of the American Control Conference*, p. 3334. 2012. arXiv:1112.0741 [math.OC].

[174] N. Chow and J. Khoury, "Galileon Cosmology," *Phys. Rev.* **D80** (2009) 024037, arXiv:0905.1325 [hep-th].

[175] C. de Rham, G. Gabadadze, L. Heisenberg, and D. Pirtskhalava, "Cosmic Acceleration and the Helicity-0 Graviton," *Phys. Rev.* **D83** (2011) 103516, arXiv:1010.1780 [hep-th].

[176] C. de Rham and L. Heisenberg, "Cosmology of the Galileon from Massive Gravity," *Phys. Rev.* **D84** (2011) 043503, arXiv:1106.3312 [hep-th].

[177] C. Cheung, K. Kampf, J. Novotny, and J. Trnka, "Effective Field Theories from Soft Limits of Scattering Amplitudes," *Phys. Rev. Lett.* **114** (2015) 221602, arXiv:1412.4095 [hep-th].

[178] Z. Komargodski and A. Schwimmer, "On Renormalization Group Flows in Four Dimensions," *JHEP* **12** (2011) 099, arXiv:1107.3987 [hep-th].

[179] B. Bellazzini, "Softness and Amplitudes' Positivity for Spinning Particles," *JHEP* **02** (2017) 034, arXiv:1605.06111 [hep-th].

[180] Y. Kats and P. Petrov, "Effect of Curvature Squared Corrections in AdS on the Viscosity of the Dual Gauge Theory," *JHEP* **01** (2009) 044, arXiv:0712.0743 [hep-th].

[181] S. Weinberg, *Gravitation and Cosmology*. John Wiley and Sons, New York, 1972.

[182] G. Källén, "On the Definition of the Renormalization Constants in Quantum Electrodynamics," *Helv. Phys. Acta* **25** (1952) 417.

[183] H. Lehmann, "On the Properties of Propagation Functions and Renormalization Contants of Quantized Fields," *Nuovo Cim.* **11** (1954) 342.

[184] M. Reece et al., *forthcoming*.

[185] J. J. M. Carrasco, "Gauge and Gravity Amplitude Relations," arXiv:1506.00974 [hep-th].

[186] R. Monteiro, D. O'Connell, and C. D. White, "Black Holes and the Double Copy," *JHEP* **12** (2014) 056, arXiv:1410.0239 [hep-th].

[187] A. K. Ridgway and M. B. Wise, "Static Spherically Symmetric Kerr-Schild Metrics and Implications for the Classical Double Copy," *Phys. Rev.* **D94** (2016) 044023, arXiv:1512.02243 [hep-th].

[188] A. Luna, R. Monteiro, D. O'Connell, and C. D. White, "The Classical Double

Copy for Taub-NUT Spacetime," *Phys. Lett.* **B750** (2015) 272, arXiv:1507.01869 [hep-th].

[189] A. Luna, R. Monteiro, I. Nicholson, D. O'Connell, and C. D. White, "The Double Copy: Bremsstrahlung and Accelerating Black Holes," *JHEP* **06** (2016) 023, arXiv:1603.05737 [hep-th].

[190] Y.-Z. Chu, "More on Cosmological Gravitational Waves and their Memories," arXiv:1611.00018 [gr-qc].

[191] W. D. Goldberger and A. K. Ridgway, "Radiation and the Classical Double Copy for Color Charges," arXiv:1611.03493 [hep-th].

[192] F. Cachazo, S. He, and E. Y. Yuan, "Scattering of Massless Particles in Arbitrary Dimensions," *Phys. Rev. Lett.* **113** (2014) 171601, arXiv:1307.2199 [hep-th].

[193] Z. Bern and A. K. Grant, "Perturbative Gravity from QCD Amplitudes," *Phys. Lett.* **B457** (1999) 23, arXiv:hep-th/9904026 [hep-th].

[194] C. Hull and B. Zwiebach, "Double Field Theory," *JHEP* **09** (2009) 099, arXiv:0904.4664 [hep-th].

[195] O. Hohm, C. Hull, and B. Zwiebach, "Background Independent Action for Double Field Theory," *JHEP* **07** (2010) 016, arXiv:1003.5027 [hep-th].

[196] O. Hohm, C. Hull, and B. Zwiebach, "Generalized Metric Formulation of Double Field Theory," *JHEP* **08** (2010) 008, arXiv:1006.4823 [hep-th].

[197] O. Hohm, "On Factorizations in Perturbative Quantum Gravity," *JHEP* **04** (2011) 103, arXiv:1103.0032 [hep-th].

[198] W. Siegel, "Two Vierbein Formalism for String-Inspired Axionic Gravity," *Phys. Rev.* **D47** (1993) 5453, arXiv:hep-th/9302036 [hep-th].

[199] W. Siegel, "Superspace Duality in Low-Energy Superstrings," *Phys. Rev.* **D48** (1993) 2826, arXiv:hep-th/9305073 [hep-th].

[200] W. Siegel, "Manifest Duality in Low-Energy Superstrings," in *International Conference on Strings 93, Berkeley, California, May 24-29, 1993*, p. 353. 1993. arXiv:hep-th/9308133 [hep-th].

[201] K. Kampf, J. Novotny, and J. Trnka, "Tree-Level Amplitudes in the Nonlinear Sigma Model," *JHEP* **05** (2013) 032, arXiv:1304.3048 [hep-th].

[202] R. Haag, "Quantum Field Theories with Composite Particles and Asymptotic Conditions," *Phys. Rev.* **112** (1958) 669.

[203] J. F. Donoghue, E. Golowich, and B. R. Holstein, *Dynamics of the Standard Model*. Camb. Monogr. Part. Phys. Nucl. Phys. Cosmol. Cambridge University Press, 1992.

[204] H. Elvang and D. Z. Freedman, *unpublished notes*. 2007.

[205] R. Britto, F. Cachazo, and B. Feng, "New Recursion Relations for Tree Amplitudes of Gluons," *Nucl. Phys.* **B715** (2005) 499, arXiv:hep-th/0412308 [hep-th].

[206] R. Britto, F. Cachazo, B. Feng, and E. Witten, "Direct Proof of Tree-Level Recursion Relation in Yang-Mills Theory," *Phys. Rev. Lett.* **94** (2005) 181602, arXiv:hep-th/0501052 [hep-th].

[207] N. Arkani-Hamed and J. Kaplan, "On Tree Amplitudes in Gauge Theory and Gravity," *JHEP* **04** (2008) 076, arXiv:0801.2385 [hep-th].

[208] C. Cheung, "On-Shell Recursion Relations for Generic Theories," *JHEP* **03** (2010) 098, arXiv:0808.0504 [hep-th].

[209] C. Cheung and C.-H. Shen, "Symmetry and Action for Flavor-Kinematics Duality," *Phys. Rev. Lett.* **118** (2017) 121601, arXiv:1612.00868 [hep-th].

[210] J. M. Bardeen, B. Carter, and S. Hawking, "The Four Laws of Black Hole Mechanics," *Commun.Math.Phys.* **31** (1973) 161.

[211] J. D. Bekenstein, "Black Holes and Entropy," *Phys.Rev.* **D7** (1973) 2333.

[212] T. Faulkner, M. Guica, T. Hartman, R. C. Myers, and M. Van Raamsdonk, "Gravitation from Entanglement in Holographic CFTs," *JHEP* **1403** (2014) 051, arXiv:1312.7856 [hep-th].

[213] B. Swingle, "Constructing Holographic Spacetimes Using Entanglement Renormalization," arXiv:1209.3304 [hep-th].

[214] S. Hawking, "Breakdown of Predictability in Gravitational Collapse," *Phys.Rev.* **D14** (1976) 2460.

[215] A. Almheiri, D. Marolf, J. Polchinski, and J. Sully, "Black Holes: Complementarity or Firewalls?," *JHEP* **1302** (2013) 062, arXiv:1207.3123 [hep-th].

[216] S. L. Braunstein, "Black Hole Entropy as Entropy of Entanglement, or It's Curtains for the Equivalence Principle," arXiv:0907.1190v1 [quant-ph].

[217] K. Jensen and A. Karch, "Holographic Dual of an Einstein-Podolsky-Rosen Pair has a Wormhole," *Phys.Rev.Lett.* **111** (2013) 211602, arXiv:1307.1132 [hep-th].

[218] J. Sonner, "Holographic Schwinger Effect and the Geometry of Entanglement," *Phys.Rev.Lett.* **111** (2013) 211603, arXiv:1307.6850 [hep-th].

[219] J. Bell, "On the Einstein-Podolsky-Rosen Paradox," *Physics* **1** (1964) 195.

[220] W. Wootters and W. Zurek, "A Single Quantum Cannot be Cloned," *Nature* **299** (1982) 802.

[221] R. Geroch, *Singularities in the Spacetime of General Relativity: Their Definition, Existence, and Local Characterization.* PhD thesis, Princeton University, 1967.

[222] R. Geroch and G. T. Horowitz, "Global Structure of Spacetimes," in *General Relativity: An Einstein Centenary Survey*, S. W. Hawking and W. Israel, eds., p. 212. Cambridge University Press, 1979.

[223] F. Tipler, "Singularities and Causality Violation," *Annals Phys.* **108** (1977) 1.

[224] F. Tipler, "Singularities and Causality Violation," *General Relativity and Gravitation* **10** (1979) 983.

[225] P. Hájíček, "Causality in Non-Hausdorff Space-Times," *Comm. Math. Phys.* **21** (1971) 75.

[226] J. L. Friedman, K. Schleich, and D. M. Witt, "Topological Censorship," *Phys.Rev.Lett.* **71** (1993) 1486, arXiv:gr-qc/9305017 [gr-qc].

[227] G. Galloway, K. Schleich, D. Witt, and E. Woolgar, "Topological Censorship and Higher Genus Black Holes," *Phys.Rev.* **D60** (1999) 104039, arXiv:gr-qc/9902061 [gr-qc].

[228] D. Dieks, "Communication by EPR Devices," *Phys.Lett.* **A92** (1982) 271.

[229] G. T. Horowitz and D. Marolf, "A New Approach to String Cosmology," *JHEP* **9807** (1998) 014, arXiv:hep-th/9805207 [hep-th].

[230] R. P. Geroch, "Domain of Dependence," *J. Math. Phys.* **11** (1970) 437.

[231] V. Rubakov, "The Null Energy Condition and its Violation," *Phys.Usp.* **57** (2014) 128, arXiv:1401.4024 [hep-th].

[232] T. Roman, "Quantum Stress Energy Tensors and the Weak Energy Condition," *Phys.Rev.* **D33** (1986) 3526.

[233] R. Horodecki, P. Horodecki, M. Horodecki, and K. Horodecki, "Quantum Entanglement," *Rev.Mod.Phys.* **81** (2009) 865, arXiv:quant-ph/0702225 [quant-ph].

[234] V. Frolov and A. Zelnikov, *Introduction to Black Hole Physics*. Oxford University Press, 2011.

[235] T. A. Brun, M. M. Wilde, and A. Winter, "Quantum State Cloning Using Deutschian Closed Timelike Curves," *Phys. Rev. Lett.* **111** (2013) 190401, arXiv:1306.1795 [quant-ph].

[236] G. 't Hooft, "On the Quantum Structure of a Black Hole," *Nucl. Phys.* **B256** (1985) 727.

[237] G. 't Hooft, "The Black Hole Interpretation of String Theory," *Nucl. Phys.* **B335** (1990) 138.

[238] K. Papadodimas and S. Raju, "An Infalling Observer in AdS/CFT," *JHEP* **10** (2013) 212, arXiv:1211.6767 [hep-th].

[239] K. Papadodimas and S. Raju, "Black Hole Interior in the Holographic Correspondence and the Information Paradox," *Phys. Rev. Lett.* **112** (2014) 051301, arXiv:1310.6334 [hep-th].

[240] K. Papadodimas and S. Raju, "State-Dependent Bulk-Boundary Maps and Black Hole Complementarity," *Phys. Rev.* **D89** (2014) 086010, arXiv:1310.6335 [hep-th].

[241] K. Papadodimas and S. Raju, "Local Operators in the Eternal Black Hole," *Phys. Rev. Lett.* **115** (2015) 211601, arXiv:1502.06692 [hep-th].

[242] K. Papadodimas and S. Raju, "Remarks on the Necessity and Implications of State-Dependence in the Black Hole Interior," *Phys. Rev.* **D93** (2016) 084049, arXiv:1503.08825 [hep-th].

[243] R. Bousso, "Firewalls from Double Purity," *Phys. Rev.* **D88** (2013) 084035, arXiv:1308.2665 [hep-th].

[244] D. Marolf and J. Polchinski, "Violations of the Born Rule in Cool State-Dependent Horizons," *JHEP* **01** (2016) 008, arXiv:1506.01337 [hep-th].

[245] J. C. Baez and J. Vicary, "Wormholes and Entanglement," *Class. Quant. Grav.* **31** (2014) 214007, arXiv:1401.3416 [gr-qc].

[246] H. Gharibyan and R. F. Penna, "Are Entangled Particles Connected by Wormholes? Evidence for the ER=EPR Conjecture from Entropy Inequalities," *Phys. Rev.* **D89** (2014) 066001, arXiv:1308.0289 [hep-th].

[247] G. Mandal, R. Sinha, and N. Sorokhaibam, "The Inside Outs of AdS_3/CFT_2: Exact AdS Wormholes with Entangled CFT Duals," *JHEP* **01** (2015) 036, arXiv:1405.6695 [hep-th].

[248] M. A. Nielsen and I. L. Chuang, *Quantum Computation and Quantum Information*. Cambridge University Press, 2010.

[249] J. Griffiths and J. Podolský, *Exact Space-Times in Einstein's General Relativity*. Cambridge University Press, 2009.

[250] K. Lake and R. C. Roeder, "Effects of a Nonvanishing Cosmological Constant on the Spherically Symmetric Vacuum Manifold," *Phys. Rev.* **D15** (1977) 3513.

[251] S. Hemming and E. Keski-Vakkuri, "Hawking Radiation from AdS Black Holes," *Phys. Rev.* **D64** (2001) 044006, arXiv:gr-qc/0005115 [gr-qc].

[252] G. T. Horowitz and V. E. Hubeny, "Quasinormal Modes of AdS Black Holes and the Approach to Thermal Equilibrium," *Phys. Rev.* **D62** (2000) 024027, arXiv:hep-th/9909056 [hep-th].

[253] L. Fidkowski, V. Hubeny, M. Kleban, and S. Shenker, "The Black Hole Singularity in AdS/CFT," *JHEP* **02** (2004) 014, arXiv:hep-th/0306170 [hep-th].

[254] V. Coffman, J. Kundu, and W. K. Wootters, "Distributed Entanglement," *Phys. Rev.* **A61** (2000) 052306, arXiv:quant-ph/9907047 [quant-ph].

[255] E. H. Lieb and M. B. Ruskai, "Proof of the Strong Subadditivity of Quantum-Mechanical Entropy," *J. Math. Phys.* **14** (1973) 1938.

[256] C. Holzhey, F. Larsen, and F. Wilczek, "Geometric and Renormalized Entropy in Conformal Field Theory," *Nucl.Phys.* **B424** (1994) 443, arXiv:hep-th/9403108 [hep-th].

[257] P. Calabrese and J. L. Cardy, "Entanglement Entropy and Quantum Field Theory," *J. Stat. Mech.* **0406** (2004) P06002, arXiv:hep-th/0405152 [hep-th].

[258] R. Bousso, H. Casini, Z. Fisher, and J. Maldacena, "Entropy on a Null Surface for Interacting Quantum Field Theories and the Bousso Bound," *Phys. Rev.* **D91** (2015) 084030, arXiv:1406.4545 [hep-th].

[259] P. Hayden, M. Headrick, and A. Maloney, "Holographic Mutual Information is Monogamous," *Phys. Rev.* **D87** (2013) 046003, arXiv:1107.2940 [hep-th].

[260] N. Bao, S. Nezami, H. Ooguri, B. Stoica, J. Sully, and M. Walter, "The Holographic Entropy Cone," *JHEP* **09** (2015) 130, arXiv:1505.07839 [hep-th].

[261] M. Srednicki, "Entropy and Area," *Phys.Rev.Lett.* **71** (1993) 666, arXiv:hep-th/9303048 [hep-th].

[262] L. Susskind, "Copenhagen vs Everett, Teleportation, and ER=EPR," *Fortsch. Phys.* **64** (2016) 551, arXiv:1604.02589 [hep-th].

[263] S. W. Hawking, "Black Holes in General Relativity," *Commun. Math. Phys.* **25** (1972) 152.

[264] A. Królak, "Definitions of Black Holes without Use of the Boundary at Infinity," *General Relativity and Gravitation* **14** (1982) 793.

[265] R. Bousso, H. Casini, Z. Fisher, and J. Maldacena, "Proof of a Quantum Bousso Bound," *Phys. Rev.* **D90** (2014) 044002, arXiv:1404.5635 [hep-th].

[266] R. Bousso and N. Engelhardt, "New Area Law in General Relativity," *Phys. Rev. Lett.* **115** (2015) 081301, arXiv:1504.07627 [hep-th].

[267] V. E. Hubeny, M. Rangamani, and T. Takayanagi, "A Covariant Holographic Entanglement Entropy Proposal," *JHEP* **07** (2007) 062, arXiv:0705.0016 [hep-th].

[268] A. C. Wall, "Maximin Surfaces, and the Strong Subadditivity of the Covariant Holographic Entanglement Entropy," *Class. Quant. Grav.* **31** (2014) 225007, arXiv:1211.3494 [hep-th].

[269] V. Balasubramanian, P. Hayden, A. Maloney, D. Marolf, and S. F. Ross, "Multiboundary Wormholes and Holographic Entanglement," *Class. Quant. Grav.* **31** (2014) 185015, arXiv:1406.2663 [hep-th].

[270] M. Headrick, V. E. Hubeny, A. Lawrence, and M. Rangamani, "Causality & Holographic Entanglement Entropy," *JHEP* **12** (2014) 162, arXiv:1408.6300 [hep-th].

[271] S. M. Carroll, *Spacetime and Geometry: An Introduction to General Relativity*. Addison Wesley, 2004.

[272] A. Lewkowycz and J. Maldacena, "Generalized Gravitational Entropy," *JHEP* **1308** (2013) 090, arXiv:1304.4926 [hep-th].

[273] R. Orús, "A Practical Introduction to Tensor Networks: Matrix Product States and Projected Entangled Pair States," *Ann.Phys.* **349** (2014) 117, arXiv:1306.2164 [cond-mat.str-el].

[274] G. Vidal, "Class of Quantum Many-Body States that Can be Efficiently Simulated," *Phys.Rev.Lett.* **101** (2008) 110501, quant-ph/0610099.

[275] B. Swingle, "Entanglement Renormalization and Holography," *Phys.Rev.* **D86** (2012) 065007, arXiv:0905.1317 [cond-mat.str-el].

[276] X.-L. Qi, "Exact Holographic Mapping and Emergent Space-Time Geometry," arXiv:1309.6282 [hep-th].

[277] 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 [hep-th].

[278] B. Czech, L. Lamprou, S. McCandlish, and J. Sully, "Integral Geometry and Holography," *JHEP* **10** (2015) 175, arXiv:1505.05515 [hep-th].

[279] C. Bény, "Causal Structure of the Entanglement Renormalization Ansatz," *New J. Phys.* **15** (2013) 023020, arXiv:1110.4872 [quant-ph].

[280] G. Evenbly and G. Vidal, "Algorithms for Entanglement Renormalization," *Phys.Rev.* **B79** (2009) 144108, arXiv:0707.1454 [cond-mat.str-el].

[281] R. N. C. Pfeifer, G. Evenbly, and G. Vidal, "Entanglement Renormalization, Scale Invariance, and Quantum Criticality," *Phys.Rev.* **A79** (2009) 040301, arXiv:0810.0580 [cond-mat.str-el].

[282] G. Evenbly, P. Corboz, and G. Vidal, "Nonlocal Scaling Operators with Entanglement Renormalization," *Phys.Rev.* **B82** (Oct., 2010) 132411, arXiv:0912.2166 [cond-mat.str-el].

[283] Y.-Y. Shi, L.-M. Duan, and G. Vidal, "Classical Simulation of Quantum Many-Body Systems with a Tree Tensor Network," *Phys.Rev.* **A74** (2006) 022320, quant-ph/0511070.

[284] P. Pfeuty, "The One-Dimensional Ising Model with a Transverse Field," *Ann.Phys.* **57** (1970) 79.

[285] T. Hartman and J. Maldacena, "Time Evolution of Entanglement Entropy from Black Hole Interiors," *JHEP* **1305** (2013) 014, arXiv:1303.1080 [hep-th].

[286] G. Evenbly and G. Vidal, "Scaling of Entanglement Entropy in the (Branching) Multiscale Entanglement Renormalization Ansatz," *Phys.Rev.* **B89** (2014) 235113, arXiv:1310.8372 [quant-ph].

[287] J. Haegeman, T. J. Osborne, H. Verschelde, and F. Verstraete, "Entanglement Renormalization for Quantum Fields in Real Space," *Phys.Rev.Lett.* **110** (2013) 100402, arXiv:1102.5524 [hep-th].

[288] M. Nozaki, S. Ryu, and T. Takayanagi, "Holographic Geometry of Entanglement Renormalization in Quantum Field Theories," *JHEP* **1210** (2012) 193, arXiv:1208.3469 [hep-th].

[289] A. Mollabashi, M. Nozaki, S. Ryu, and T. Takayanagi, "Holographic Geometry of cMERA for Quantum Quenches and Finite Temperature," *JHEP* **1403** (2014) 098, arXiv:1311.6095 [hep-th].

[290] 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.

[291] S. Ryu and T. Takayanagi, "Aspects of Holographic Entanglement Entropy," *JHEP* **0608** (2006) 045, arXiv:hep-th/0605073 [hep-th].

[292] J. Bekenstein, "Black Holes and the Second Law," *Lett.Nuovo Cim.* **4** (1972) 737.

[293] M. Bañados, C. Teitelboim, and J. Zanelli, "Black Hole in Three-Dimensional Spacetime," *Phys.Rev.Lett.* **69** (1992) 1849, arXiv:hep-th/9204099 [hep-th].

[294] C. Martinez, C. Teitelboim, and J. Zanelli, "Charged Rotating Black Hole in Three Spacetime Dimensions," *Phys.Rev.* **D61** (2000) 104013, arXiv:hep-th/9912259 [hep-th].

[295] G. Evenbly and G. Vidal, "Tensor Network States and Geometry," *J.Stat.Phys.* **145** (2011) 891, arXiv:1106.1082 [quant-ph].

[296] A. Almheiri, X. Dong, and D. Harlow, "Bulk Locality and Quantum Error Correction in AdS/CFT," *JHEP* **04** (2015) 163, arXiv:1411.7041 [hep-th].

[297] G. Evenbly and G. Vidal, "Tensor Network Renormalization Yields the Multiscale Entanglement Renormalization Ansatz," *Phys.Rev.Lett.* **115** (2015) 200401, 1502.05385 [cond-mat.str-el].

[298] J. Molina-Vilaplana and J. Prior, "Entanglement, Tensor Networks and Black Hole Horizons," *Gen.Rel.Grav.* **46** (2014) 1823, arXiv:1403.5395 [hep-th].

[299] S. Hawking, "Black Holes and Thermodynamics," *Phys.Rev.* **D13** (1976) 191.

[300] N. Lashkari, M. B. McDermott, and M. Van Raamsdonk, "Gravitational Dynamics from Entanglement 'Thermodynamics'," *JHEP* **04** (2014) 195, arXiv:1308.3716 [hep-th].

[301] B. Swingle and M. Van Raamsdonk, "Universality of Gravity from Entanglement," arXiv:1405.2933 [hep-th].

[302] E. P. Verlinde, "On the Origin of Gravity and the Laws of Newton," *JHEP* **1104** (2011) 029, arXiv:1001.0785 [hep-th].

[303] T. Padmanabhan, "Thermodynamical Aspects of Gravity: New Insights," *Rept. Prog. Phys.* **73** (2010) 046901, arXiv:0911.5004 [gr-qc].

[304] Y. Tian and X.-N. Wu, "Thermodynamics of Black Holes from Equipartition of Energy and Holography," *Phys. Rev.* **D81** (2010) 104013, arXiv:1002.1275 [hep-th].

[305] R.-G. Cai and S. P. Kim, "First Law of Thermodynamics and Friedmann Equations of Friedmann-Robertson-Walker Universe," *JHEP* **02** (2005) 050, arXiv:hep-th/0501055 [hep-th].

[306] S. Hawking, "Particle Creation by Black Holes," *Commun.Math.Phys.* **43** (1975) 199.

[307] A. Strominger and C. Vafa, "Microscopic Origin of the Bekenstein-Hawking Entropy," *Phys. Lett.* **B379** (1996) 99, arXiv:hep-th/9601029 [hep-th].

[308] J. M. Maldacena, A. Strominger, and E. Witten, "Black Hole Entropy in M-Theory," *JHEP* **12** (1997) 002, arXiv:hep-th/9711053 [hep-th].

[309] G. W. Gibbons and S. W. Hawking, "Cosmological Event Horizons, Thermodynamics, and Particle Creation," *Phys. Rev.* **D15** (1977) 2738.

[310] W. Unruh, "Notes on Black Hole Evaporation," *Phys.Rev.* **D14** (1976) 870.

[311] J. Bisognano and E. Wichmann, "On the Duality Condition for Quantum Fields," *J.Math.Phys.* **17** (1976) 303.

[312] L. Bombelli, R. K. Koul, J. Lee, and R. D. Sorkin, "A Quantum Source of Entropy for Black Holes," *Phys.Rev.* **D34** (1986) 373.

[313] H. Casini, M. Huerta, and R. C. Myers, "Towards a Derivation of Holographic Entanglement Entropy," *JHEP* **05** (2011) 036, arXiv:1102.0440 [hep-th].

[314] D. D. Blanco, H. Casini, L.-Y. Hung, and R. C. Myers, "Relative Entropy and Holography," *JHEP* **08** (2013) 060, arXiv:1305.3182 [hep-th].

[315] V. Balasubramanian, M. B. McDermott, and M. Van Raamsdonk, "Momentum-Space Entanglement and Renormalization in Quantum Field Theory," *Phys. Rev.* **D86** (2012) 045014, arXiv:1108.3568 [hep-th].

[316] J. von Neumann, *Mathematical Foundations of Quantum Mechanics*. Princeton University Press, 1955.

[317] H. Casini, D. A. Galante, and R. C. Myers, "Comments on Jacobson's 'Entanglement Equilibrium and the Einstein Equation'," *JHEP* **03** (2016) 194, arXiv:1601.00528 [hep-th].

[318] C. G. Callan, Jr. and F. Wilczek, "On Geometric Entropy," *Phys. Lett.* **B333** (1994) 55, arXiv:hep-th/9401072 [hep-th].

[319] P. Calabrese and J. Cardy, "Entanglement Entropy and Conformal Field Theory," *J. Phys.* **A42** (2009) 504005, arXiv:0905.4013 [cond-mat.stat-mech].

[320] M. M. Anber and J. F. Donoghue, "On the Running of the Gravitational Constant," *Phys. Rev.* **D85** (2012) 104016, arXiv:1111.2875 [hep-th].

[321] C. Eling, R. Guedens, and T. Jacobson, "Nonequilibrium Thermodynamics of Spacetime," *Phys. Rev. Lett.* **96** (2006) 121301, arXiv:gr-qc/0602001 [gr-qc].

[322] Z.-K. Guo, Y.-S. Piao, R.-G. Cai, and Y.-Z. Zhang, "Inflationary Attractor from Tachyonic Matter," *Phys. Rev.* **D68** (2003) 043508, arXiv:hep-ph/0304236 [hep-ph].

[323] L. A. Ureña-López and M. J. Reyes-Ibarra, "On the Dynamics of a Quadratic Scalar Field Potential," *Int. J. Mod. Phys.* **D18** (2009) 621, arXiv:0709.3996 [astro-ph].

[324] V. A. Belinsky, L. P. Grishchuk, I. M. Khalatnikov, and Ya. B. Zeldovich, "Inflationary Stages in Cosmological Models with a Scalar Field," *Phys. Lett.* **B155** (1985) 232.

[325] T. Piran and R. M. Williams, "Inflation in Universes with a Massive Scalar Field," *Phys. Lett.* **B163** (1985) 331.

[326] B. Ratra and P. J. E. Peebles, "Cosmological Consequences of a Rolling Homogeneous Scalar Field," *Phys. Rev.* **D37** (1988) 3406.

[327] A. Liddle and D. Lyth, *Cosmological Inflation and Large-Scale Structure*. Cambridge University Press, 2000.

[328] A. R. Liddle, P. Parsons, and J. D. Barrow, "Formalizing the Slow-Roll Approximation in Inflation," *Phys. Rev.* **D50** (1994) 7222, arXiv:astro-ph/9408015 [astro-ph].

[329] V. Kiselev and S. Timofeev, "Quasiattractor Dynamics of $\lambda\phi^4$-Inflation," arXiv:0801.2453 [gr-qc].

[330] P. G. Ferreira and M. Joyce, "Cosmology with a Primordial Scaling Field," *Phys. Rev.* **D58** (1998) 023503, arXiv:astro-ph/9711102 [astro-ph].

[331] S. Downes, B. Dutta, and K. Sinha, "Attractors, Universality, and Inflation," *Phys. Rev.* **D86** (2012) 103509, arXiv:1203.6892 [hep-th].

[332] J. Khoury and P. J. Steinhardt, "Generating Scale-Invariant Perturbations from Rapidly-Evolving Equation of State," *Phys. Rev.* **D83** (2011) 123502, arXiv:1101.3548 [hep-th].

[333] S. Clesse, C. Ringeval, and J. Rocher, "Fractal Initial Conditions and Natural Parameter Values in Hybrid Inflation," *Phys. Rev.* **D80** (2009) 123534, arXiv:0909.0402 [astro-ph.CO].

[334] J. Milnor, "On the Concept of Attractor," *Commun. Math. Phys.* **99** (1985) 177.

[335] J. Auslander, N. Bhatia, and P. Seibert, "Attractors in Dynamical Systems," *Bol. Soc. Mat. Mexicana (2)* **9** (1964) 55.

[336] J. Gibbs, *Elementary Principles in Statistical Mechanics*. Yale University Press, 1902.

[337] C.-Y. Tseng, *Deviation from Standard Inflationary Cosmology and Problems in Ekpyrosis*. PhD thesis, California Institute of Technology, 2013.

[338] R. Wald, *General Relativity*. University of Chicago Press, 1984.

[339] S. M. Carroll and H. Tam, "Unitary Evolution and Cosmological Fine-Tuning," arXiv:1007.1417 [hep-th].

[340] G. Gibbons and N. Turok, "The Measure Problem in Cosmology," *Phys.Rev.* **D77** (2008) 063516, arXiv:hep-th/0609095 [hep-th].

[341] J. S. Schiffrin and R. M. Wald, "Measure and Probability in Cosmology," *Phys. Rev.* **D86** (2012) 023521, arXiv:1202.1818 [gr-qc].

[342] S. Hawking and D. N. Page, "How Probable is Inflation?," *Nucl.Phys.* **B298** (1988) 789.

[343] J. Douglas, "Solution of the Inverse Problem of the Calculus of Variations," *Trans. Amer. Math. Soc.* **50** (1941) 71.

[344] R. Santilli, *Foundations of Theoretical Mechanics I: The Inverse Problem in Newtonian Mechanics*. Springer-Verlag, 1978.

[345] S. Kowalevski, "Zur Theorie der Partiellen Differentialgleichung," *J. Reine Angew. Math.* **80** (1875) 1.

[346] A. Lyapunov, *The General Problem of the Stability of Motion*. Control Theory and Applications Series. Taylor & Francis, 1992, orig. 1892. translated by A.T. Fuller.

[347] **BICEP2** Collaboration, P. Ade et al., "Detection of B-Mode Polarization at Degree Angular Scales by BICEP2," *Phys.Rev.Lett.* **112** (2014) 241101, arXiv:1403.3985 [astro-ph.CO].

[348] **BICEP2 and Planck** Collaboration, P. A. R. Ade et al., "Joint Analysis of

BICEP2/Keck Array and Planck Data," *Phys. Rev. Lett.* **114** (2015) 101301, arXiv:1502.00612 [astro-ph.CO].

[349] D. Baumann, "Inflation," in *Physics of the Large and the Small: TASI 09, Proceedings of the Theoretical Advanced Study Institute in Elementary Particle Physics, Boulder, Colorado, USA, 1-26 June 2009*, p. 523. 2011. arXiv:0907.5424 [hep-th].

[350] K. Freese and W. H. Kinney, "Natural Inflation: Consistency with Cosmic Microwave Background Observations of Planck and BICEP2," *JCAP* **1503** (2015) 044, arXiv:1403.5277 [astro-ph.CO].

[351] R. Penrose, "Difficulties with Inflationary Cosmology," *Annals N.Y.Acad.Sci.* **571** (1989) 249.

[352] A. D. Linde, "Recent Progress in the Inflationary Universe Scenario," *Nucl.Phys.* **B252** (1985) 153.

[353] M. Madsen and P. Coles, "Chaotic Inflation," *Nucl.Phys.* **B298** (1988) 701.

[354] M. Kawasaki, M. Yamaguchi, and T. Yanagida, "Natural Chaotic Inflation in Supergravity," *Phys.Rev.Lett.* **85** (2000) 3572, arXiv:hep-ph/0004243 [hep-ph].

[355] J. Ellis, M. Fairbairn, and M. Sueiro, "Rescuing Quadratic Inflation," *JCAP* **1402** (2014) 044, arXiv:1312.1353 [astro-ph.CO].

[356] F. Marchesano, G. Shiu, and A. M. Uranga, "F-term Axion Monodromy Inflation," *JHEP* **09** (2014) 184, arXiv:1404.3040 [hep-th].

[357] F. C. Adams, J. R. Bond, K. Freese, J. A. Frieman, and A. V. Olinto, "Natural Inflation: Particle Physics Models, Power-Law Spectra for Large-Scale Structure, and Constraints from the Cosmic Background Explorer," *Phys.Rev.* **D47** (1993) 426, arXiv:hep-ph/9207245 [hep-ph].

[358] K. Freese, J. A. Frieman, and A. V. Olinto, "Natural Inflation with Pseudo Nambu-Goldstone Bosons," *Phys.Rev.Lett.* **65** (1990) 3233.

[359] N. Arkani-Hamed, H.-C. Cheng, P. Creminelli, and L. Randall, "Extranatural Inflation," *Phys.Rev.Lett.* **90** (2003) 221302, arXiv:hep-th/0301218 [hep-th].

[360] C. Germani and A. Kehagias, "UV-Protected Inflation," *Phys.Rev.Lett.* **106** (2011) 161302, arXiv:1012.0853 [hep-ph].

[361] R. Kallosh, A. Linde, and B. Vercnocke, "Natural Inflation in Supergravity and Beyond," *Phys. Rev.* **D90** (2014) 041303, arXiv:1404.6244 [hep-th].