From Quantum Information to Cosmic Censorship: Emergent Spacetimes and Their Surfaces

by

Åsmund Schiager Folkestad

M.Sc. Physics, NTNU, 2018

Submitted to the Department of Physics
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

May 2024

Authored by:	Åsmund Schiager Folkestad Department of Physics May 17, 2024
Certified by:	Netta Engelhardt Professor of Physics, Thesis Supervisor
Accepted by:	Lindley Winslow Professor of Physics Associate Department Head, Department of Physics

From Quantum Information to Cosmic Censorship: Emergent Spacetimes and Their Surfaces

by

Åsmund Schiager Folkestad

Submitted to the Department of Physics
on May 17, 2024 in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

ABSTRACT

In this thesis, we explore classical and semiclassical gravity from the perspective of the AdS/CFT correspondence. We leverage global methods in General Relativity (GR) together with quantum information– and complexity–theoretic properties of the conformal field theory (CFT) dual to obtain novel results in classical and semiclassical gravity.

In the first part, we obtain a collection of results suggesting that holography enforces a refined version of Cosmic Censorship that potentially can replace the Weak Cosmic Censorship (WCC) conjecture, which has been disproven in Anti-de Sitter (AdS) spacetimes. We show that certain important GR results usually proven assuming WCC can instead be derived from consistency of the AdS/CFT dictionary. We also construct new likely violations of WCC in asymptotically AdS_4 spacetimes, but show that these cannot have a holographic dual; this provides evidence that singularities are better behaved in holographic theories, compared to GR with generic matter. Finally, we show a connection between event horizons and CFT pseudorandomness, and we construct a new measure of the size of a naked singularity. We conjecture that quantum gravity only forbids macroscopic naked singularities, according to this measure.

In the second part, we derive new properties of various extremal submanifolds, with several consequences for AdS/CFT. For example, we provide a physically intuitive explanation for why extremal surfaces are natural boundaries between independent subsystems. We also prove results that constrain far-from-equilibrium dynamics in gravity and CFTs. Finally, we construct a puzzle showing that geometric states with large entanglement need not correspond to a wormhole, highlighting subtleties in the ER=EPR proposal.

Thesis supervisor: Netta Engelhardt
Title: Professor of Physics

Acknowledgments

First, I want to thank my advisor, Professor Netta Engelhardt, for her continuous support, collaboration, and insights. I am deeply grateful for her unflinching encouragement and the many invaluable things she has taught me along the way.

Next, I want to thank the many great friends, collaborators, and colleagues in Cambridge and the Centre for Theoretical Physics. I am especially grateful for the support from Professors Daniel Harlow, Hong Liu, and Krishna Rajagopal, and for the friendship and interesting discussions with Pierre Barral, Sam Leutheusser, Shreya Vardhan, Eric Anscheutz, Dimitra Pefkou, Atakan Hilmi Firat, Sam Alipour-fard, Nico Valdes-Meller, Elba Alonso-Monsalve, and Ana Trisović. I am also thankful for the support provided by the Aker Scholarship and the CTP administrators, Scott Morley and Charles Suggs.

Finally, I want to express my gratitude to my family for their continuous support in my journey to do physics, and to my fiancée Elisa for her love, friendship, and encouragement.

Publication List

This thesis presents the following publications.

References

[1] N. Engelhardt, Å. Folkestad, A. Levine, E. Verheijden, and L. Yang, *Cryptographic Censorship*, arXiv:2402.03425.

[2] Å. Folkestad, *Subregion Independence in Gravity*, arXiv:2311.09403.

[3] Å. Folkestad and A. Dhumuntarao, *Maximal entangling rates from holography*, *Phys. Rev. D* **108** (2023), no. 8 086032, [arXiv:2211.07654].

[4] Å. Folkestad, *Penrose Inequality as a Constraint on the Low Energy Limit of Quantum Gravity*, *Phys. Rev. Lett.* **130** (2023), no. 12 121501, [arXiv:2209.00013].

[5] N. Engelhardt and Å. Folkestad, *Canonical purification of evaporating black holes*, *Phys. Rev. D* **105** (2022), no. 8 086010, [arXiv:2201.08395].

[6] N. Engelhardt and Å. Folkestad, *Negative complexity of formation: the compact dimensions strike back*, *JHEP* **07** (2022) 031, [arXiv:2111.14897].

[7] N. Engelhardt and Å. Folkestad, *General bounds on holographic complexity*, *JHEP* **01** (2022) 040, [arXiv:2109.06883].

[8] Å. Folkestad and S. Hernández-Cuenca, *Conformal Rigidity from Focusing*, *Classical and Quantum Gravity* (2021) [arXiv:2106.09037].

[9] N. Engelhardt and Å. Folkestad, *Holography abhors visible trapped surfaces*, *JHEP* **07** (2021) 066, [arXiv:2012.11445].

Contents

List of Figures

List of Tables

Chapter 1

Introduction

Before the discovery of General Relativity (GR), space and time constituted a fixed stage upon which physics played out. Together, they formed a static background to which we could anchor our description of observable quantities; the background provided fixed clocks and rods allowing universal measurements comparable across different configurations of matter. However, with the advent of Einstein's theory of gravity, the picture radically changed. The clean separation between matter and the stage it acts on was broken. Spacetime is a curved geometry that influences the behavior of matter, while matter dynamically alters the geometry of spacetime. Crucially, matter appearing within spacetime no longer constitutes degrees of freedom independent from spacetime itself. This deep fact leads us to the following question: what are the *fundamental* degrees of freedom when gravity is taken into account?

It is tempting to propose that the fundamental degrees of freedom include the metric at each point p in spacetime. But this view is untenable. If a degree of freedom is deemed fundamental, we should at least be allowed to ask about its state for every possible configuration of the universe, and we should have a definite answer independent of arbitrary convention. The spacetime metric g_{ab} at p satisfies none of these criteria; its value depends on the particular choice of coordinates, and in GR, coordinates on their own hold no physical significance – the theory is diffeomorphism invariant. This particular issue is readily dealt with, however; there exist curvature invariants that are scalars, such as the squared Riemann tensor $R_{abcd}R^{abcd}$, so perhaps these constitute fundamental degrees of freedom? But a deeper issue remains: how to identify the point p in a way that does not make reference to any particular spacetime and particular choice of coordinates?

One partial answer to this question is known for asymptotically flat (AF) and asymptotically AdS (AAdS) spacetimes, which correspond to universes with a bounded amount of matter and a vanishing or negative cosmological constant Λ, respectively. In these classes of spacetimes, the structure at conformal infinity $\mathscr{I}$ is fixed, providing a background structure that we can anchor observables to. We can unambiguously identify points q in $\mathscr{I}$, regardless of what happens in the bulk of the spacetime. This enables us to identify bulk spacetime points p in a diffeomorphism invariant way using relational prescriptions like the following: "fire a spacelike geodesic from q in a given direction, and follow the geodesic for a renormalized proper distance s; the resulting event is p.". We can now ask about the value of $R_{abcd}R^{abcd}$ at p, and the final result is diffeomorphism invariant. With this type of procedure, we might hope to save the idea that the fundamental degrees of freedom include the spacetime geometry,

albeit in an indirect way where everything is anchored to $\mathscr{I}$. However, black holes and the requirement that gravity should be quantum mechanical kills this hope once and for all: spacetime as a fundamental concept appears to be doomed. Nevertheless, $\mathscr{I}$ survives, and its importance only increases when we quantize gravity.[1]

When a large amount of energy concentrates in a sufficiently small region, the conventional wisdom says that we tend to form black holes. A remarkable property of standard black hole solutions is that they contain a singularity – a location where time (or spacetime) itself can end. Specifically, geodesics will terminate at a singularity and cannot be continued further, so particles travelling on geodesics appear to get frozen in time when they hit the singularity. This is a hint that spacetime as a fundamental concept breaks down. Another notable consequence of singularities is that the diffeomorphism invariant event that we defined above no longer exist in general. Their existence becomes state dependent – it might not be possible to travel a given renormalized proper distance along some geodesic fired from $q \in \mathscr{I}$, since a singularity could terminate this geodesic preemptively. Thus, already in classical GR, there appears to be no robust state-independent way to talk about a given point in spacetime, unless this point is at $\mathscr{I}$ itself. Attempts to reconcile quantum mechanics with GR makes the situation worse. In units of $c = 1$, the Heisenberg uncertainty principle implies that we need wave packets with energies of order $E \gtrsim \hbar/\Delta x$ to probe distances of order Δx. But once Δx is on the order of the Planck length $\ell_{\rm Pl} \equiv \sqrt{\hbar G_N}$, the size of the wave packet is comparable to its would-be Schwarzschild radius $2G_N E$, leading to the likely formation of a black hole.[2] This raises serious doubts about the operational meaning of distances, and thus of spacetime geometry itself, at length scales at or below the Planck scale. The naive quantization of GR leads to the same conclusion. The effective coupling constant of this theory is $(E\ell_{\rm Pl}/\hbar)^2$, so the quantum theory becomes strongly coupled at Planckian energies. Additionally, the theory is non-renormalizable, meaning that an infinite series of counterterms become important in this regime. Since each is associated to a free parameter, we get a complete breakdown of the theory's predictability. The same thing happens as we approach curvature singularities, since the blowup of invariants like $R_{abcd}R^{abcd}$ signals that UV-sensitive higher derivative corrections to the GR Lagrangian become important.

While black holes and quantum mechanics spell trouble for spacetime as a fundamental concept, together they also point towards the solution. The remarkable discovery that black holes are thermal systems [1–3] with an entropy given by

$$S_{\rm BH} = \frac{\text{Area[event horizon]}}{4G_N\hbar} \tag{1.1}$$

shows that the fundamental degrees of freedom of gravity are radically different from ordinary non-gravitational quantum field theory (QFT). Thermal states in gravity have entropy proportional to the surface area of the relevant region, rather than the volume scaling of standard non-gravitational QFT, such as quantum electrodynamics. This suggests that the fundamental degrees of freedom of a region of spacetime is somehow associated with its

[1]We do not claim that gravity in principle cannot be quantized in universes with no conformal boundary. However, quantum gravity in these universes is at present poorly understood.

[2]This argument assumes that the wave packet does not have a parametrically asymmetric spatial profile, like a highly elongated ellipse. However, in this case, there is no sense in which the wave packet is close to probing a local region in spacetime.

boundary. This is the idea of the holographic principle [4, 5]. Note however that the boundary of all of spacetime is in a certain sense $\mathscr{I}$. Thus the Bekenstein-Hawking entropy (1.1), together with the fact that conformal infinity is the only spacetime structure with a robust state-independent existence, hints that the degrees of freedom of quantum gravity could be localized at $\mathscr{I}$ – at least in cases where $\mathscr{I}$ is non-empty. Further evidence in favor of it comes from the fact that the Hamiltonian of GR is a boundary term located at conformal infinity.

The 1997 discovery of the celebrated AdS/CFT correspondence [6–8] gave a precise realization of the holographic principle in the case of a negative cosmological constant. Not only are the fundamental degrees of freedom localized at conformal infinity, as hinted to by GR itself, but remarkably, they form a relativistic non-gravitational QFT that is local, unitary, and conformally invariant – a so-called holographic conformal field theory (CFT). Furthermore, the gravitational bulk is a string theory, which reduces to GR coupled to certain matter fields in the limit of low curvature and weakly coupled strings. While many aspects of string theory are quite well understood at the level of perturbation theory around certain spacetimes, significant difficulties remain when trying to understand the theory non-perturbatively. One of the most significant facts about the AdS/CFT correspondence is that it provides a completely non-perturbative UV-complete definition of quantum gravity via a type of theory we understand relatively well. The price we pay for a non-perturbative description is that spacetime is encoded in the holographic CFT in an indirect and often hard-to-access way; the description brings the emergent nature of spacetime to the front. However, over the last 20 years, tools from quantum information theory have revolutionized our understanding of how spacetime emerges from the holographic CFT, leading to remarkable new insights into quantum gravity, including in the regime where classical or semiclassical GR is a good description.

A central theme of this thesis is the use of perspectives and tools made available by AdS/CFT to obtain new insights and constraints on gravitational theories that arise as the classical or semiclassical limits of holography. Another theme is using AdS/CFT to learn new lessons about CFTs in regimes where direct CFT calculations are challenging. Examples of questions to which this thesis provides new insights, answers, or partial answers are the following:

- Chapters 2–4: Does holography enforce some notion of Cosmic Censorship?
- Chapter 3: Does the existence of a holographic CFT impose quantitative bounds on the coupling constants of matter?
- Chapter 5: How can we reconstruct the spacetime geometry directly from CFT data?
- Chapter 6: Do the surfaces that compute entropies in holography serve a functional role already at the level of classical GR?
- Chapter 7: For entangled CFTs, how can we diagnose from the CFT whether the emergent geometry is a wormhole rather than multiple disconnected spacetimes?
- Chapter 8: Is there a general limit to how fast the volume of black hole interiors can grow?

- Chapter 9: Does the (renormalized) volume of AAdS spacetimes have a lower bound?
- Chapter 10: How can AdS/CFT be used to derive bounds on far-from-equilibrium dynamics in QFT many-body states?

An underlying theme unifying all these questions is that they can be understood to be intimately tied to the dynamics of spacetime submanifolds with certain extremality properties.

Before turning to this work, we present a short review of selected aspects AdS/CFT and Weak Cosmic Censorship, contextualized by the problems solved in this thesis.

1.1 The AdS/CFT Correspondence

AdS/CFT [6–8] refers to a conjectured duality between certain CFTs and string theories with a negative cosmological constant. Duality here means that the theories are equivalent, in the sense that the Hilbert spaces and gauge-invariant observables are in one-to-one correspondence, so that physical observables in one description can also be computed in the other.[3] The best known example of AdS/CFT is the duality between Type IIB string theory on asymptotically $\mathrm{AdS}_5 \times S^5$ spacetimes and four-dimensional Yang-Mills gauge theory with gauge group $SU(N)$ and $\mathcal{N} = 4$ supersymmetry. In this case, N is related to the ratio of the Planck scale to the AdS length scale $L \propto 1/\sqrt{-\Lambda}$ as

$$\left(\frac{\ell_{\mathrm{Pl}}}{L}\right)^8 \propto \frac{1}{N^2}. \tag{1.2}$$

The effective length of strings ℓ_{s} is related to the gauge coupling g of the $\mathcal{N} = 4$ super Yang-Mills theory (SYM) by

$$\left(\frac{\ell_s}{L}\right)^4 \propto \frac{1}{g^2 N}, \tag{1.3}$$

where the combination $\lambda \equiv g^2 N$ is known as the 't Hooft coupling. Thus, N controls the importance of bulk quantum effects, while λ controls the importance of higher-curvature corrections to GR caused by the extended nature of strings.

In the $N, \lambda \to \infty$ limit, $\mathcal{N} = 4$ SYM becomes a strongly coupled gauge theory with a large gauge group; the bulk approaches classical GR coupled to matter fields, together with a free QFT living on top of the background provided by the classical sector. The matter fields in the bulk are in one-to-one correspondence with a privileged family of local CFT operators, so-called single-trace conformal primary operators. At strict $N = \infty$, the classical sector of the theory determines the one-point functions of these operators, while their connected two-point functions are determined by the QFT sector. Including perturbative corrections in $1/N$ corresponds to including bulk interactions and loops, and it makes connected n-point functions on the boundary for $n \geq 3$ non-zero.[4] In general, even when we are not in Type IIB string theory, the limit where we have (semi)classical GR in the bulk corresponds to a strongly coupled boundary theory in a large$-N$ limit, where N counts the effective number

[3] At least in the regime where both theories are well defined. If no alternative non-perturbative description of string theory exists, then the CFT simply defines the gravity theory in the non-perturbative regime.

[4] In known examples of AdS/CFT, the coupling constants of matter are of gravitational strength, scaling with $G_N^{a>0}$, so G_N simultaneously controls the size of tree level matter couplings and the importance of loops.

of local CFT degrees of freedom. In this thesis, we always work to leading order in $\lambda = \infty$, and we work to either leading or subleading order in $1/N$. We will not rely on any of the particulars of $\mathcal{N} = 4$ SYM, but for a review of these details and AdS/CFT more generally, see [9, 10].

The vacuum state of a holographic CFT in d spacetime dimensions is dual to pure AdS_{d+1}, after Kaluza-Klein reduction of any potential compact bulk dimensions (such as the S^5 in Type IIB string theory). Its metric has a coordinate representation

$$\mathrm{d}s^2 = -\left(1 + \frac{r^2}{L^2}\right)\mathrm{d}t^2 + \frac{\mathrm{d}r^2}{1 + \frac{r^2}{L^2}} + r^2\mathrm{d}\Omega^2, \tag{1.4}$$

where $\mathrm{d}\Omega^2$ is the metric of the round $(d-1)$–sphere. Conformal infinity $\mathscr{I}$ lies at $r = \infty$ and has topology $\mathbb{R} \times S^{d-1}$. A representative of the conformal class of metrics at conformal infinity is the Einstein static universe: $-\mathrm{d}t^2 + L^2\mathrm{d}\Omega^2$. This is a Lorentzian manifold, and the holographic CFT can be taken to live on this spacetime, or any spacetime conformally equivalent to it. As we change the CFT state, the spacetime generally changes, but the metric still approaches (1.4) as $r \to \infty$, preserving the structure of $\mathscr{I}$. This is a consequence of the IR-UV correspondence in AdS/CFT [11]. Namely, phenomena occurring at large radii in the bulk correspond to phenomena happening at high energies in the CFT.[5] But since finite energy CFT states resemble the vacuum at sufficiently small length scales, their spacetime duals should look the same at large r.

1.1.1 Local observables

How are observables in the CFT and gravity related? The easiest case to determine are the expectation values of the single trace primary operators. For example, working in an $1/r$ expansion about $r = \infty$, the first state-dependent term in the metric expansion is proportional to the expectation value of CFT stress tensor [12]. For matter fields, the coefficient of the leading term in the $1/r$ expansion computes the expectation value of the corresponding primary operator [8]. For example, a minimally coupled scalar field with mass m^2 will usually have a leading falloff at large r given by[6]

$$\langle\phi(r,x)\rangle = \frac{\alpha(x)}{r^\Delta} + \ldots, \qquad \Delta = \left(\frac{d}{2}\right)^2 + \sqrt{\left(\frac{d}{2}\right)^2 + m^2L^2}, \tag{1.5}$$

where x serve as coordinates on $\mathscr{I}$; in (1.4) x would label t and the angular coordinates Ω^i. We have that $\alpha(x) \propto \langle O(x)\rangle$, where $O(x)$ is a CFT scalar single trace primary operator with scaling dimension Δ. For quantum fields deeper in the bulk (i.e. at finite r), but not so deep that they are behind event horizons, the relation to boundary operators is non-local, but still

[5] The scale-radius duality is a heuristic rather than a precise statement, and while it is a good rule of thumb, there are a few minor subtleties. For example, a massless particle can reach the conformal boundary and bounce off it in finite time. However, these massless fields do not backreact to change the geometry at $\mathscr{I}$.

[6] There are other possible falloffs possible at special values of m^2. In addition, we can turn on sources in the CFT, in which case the leading coefficient is the source.

relatively well understood. The relationship reads [13–15][7]

$$\phi(y) = \int \mathrm{d}^d x f(y,x) O(x) + \mathcal{O}\left(1/N\right) \tag{1.6}$$

where y is a bulk point, x a boundary point, and $f(y,x)$ is some non-unique smearing kernel that is determined by the bulk geometry. In examples where $f(y,x)$ has been determined, it is obtained as a Green's function of the free wave equation on the relevant geometry, albeit with unusual boundary conditions.

If we we only have access to the CFT state $|\psi\rangle$, how would we actually proceed to determine $f(y,x)$ in practice? This question is intimately related to the problem determining the geometry that is dual to $|\psi\rangle$ from basic CFT measurements. If we want to simulate holographic CFTs on a quantum computer in the future and then try to image the dual bulk, perhaps to verify the holographic nature of our simulated system, we ought to solve this problem. A partial answer was provided by [16, 17], which showed that measuring $(d+3)$-point correlators of local CFT operators could be used to reconstruct the metric in a large portion of the spacetime lying outside event horizons, up to an undetermined conformal factor. One contribution of this thesis is to show, in Chapter 5, how to determine the missing conformal factor in the special case where the spacetime is a vacuum spacetime, i.e. where the bulk matter is turned off.

1.1.2 Negative energy densities

A notable fact about AdS is that the scalar squared mass m^2 need not to be positive, as long as it satisfies the Breitenlohner-Freedman bound [18, 19],

$$m^2 L^2 \geq -\left(\frac{d}{2}\right)^2, \tag{1.7}$$

so that the scaling dimension in (1.5) is real. Since the potential of the scalar field has a leading contribution $m^2\phi^2/2$ at small ϕ, having $m^2 < 0$ leads to negative energy densities. In AAdS spacetimes, this need not destabilize the theory. Furthermore, it turns out that $m^2 < 0$ is not an exotic phenomenon in AdS/CFT; whenever the CFT has a single trace scalar operator that is relevant, i.e. $\Delta < d$, these scalar fields are present. Explicit examples from string theory in fact give rise not only to $m^2 < 0$, but also to lower unbounded scalar potentials [20–22]. The fact that we must live with arbitrarily negative energy densities in AdS causes interesting effects that we study in Chapters 3 and 9. For example, in Chapter 3 we show that while energy densities can be lower unbounded, there are still constraints on the legal forms of scalar potentials. We show that holography can be used to derive novel bounds similar to (1.7), but for coupling constants of scalar self-interactions rather than mass terms. Then, in Chapter 9 we show that negative energy densities cause the (renormalized) volume of volume-maximizing Cauchy slices to be lower-unbounded. The latter has implications for a conjectured duality [23] between volumes and certain complexity-theoretic quantities in the CFT, introduced below.

[7] The $\mathcal{O}(1/N)$ correction ensures that this operator is perturbatively diffeomorphism invariant.

1.1.3 Entanglement entropy and ER=EPR

One of the most important insights into AdS/CFT comes from the quantum extremal surface (QES) prescription [24–30]. To explain it, consider a spatial region R that is an open subset of some Cauchy slice A of $\mathscr{I}$. We can then consider the reduced density matrix $\rho_R = \mathrm{Tr}_{\bar{R}} |\psi\rangle \langle\psi|$ of the CFT state $|\psi\rangle$, where $\bar{R}$ is the complement of R within A. Assuming we are in the semiclassical regime of AdS/CFT, the QES prescription then says that the von Neumann entropy of ρ_R can be computed as

$$S_{\mathrm{vN}}[\rho_R] = \frac{\mathrm{Area}[X_R]}{4G_N\hbar} + S_{\mathrm{vN}}[\rho_r]. \tag{1.8}$$

Here X_R is a special codimension-2 spacelike surface that is known as a QES; ρ_r is the state of the bulk quantum fields on any choice of spacelike hypersurface Σ_R that is bounded by X_R in the bulk and by R at conformal infinity. X_R satisfies three conditions:

1. X_R is homologous to R.
2. X_R extremizes the functional (1.8).
3. Of all X_R satisfying 1. and 2., X_R is the one giving smallest $S_{\mathrm{vN}}[\rho_R]$.

The homology condition means that there exists a hypersurface Σ_R whose boundary is identically X_R and whose conformal boundary is identically R. It is worth noting that for a generic R, the surface X_R stretches out to $\mathscr{I}$, causing the area term in (1.8) to diverge. This can be dealt with using a cutoff on r at some large but finite value, which in the CFT is equivalent to imposing a UV cutoff. The divergence of S_{vN} for a general region R is caused by entanglement of short wavelength modes across ∂R in the CFT. In many cases, we can extract finite quantities from S_{vN} by doing a vacuum subtraction or computing certain derivatives.

The formula (1.8) is a remarkable generalization of the Bekenstein-Hawking entropy formula. In the special case where the CFT state is dual to a stationary two-sided black hole, and where R is a complete slice Cauchy slice of one of the two connected components of $\mathscr{I}$, the QES formula reduces to the Bekenstein-Hawking entropy [31]. But it also gives a new interpretation to this entropy: it is the entropy of entanglement between two CFTs, each associated to one of the two asymptotic regions. If we take R to be different region, or if we take the spacetime to be a more general one, the QES formula goes far beyond what is contained in Bekenstein-Hawking entropy. Notably, it was recently used to make remarkable progress on the black hole information problem [32, 33].

The QES formula shows that a large class of surfaces in general spacetimes have a fine-grained entropic interpretation in terms of the CFT degrees of freedom. This entropic interpretation will be central to many results obtained in this thesis. In Chapter 2 we use the invariance of the von Neumann entropy under local unitaries to show that holography in the classical limit imposes that trapped surfaces lie behind event horizons. This is an important consequence of the Weak Cosmic Censorship (WCC) conjecture (see discussion below), but crucially, we derive this result without assuming WCC. This is significant, since WCC in its usual formulation appears to be false in AdS [34, 35]. Next, to obtain our coupling constant bounds in Chapter 3, we are again reliant on (1.8). Then, in Chapter 10, we use the QES

formula to derive bounds on time- and space-derivatives of S_{vN} in large classes of CFT states out of equilibrium. These results can for example be used to lower bound thermalization times in far-from-equilibrium CFT states.

In the case of two entangled CFTs in the regime where the QES formula applies, (1.8) shows that if the state under consideration is dual to a wormhole, then the entanglement must be large. This is because a wormhole dictates that there is a non-trivial area term in (1.8), and thus there is an order $\mathcal{O}(G_N^{-1}) \sim \mathcal{O}(N^2)$ amount of entanglement between the CFTs.[8] But what about the converse? This question is intimately related to an idea that has been central in the study of the connection between entanglement and gravity: the idea of ER=EPR [36, 37]. This is a heuristic proposal stating a wormhole connecting two systems is a geometric manifestation of the entanglement between these two systems. The above observation shows that ER⇒EPR. But if we know that two CFTs have an $\mathcal{O}(N^2)$ amount of entanglement, and furthermore that they have a geometric bulk dual so that (1.8) is applicable, then must this geometry be a wormhole? This would be suggested by the most naive interpretation of ER=EPR. In Chapter 7 we show that this is false by constructing an explicit counterexample: it is possible to have a geometric CFT state with an $\mathcal{O}(N^2)$ amount of entanglement that is not geometrized by a wormhole. This has implications for possible sharp formulations of ER=EPR.

1.1.4 Entanglement wedges

A central concept in holography intimately tied to entanglement and QESs is that of entanglement wedges [27, 38, 39]. The entanglement wedge E_R is a bulk region associated to a boundary region R; it is defined as the domain of dependence of any of the homology hypersurfaces Σ_R described above. There is now substantial evidence [40–45] supporting that the CFT reduced density matrix ρ_R encodes all the information about the semiclassical physics in E_R. This means that with only access to the CFT state ρ_R, we can reconstruct the state of the quantum fields in E_R. Furthermore, we can implement the action of bulk semiclassical operators in E_R by acting with CFT operators that have support only on R. An implication of this is that the entanglement wedges E_{R_1} and E_{R_2} of two disjoint and spacelike separated boundary regions R_1 and R_2 are independent. This follows from the independence of R_1 and R_2 in a relativistic local theory. But this is puzzling from the bulk perspective, since the Einstein constraint equations are elliptic equations, and a change in the fields in one region might demand a change in the gravitational fields spacelike separated from it. An analogous effect is present in electromagnetism, where the addition of a charged particle requires the addition of electric fields reaching out to infinity. However, in electromagnetism, due to the presence of particles with both negative and positive charge, spatially separated regions R_1 and R_2 preserve their independence: we can always add some particles with the appropriate sign of the charge between R_1, R_2 to soak up electric field lines. In gravity this solution is not available to us since energy densities often are restricted to be positive, and even if they are not, matter energies have global positivity properties. In Chapter 6 we provide an explanation for how E_{R_1} and E_{R_2} preserve their independence at the classical level, despite the elliptic nature of the constraint equations. We do this by showing that

[8] We are assume that the wormhole is of macroscopic size, so $A \sim \mathcal{O}(1)$.

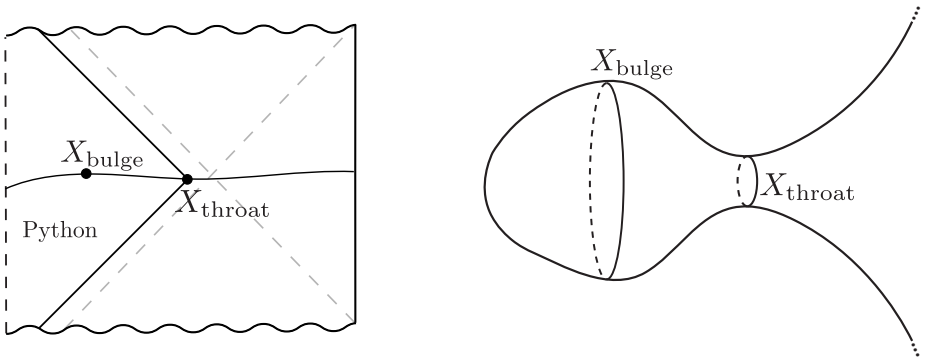


Figure 1.1: (Left) Illustration of a spacetime with a Python's lunch region. (Right) Illustration of a spatial slice through the extremal surfaces X_{bulge} and X_{throat}.

the extremal surfaces that appear in (1.8) play a specific functional purpose that enables E_{R_1} and E_{R_2} to stay independent. This provides a new consistency check on the concept of entanglement wedges, and more importantly, elucidates the functional role of the extremal surfaces that are so central to AdS/CFT.

1.1.5 Complexity

Recently, several connections have been proposed between the gravitational bulk and aspects of computational complexity theory in the CFT. For example, based initially on findings in toy models of AdS/CFT [46] and later supported by evidence in GR [47], a conjecture known as the "Python's lunch conjecture" [46] states that certain subregions of the entanglement wedge E_R require exponential computational resources to reconstruct from a boundary region R. Let us describe the conjecture for the special case of a single CFT in a pure state. In this case, the true QES that computes the von Neumann entropy of the full CFT state is the empty set. However, assume that we have two additional QESs X_{bulge} and X_{throat}, such that there exists a Cauchy slice Σ containing both surfaces, and with X_{bulge} and X_{throat} having locally maximal and minimal generalized entropy S_{gen} on the slice, respectively.[9] Then the conjecture states that the computational complexity $\mathcal{C}$ of implementing bulk operators in the region behind X_{throat} from the CFT perspective scales as

$$\mathcal{C} \propto \exp\left[\frac{1}{2}\left(S_{\text{gen}}[X_{\text{bulge}}] - S_{\text{gen}}[X_{\text{throat}}]\right)\right] \tag{1.9}$$

See Fig. 1.1 for an illustration of a Python's lunch geometry. In particular, it is believed that there is no simple formula like (1.6) that can be used to write a bulk local operator in the Python region in terms of local operators on the boundary.

In Chapter 4 we show that if the Python's lunch conjecture is true, then certain regions of spacetime must lie behind horizons, again making no assumption of WCC. Furthermore, using results from the theory of quantum learning [48] and certain assumptions about the nature singularities, bulk reconstruction, and the large$-N$ limit of AdS/CFT, we show that horizons

[9]The generalized entropy of a general QES is given by the right hand side of (1.8).

always must exist when the CFT time-evolution operator looks sufficiently random to a computationally bounded observer. This is significant, since guarantees of horizon formation are hard to come by, given the teleological nature of event horizons.

There are also other notions of complexity that are proposed to have an interpretation in AdS/CFT. A puzzling fact about black holes in AdS is that their volume, measured using maximal-volume hypersurfaces anchored at boundary time t, grows linearly forever. This is due to the fact that the length of the wormhole grows linearly with time. Taking non-perturbative quantum effects into account, the volume does eventually saturate [49]. However, this process takes a time exponential in the black hole entropy, which is a puzzle from the CFT perspective. What quantity in the CFT could evolve for so long, way past timescales associated with thermalization and scrambling of quantum information? One proposal [23] is that the volume computes the complexity $\mathcal{C}$ of preparing the CFT state $|\psi(t)\rangle$:

$$\mathcal{C}\left(|\psi(t)\rangle\right) = \frac{\text{Vol}[\Sigma_t]}{G_N L}, \tag{1.10}$$

where Σ_t is the maximal volume hypersurface anchored at time t on $\mathscr{I}$. $\mathcal{C}$ is the minimal number of elementary gates needed to create the state $|\psi(t)\rangle$ from some reference state $|R\rangle$, assuming some error tolerance and set of allowed gates. There is however an obvious issue with this definition: a precise definition of $\mathcal{C}$ requires many arbitrary choices that are not present in the right hand side of (1.10). Thus, a precise dual to the volume is yet to be understood. Nevertheless, the maximal volume is a natural diffeomorphism invariant observable that, unlike the entropy, is greatly sensitive to the black hole interior. Consequently, to understand the emergence of the black hole interior from the CFT, it is of great interest to pin down the dual CFT quantity. In Chapter 8 we prove a series of rigorous results about volumes in general classes of spacetimes. This provides a set of constraints that any proposed dual CFT quantity must match. Then in Chapter 9, we show that volumes have qualitatively new behaviors when we allow for the negative energy densities usually present in AdS/CFT, providing additional qualitative features that must be matched by any proposed CFT dual quantity.

1.2 Cosmic Censorship

One of the most important open problems in classical GR is the Weak Cosmic Censorship conjecture, proposed by Penrose in 1965 [50] and further refined by [51]. In AF and AAdS spacetimes, the conjecture states that in GR coupled to matter satisfying an energy condition,[10] the maximal evolution of a generic complete regular initial dataset is a spacetime with a complete null infinity. That is, singularities resulting from generic gravitational collapse are hidden inside black holes. Since the singularity theorems [50, 53, 54] show that singularity formation is generic in GR, the status of the WCC conjecture is of fundamental importance. A singularity implies a breakdown in the predictability of GR, and if singularities generically form outside black holes, then GR is inadequate for describing generic physics outside black

[10]This matter must be field theory matter, and not fluid-like matter models, as these can be singular on their own [52]. The energy condition is often taken to be the dominant energy condition, although this condition is too strong in AdS/CFT, as it is violated by the $m^2 < 0$ scalars discussed above.

holes. This is both a curse and a blessing; our best empirically backed theory of gravity could be inadequate for explaining observable phenomena in the night sky, any many theoretical results relying on WCC are in jeopardy. On the flip side, quantum gravitational effects could be more readily observable.

Before discussing WCC in more detail, let us comment on the question begged by the name *weak* cosmic censorship. What exactly is weak about it? The term weak is included to separate it from a different censorship conjecture, known as the strong cosmic censorship (SCC) [51, 55–57]. SCC is a logically distinct conjecture that does not imply WCC, so the terms weak and strong are somewhat misleading; however, SCC and WCC derive their names from the fact that the former is perceived by some to be a morally stronger constraint on the predictability of GR. The conjecture states that generic initial data evolve to spacetimes with no Cauchy horizons – that is, maximal Cauchy evolution generically cannot be extended to a larger spacetime by appending portions that are not uniquely determined by the initial data. Generic occurrence of Cauchy horizons in the interior of black holes would violate SCC but not WCC; generic formation of null- or timelike singularities reaching all the way out to $\mathscr{I}$, causing its incompleteness but no Cauchy horizon, would violate WCC but not SCC.

Let us now return to the WCC. In the case of four-dimensional AF spacetimes, there exist several results supporting WCC. Examples include rigorous mathematical proofs of the non-linear stability of Schwarzschild$_4$ [58, 59] and slowly rotating Kerr$_4$ [60–62] in vacuum GR, proof of the time-symmetric Penrose inequality [63, 64], the resistance of black holes to being overspun and overcharged [65, 66], and the fact that numerical GR simulations of black hole mergers [67–70] do not seem to produce naked singularities. However, for spacetime dimension five or higher [71–77], or if we consider asymptotically AdS$_4$ [34, 35, 78, 79] spacetimes, the WCC has already been shown to be false. There also exist work [80] arguing for hints of violation of WCC in four asymptotically flat dimensions, although this work is inconclusive.

In light of the violations of WCC outside 4d AF spacetime, should we abandon the idea behind the WCC in the general case, and retreat to caring about WCC only for $D = 4$ AF spacetimes? A careful look at the counterexamples suggests otherwise; some refined version of Cosmic Censorship is likely true more generally.

First, in the AdS$_4$ counterexamples to WCC [34, 35, 78, 79], a remarkable connection to quantum gravity was found. By dialing the matter coupling constants of their theory, the authors of [78, 79] discovered that their naked singularities stopped existing precisely when their theory entered the regime where it satisfied a version of the Weak Gravity Conjecture (WGC) [81] (see [82] for a recent review and [83] for a proof in certain cases). The WGC heuristically states that in any consistent theory of quantum gravity, gravity must be the weakest force. There are several related precise versions of the statement, and the one relevant here is the statement that any consistent theory of quantum gravity with a $U(1)$ gauge field must have a particle that is superextremal, i.e. a particle with a charge-to-mass ratio greater than that of an extremal black hole.[11] Thus, the findings of [78, 79] suggest that singularities might behave better when GR is coupled to matter arising as the classical limit of a consistent theory of quantum gravity, rather than to arbitrary matter. We provide additional evidence for this in Chapter 3, where we stress test the so-called Penrose inequality. This inequality

[11] For two such particles, gravitational attraction is weaker than electromagnetic repulsion. When a theory has superextremal particles, extremal black holes can discharge through radiation of these particles.

is a conjectured inequality that can be derived from WCC,[12] and a violation of it implies a violation of WCC. We are able to construct violations of the Penrose inequality by developing an automatized numerical search for counterexamples. However, the Penrose inequality can be derived from AdS/CFT [84] via (1.8), so our theories that violate the WCC through the PI cannot be holographic. Furthermore, matter arising from string theory resist all our attempts at violating the Penrose inequality, providing an additional consistency check on the idea that that singularities behave better in theories consistent with UV completion in quantum gravity.

Second, the singularities in the higher-dimensional counterexamples to WCC are, in some vague sense, small. In these examples, known as Gregory-LaFlamme (GL) instabilities [71–73], black strings of initially even thickness evolve towards lumpy black strings, which in turn try to pinch off and then bifurcate into multiple disconnected black holes. This causes the formation of singularities, since horizons cannot bifurcate in GR.[13] As suggested in [85], these singularities are perhaps morally similar to the breaking of a single water droplet into two. In this case, hydrodynamics breaks down only for a short amount of time and in a small region, after which the hydrodynamic description resumes its validity everywhere. Continuing hydrodynamic evolution requires that only a small non-extensive amount of information must be supplied by the microscopic molecular theory, here analogous to a UV complete quantum gravity theory. Thus, the GL singularities might give a breakdown of GR for only a small amount of time and in a small region. This is qualitatively different from a potential naked singularity caused by a violation of the Penrose inequality, which would require a large scale violation of GR.[14]

In light of this, an upgraded version of Cosmic Censorship might reasonably be expected to take into account the size of a singularity. However, a challenge with this idea has been to find a well defined way to measure the size of a singularity. In Chapter 4, motivated by AdS/CFT, we propose a diffeomorphism invariant way to measure the size of a naked singularity in AAdS spacetimes. This measure, while defined purely on the gravity side, is constructed so as to also measure the minimal amount of CFT time evolution (i.e. the UV complete description) that is needed to evolve through the naked singularity. Thus, it measures the amount of information that the UV theory must provide in order to evolve through the singularity. This leads us to a classification of three types of naked singularities: classical, Planckian, and semi-Planckian. While difficult to verify directly in the low-energy theory, GL singularities and the singularities at the endpoint of black hole evaporation could plausibly be of Planckian or semi-Planckian size. With this knowledge, together with the observations in the previous paragraph, we conjecture in Chapter 4 that in quantum gravity with a holographic UV-completion, only classical naked singularities are forbidden. Indirect support for this conjecture is provided by findings in Chapters 2–4.

[12] Together with certain other reasonable assumptions, such as the non-existence black holes with eternally time-evolving exteriors.

[13] Or more precisely, a future horizon cannot bifurcate towards the future [57].

[14] Assuming the violation is of classical $O(1)$ magnitude, as is the case for the counterexamples constructed in this thesis.

Permissions

The following Chapters, with the exception of Chapters 3 and 6, present work carried out with collaborators. These authors are listed in the publication section above, and the papers are reproduced with their permission.

References

[1] J. D. Bekenstein, *Black holes and the second law*, *Nuovo Cim. Lett.* **4** (1972) 737–740.

[2] J. D. Bekenstein, *Black holes and entropy*, *Phys. Rev. D* **7** (1973) 2333.

[3] S. W. Hawking, *Particle creation by black holes*, *Commun. Math. Phys.* **43** (1975) 199.

[4] G. 't Hooft, *Dimensional reduction in quantum gravity*, in *Salamfest 1993:0284-296*, pp. 0284–296, 1993. gr-qc/9310026.

[5] L. Susskind, *The world as a hologram*, *J. Math. Phys.* **36** (1995) 6377–6396, [hep-th/9409089].

[6] J. Maldacena, *The large N limit of superconformal field theories and supergravity*, *Adv. Theor. Math. Phys.* **2** (1998) 231, [hep-th/9711200].

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

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

[9] O. Aharony, S. S. Gubser, J. Maldacena, H. Ooguri, and Y. Oz, *Large N field theories, string theory and gravity*, *Phys. Rept.* **323** (2000) 183–386, [http://arXiv.org/abs/hep-th/9905111].

[10] E. D'Hoker and D. Z. Freedman, *Supersymmetric gauge theories and the AdS / CFT correspondence*, in *Theoretical Advanced Study Institute in Elementary Particle Physics (TASI 2001): Strings, Branes and EXTRA Dimensions*, pp. 3–158, 1, 2002. hep-th/0201253.

[11] L. Susskind and E. Witten, *The holographic bound in Anti-de Sitter space*, hep-th/9805114.

[12] V. Balasubramanian and P. Kraus, *A Stress tensor for Anti-de Sitter gravity*, *Commun. Math. Phys.* **208** (1999) 413–428, [hep-th/9902121].

[13] T. Banks, M. R. Douglas, G. T. Horowitz, and E. J. Martinec, *AdS dynamics from conformal field theory*, hep-th/9808016.

[14] A. Hamilton, D. N. Kabat, G. Lifschytz, and D. A. Lowe, *Local bulk operators in AdS/CFT: A Boundary view of horizons and locality*, *Phys.Rev.* **D73** (2006) 086003, [hep-th/0506118].

[15] A. Hamilton, D. N. Kabat, G. Lifschytz, and D. A. Lowe, *Holographic representation of local bulk operators*, *Phys.Rev.* **D74** (2006) 066009, [hep-th/0606141].

[16] N. Engelhardt and G. T. Horowitz, *Towards a Reconstruction of General Bulk Metrics*, *Class. Quant. Grav.* **34** (2017), no. 1 015004, [arXiv:1605.01070].

[17] N. Engelhardt and G. T. Horowitz, *Recovering the spacetime metric from a holographic dual*, in *International conference on string theory (Strings 2016) Beijing, China, August 1-5, 2016*, 2016. arXiv:1612.00391.

[18] P. Breitenlohner and D. Z. Freedman, *Positive energy in anti-de Sitter backgrounds AND gauged extended supergravity*, *Phys. Lett.* **B115** (1982) 197.

[19] P. Breitenlohner and D. Z. Freedman, *Stability in Gauged Extended Supergravity*, *Annals Phys.* **144** (1982) 249.

[20] H. Lu and C. N. Pope, *Exact embedding of N=1, D = 7 gauged supergravity in D = 11*, *Phys. Lett. B* **467** (1999) 67–72, [hep-th/9906168].

[21] M. Cvetic, M. J. Duff, P. Hoxha, J. T. Liu, H. Lu, J. X. Lu, R. Martinez-Acosta, C. N. Pope, H. Sati, and T. A. Tran, *Embedding AdS black holes in ten-dimensions and eleven-dimensions*, *Nucl. Phys. B* **558** (1999) 96–126, [hep-th/9903214].

[22] H. Lu, C. N. Pope, and T. A. Tran, *Five-dimensional N=4, SU(2) x U(1) gauged supergravity from type IIB*, *Phys. Lett. B* **475** (2000) 261–268, [hep-th/9909203].

[23] D. Stanford and L. Susskind, *Complexity and Shock Wave Geometries*, *Phys. Rev. D* **90** (2014), no. 12 126007, [arXiv:1406.2678].

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

[25] S. Ryu and T. Takayanagi, *Aspects of Holographic Entanglement Entropy*, *JHEP* **0608** (2006) 045, [hep-th/0605073].

[26] V. E. Hubeny, M. Rangamani, and T. Takayanagi, *A Covariant holographic entanglement entropy proposal*, *JHEP* **0707** (2007) 062, [arXiv:0705.0016].

[27] A. C. Wall, *Maximin Surfaces, and the Strong Subadditivity of the Covariant Holographic Entanglement Entropy*, *Class.Quant.Grav.* **31** (2014), no. 22 225007, [arXiv:1211.3494].

[28] A. Lewkowycz and J. Maldacena, *Generalized gravitational entropy*, *JHEP* **1308** (2013) 090, [arXiv:1304.4926].

[29] T. Faulkner, A. Lewkowycz, and J. Maldacena, *Quantum corrections to holographic entanglement entropy*, *JHEP* **1311** (2013) 074, [arXiv:1307.2892].

[30] N. Engelhardt and A. C. Wall, *Quantum Extremal Surfaces: Holographic Entanglement Entropy beyond the Classical Regime*, *JHEP* **01** (2015) 073, [arXiv:1408.3203].

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

[32] A. Almheiri, N. Engelhardt, D. Marolf, and H. Maxfield, *The entropy of bulk quantum fields and the entanglement wedge of an evaporating black hole*, *JHEP* **12** (2019) 063, [arXiv:1905.08762].

[33] G. Penington, *Entanglement Wedge Reconstruction and the Information Paradox*, arXiv:1905.08255.

[34] G. T. Horowitz, J. E. Santos, and B. Way, *Evidence for an Electrifying Violation of Cosmic Censorship*, *Class. Quant. Grav.* **33** (2016), no. 19 195007, [arXiv:1604.06465].

[35] T. Crisford and J. E. Santos, *Violating the Weak Cosmic Censorship Conjecture in Four-Dimensional Anti–de Sitter Space*, *Phys. Rev. Lett.* **118** (2017), no. 18 181101, [arXiv:1702.05490].

[36] M. Van Raamsdonk, *Building up spacetime with quantum entanglement*, *Gen. Rel. Grav.* **42** (2010) 2323–2329, [arXiv:1005.3035]. [Int. J. Mod. Phys.D19,2429(2010)].

[37] J. Maldacena and L. Susskind, *Cool horizons for entangled black holes*, arXiv:1306.0533.

[38] B. Czech, J. L. Karczmarek, F. Nogueira, and M. Van Raamsdonk, *The Gravity Dual of a Density Matrix*, *Class.Quant.Grav.* **29** (2012) 155009, [arXiv:1204.1330].

[39] M. Headrick, V. E. Hubeny, A. Lawrence, and M. Rangamani, *Causality & holographic entanglement entropy*, *JHEP* **12** (2014) 162, [arXiv:1408.6300].

[40] A. Almheiri, X. Dong, and D. Harlow, *Bulk Locality and Quantum Error Correction in AdS/CFT*, *JHEP* **04** (2015) 163, [arXiv:1411.7041].

[41] D. L. Jafferis, A. Lewkowycz, J. Maldacena, and S. J. Suh, *Relative entropy equals bulk relative entropy*, arXiv:1512.06431.

[42] X. Dong, D. Harlow, and A. C. Wall, *Reconstruction of Bulk Operators within the Entanglement Wedge in Gauge-Gravity Duality*, *Phys. Rev. Lett.* **117** (2016), no. 2 021601, [arXiv:1601.05416].

[43] D. Harlow, *The Ryu-Takayanagi Formula from Quantum Error Correction*, *Commun. Math. Phys.* **354** (2017), no. 3 865–912, [arXiv:1607.03901].

[44] T. Faulkner and A. Lewkowycz, *Bulk locality from modular flow*, *JHEP* **07** (2017) 151, [arXiv:1704.05464].

[45] J. Cotler, P. Hayden, G. Salton, B. Swingle, and M. Walter, *Entanglement Wedge Reconstruction via Universal Recovery Channels*, arXiv:1704.05839.

[46] A. R. Brown, H. Gharibyan, G. Penington, and L. Susskind, *The Python's Lunch: geometric obstructions to decoding Hawking radiation*, *JHEP* **08** (2020) 121, [arXiv:1912.00228].

[47] N. Engelhardt, G. Penington, and A. Shahbazi-Moghaddam, *A World without Pythons would be so Simple*, arXiv:2102.07774.

[48] L. Yang and N. Engelhardt, *The Complexity of Learning (Pseudo)random Dynamics of Black Holes and Other Chaotic Systems*, arXiv:2302.11013.

[49] L. V. Iliesiu, M. Mezei, and G. Sárosi, *The volume of the black hole interior at late times*, *JHEP* **07** (2022) 073, [arXiv:2107.06286].

[50] R. Penrose, *Gravitational collapse and space-time singularities*, *Phys. Rev. Lett.* **14** (Jan, 1965) 57–59.

[51] R. P. Geroch and G. T. Horowitz, *Global structure of spacetimes*, in *General Relativity: An Einstein Centenary Survey*, pp. 212–293. 1979.

[52] R. M. Wald, *Gravitational collapse and cosmic censorship*, pp. 69–85, 10, 1997. gr-qc/9710068.

[53] S. W. Hawking, *Singularities in the universe*, *Phys. Rev. Lett.* **17** (Aug, 1966) 444–445.

[54] S. W. Hawking and R. Penrose, *The Singularities of gravitational collapse and cosmology*, *Proc. Roy. Soc. Lond. A* **314** (1970) 529–548.

[55] R. Penrose, *Singularities of spacetime.*, in *Theoretical Principles in Astrophysics and Relativity* (N. R. Lebovitz, ed.), pp. 217–243. 1978.

[56] R. Penrose, *Singularities and time-asymmetry*, in *General Relativity: An Einstein Centenary Survey*, pp. 581–638, 1979.

[57] R. M. Wald, *General Relativity*. The University of Chicago Press, Chicago, 1984.

[58] M. Dafermos, G. Holzegel, and I. Rodnianski, *The linear stability of the Schwarzschild solution to gravitational perturbations*, *Acta Math.* **222** (2019) 1–214, [arXiv:1601.06467].

[59] M. Dafermos, G. Holzegel, I. Rodnianski, and M. Taylor, *The non-linear stability of the Schwarzschild family of black holes*, arXiv:2104.08222.

[60] M. Dafermos, G. Holzegel, and I. Rodnianski, *Boundedness and decay for the Teukolsky equation on Kerr spacetimes I: the case $|a| \ll M$*, arXiv:1711.07944.

[61] S. Klainerman and J. Szeftel, *Kerr stability for small angular momentum*, *Pure Appl. Math. Quart.* **19** (2023), no. 3 791–1678, [arXiv:2104.11857].

[62] E. Giorgi, S. Klainerman, and J. Szeftel, *Wave equations estimates and the nonlinear stability of slowly rotating Kerr black holes*, *arXiv e-prints* (May, 2022) arXiv:2205.14808, [arXiv:2205.14808].

[63] G. Huisken and T. Ilmanen, *The Inverse Mean Curvature Flow and the Riemannian Penrose Inequality*, *Journal of Differential Geometry* **59** (2001), no. 3 353 – 437.

[64] H. L. Bray, *Proof of the Riemannian Penrose Inequality Using the Positive Mass Theorem*, *Journal of Differential Geometry* **59** (2001), no. 2 177 – 267.

[65] R. Wald, *Gedanken experiments to destroy a black hole*, *Annals of Physics* **82** (1974), no. 2 548–556.

[66] J. Sorce and R. M. Wald, *Gedanken experiments to destroy a black hole. II. Kerr-Newman black holes cannot be overcharged or overspun*, *Phys. Rev. D* **96** (2017), no. 10 104014, [arXiv:1707.05862].

[67] F. Pretorius, *Evolution of binary black hole spacetimes*, *Phys. Rev. Lett.* **95** (2005) 121101, [gr-qc/0507014].

[68] J. G. Baker, J. Centrella, D.-I. Choi, M. Koppitz, and J. van Meter, *Gravitational wave extraction from an inspiraling configuration of merging black holes*, *Phys. Rev. Lett.* **96** (2006) 111102, [gr-qc/0511103].

[69] A. Buonanno, G. B. Cook, and F. Pretorius, *Inspiral, merger and ring-down of equal-mass black-hole binaries*, *Phys. Rev. D* **75** (2007) 124018, [gr-qc/0610122].

[70] S. Husa, S. Khan, M. Hannam, M. Pürrer, F. Ohme, X. Jiménez Forteza, and A. Bohé, *Frequency-domain gravitational waves from nonprecessing black-hole binaries. I. New numerical waveforms and anatomy of the signal*, *Phys. Rev. D* **93** (2016), no. 4 044006, [arXiv:1508.07250].

[71] R. Gregory and R. Laflamme, *Black strings and p-branes are unstable*, *Phys. Rev. Lett.* **70** (1993) 2837–2840, [hep-th/9301052].

[72] R. Gregory and R. Laflamme, *The Instability of charged black strings and p-branes*, *Nucl. Phys. B* **428** (1994) 399–434, [hep-th/9404071].

[73] L. Lehner and F. Pretorius, *Black Strings, Low Viscosity Fluids, and Violation of Cosmic Censorship*, *Phys. Rev. Lett.* **105** (2010) 101102, [arXiv:1006.5960].

[74] T. Andrade, R. Emparan, D. Licht, and R. Luna, *Cosmic censorship violation in black hole collisions in higher dimensions*, *JHEP* **04** (2019) 121, [arXiv:1812.05017].

[75] T. Andrade, R. Emparan, D. Licht, and R. Luna, *Black hole collisions, instabilities, and cosmic censorship violation at large D*, *JHEP* **09** (2019) 099, [arXiv:1908.03424].

[76] T. Andrade, P. Figueras, and U. Sperhake, *Violations of Weak Cosmic Censorship in Black Hole collisions*, arXiv:2011.03049.

[77] P. Figueras, T. França, C. Gu, and T. Andrade, *Endpoint of the Gregory-Laflamme instability of black strings revisited*, *Phys. Rev. D* **107** (2023), no. 4 044028, [arXiv:2210.13501].

[78] T. Crisford, G. T. Horowitz, and J. E. Santos, *Testing the Weak Gravity - Cosmic Censorship Connection*, *Phys. Rev. D* **97** (2018), no. 6 066005, [arXiv:1709.07880].

[79] G. T. Horowitz and J. E. Santos, *Further evidence for the weak gravity — cosmic censorship connection*, *JHEP* **06** (2019) 122, [arXiv:1901.11096].

[80] F. C. Eperon, B. Ganchev, and J. E. Santos, *Plausible scenario for a generic violation of the weak cosmic censorship conjecture in asymptotically flat four dimensions*, *Phys. Rev. D* **101** (2020), no. 4 041502, [arXiv:1906.11257].

[81] N. Arkani-Hamed, L. Motl, A. Nicolis, and C. Vafa, *The string landscape, black holes and gravity as the weakest force*, *JHEP* **06** (2007) 060, [hep-th/0601001].

[82] D. Harlow, B. Heidenreich, M. Reece, and T. Rudelius, *Weak gravity conjecture*, *Rev. Mod. Phys.* **95** (2023), no. 3 035003, [arXiv:2201.08380].

[83] B. Heidenreich and M. Lotito, *Proving the Weak Gravity Conjecture in Perturbative String Theory, Part I: The Bosonic String*, arXiv:2401.14449.

[84] N. Engelhardt and G. T. Horowitz, *Holographic argument for the Penrose inequality in AdS spacetimes*, *Phys. Rev. D* **99** (2019), no. 12 126009, [arXiv:1903.00555].

[85] R. Emparan, *Predictivity lost, predictivity regained: a Miltonian cosmic censorship conjecture*, *Int. J. Mod. Phys. D* **29** (2020), no. 14 2043021, [arXiv:2005.07389].

Part I

Cosmic Censorship from Holography

Chapter 2

Holography Abhors Visible Trapped Surfaces

Abstract: We prove that consistency of the holographic dictionary implies a hallmark prediction of the weak cosmic censorship conjecture: that in classical gravity, trapped surfaces lie behind event horizons. In particular, the existence of a trapped surface implies the existence of an event horizon, and that furthermore this event horizon must be outside of the trapped surface. More precisely, we show that the formation of event horizons outside of a strong gravity region is a direct consequence of causal wedge inclusion, which is required by entanglement wedge reconstruction. We make few assumptions beyond the absence of evaporating singularities in strictly classical gravity. We comment on the implication that spacetimes with naked trapped surfaces do not admit a holographic dual, note a possible application to holographic complexity, and speculate on the dual CFT interpretation of a trapped surface.

2.1 Introduction

Foundational results in modern gravitational physics, from black hole thermodynamics [1–3] to topological censorship [4], often rely on the absence of strong gravity outside of horizons, or more precisely the weak cosmic censorship conjecture [5]. Gedankenexperiments that rely upon the formation of black holes from generic matter collapse require an implicit assumption that matter typically coalesces into black holes and not into naked singularities.

This assumption appears reasonable given the apparent absence of naked singularities in observational data to date. The more precise statement of weak cosmic censorship (which we shall henceforth refer to simply as "cosmic censorship") is the conjecture that the Einstein equation evolves generic regular initial data with certain asymptotics to a complete asymptotic infinity $\mathscr{I}$ [6, 7]. In its initial formulation it applied specifically to asymptotically flat initial data in four dimensions [5], but it has been generalized to other dimensions and asymptotics when a suitable asymptotic infinity $\mathscr{I}$ exists (see e.g. [8]).

Developments over the past decade have eroded confidence in the validity of this statement at least in its more general form: counterexamples have been found in asymptotically AdS_4 [9–13] and in higher dimensional asymptotically flat space [14–21]. Even in four-dimensional asymptotically flat space, non-generic initial data can evolve to form naked singularities [22–24]. While it is in principle possible that cosmic censorship is in fact correct in four-dimensional asymptotically flat space (and for generic initial data), numerous extant violations in other settings suggest otherwise. Since results that rely on cosmic censorship are expected to be applicable in broad generality in arbitrary dimensions and often for AdS and dS asymptotics, violations of cosmic censorship – generic or otherwise – are problematic in any setting where classical gravity is expected to be valid.

It would be particularly unfortunate[1] if such violations were to indicate that trapped surfaces can lie outside of event horizons. In its current formulation, cosmic censorship forbids the existence of trapped surfaces – i.e. surfaces from which light rays converge in any direction due to gravitational lensing – outside of event horizons [25]. A large set of theorems in classical gravity relies on this result (see [25, 26]). Numerical relativity typically uses the detection of (marginally) trapped surfaces as an avatar for the event horizon, whose location (and existence) can only be determined in infinite time. Must we face the possibility that these results are all questionable?

Possibly not, at least in spacetimes that arise as classical limits of quantum gravity. The recent discoveries of violations of cosmic censorship in AdS_4 [9–13] have also been found to violate the Weak Gravity Conjecture [27]; cosmic censorship is restored precisely when the theory is adjusted so as to satisfy this conjecture, which is hypothesized to discriminate between spacetimes that do and do not admit valid UV completions. The confluence of validity of the cosmic censorship and the weak gravity conjecture has given rise to speculation that while classical General Relativity admits violations of cosmic censorship, the classical spacetimes that result from a truncation of a valid quantum theory of gravity do not [11, 13]: that is, that quantum gravity enforces cosmic censorship on its strict classical limit.

Here we focus on trapped surfaces rather than the statement of cosmic censorship in terms of initial data. The restriction of trapped surfaces to lie behind horizons is one of

[1] Or fortunate, from a certain point of view.

the most valuable consequences of cosmic censorship, since as noted above, it is a sine qua non for a number of results in General Relativity [26]. It is furthermore one of relatively few consequences of cosmic censorship that can be formulated in terms of the experience of a family of observers: a trapped surface outside of a horizon can in principle be detected by an asymptotic family of observers in finite time (or retarded time, in the asymptotically flat case). In fact, one could even go so far as to argue that the absence of trapped surfaces outside of horizons is a large part of the physical content of the weak cosmic censorship conjecture; namely that regions of strong gravity (usually heralded by trapped surfaces) are hidden from asymptotic observers. This statement also fortunately avoids any references to singularities, which are notoriously hard to work with – see [28–32] and references therein for literature in General Relativity attempting classify strengths and types of singularities[2].

We aim to test the hypothesis that quantum gravity forces trapped surfaces behind horizons: we use holography as a laboratory for the classical limit of quantum gravity and ask whether some principle of the AdS/CFT correspondence implies that trapped surfaces remain cloaked from the asymptotic boundary.

Concrete evidence in favor of such a conclusion was found recently in e.g. [34], which used the holographic dictionary to prove the Penrose Inequality in AdS [35–37], a key result implied by the combination of two oft-quoted but unproven conjectures: (1) that trapped surfaces lie behind horizons, and (2) that black holes equilibrate. The proof of [34] assumed neither (1) nor (2) but instead made use of the holographic entanglement entropy proposal of Ryu-Takayanagi [38] and Hubeny-Rangamani-Takayanagi [39] (HRT)

$$S_{\text{vN}}[\rho_R] = \frac{\text{Area}[X_R]}{4G\hbar}, \tag{2.1}$$

where ρ_R is the density matrix of the CFT state reduced to the region R and X_R is the minimal area stationary surface homologous to R.

Because the Penrose Inequality follows from the absence of trapped surfaces outside of horizons together with black hole equilibration, it is a good omen in favor of cosmic censorship; however, it falls well short of proving that trapped surfaces in fact must lie behind horizons[3].

In this article, we close this gap, thus proving a central consequence of cosmic censorship: the holographic dictionary implies that trapped surfaces lie behind event horizons. Our primary assumptions are (1) the HRT prescription, (2) that there exist unitary operators on the boundary whose effect in the bulk propagates causally, and (3) that singularities do not evaporate (a criterion that will be defined more rigorously in the following section) in classical gravity without violations of the null energy condition. Steps (1) and (2) together imply causal wedge inclusion [40, 41]: that the causal wedge must lie inside of the entanglement wedge. This holographic ingredient is a key step in our proof.

Before proceeding to the outline of the proof, let us briefly comment on quantum corrections. When quantum gravity effects are taken into account and violations of the null energy condition ($T_{ab}k^a k^b \geq 0$ for null k^a) are permitted, there is of course no expectation that the above formulation of cosmic censorship should remain valid; it is, after all, a prediction about the

[2]For instance, as discussed in [33], a singularity like the Gregory-Laflamme instability would by any nice definition be considered "weak" enough to be allowed by cosmic censorship. We thank R. Emparan for discussions on this topic.

[3]And falls even shorter of proving cosmic censorship.

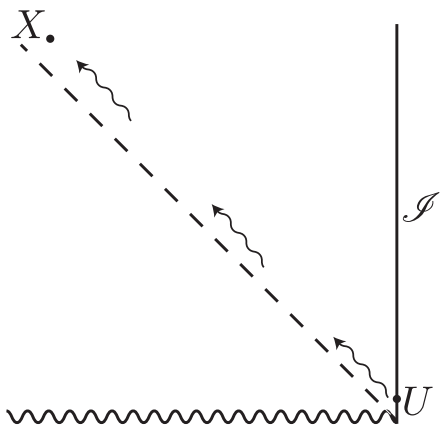


Figure 2.1: Example showing how causal connection between $\mathscr{I}$ and its HRT surface X can be used to change S_{vN} with a local unitary.

behavior of classical general relativity. Indeed, evaporating black holes do have trapped surfaces outside of horizons (see e.g. [42]). However, cosmic censorship is a statement about the classical theory; we are primarily concerned with ascertaining whether the classical limit of quantum gravity features trapped surfaces outside of event horizons. Within semiclassical gravity, however, these results may be interpreted as statements about early stages of gravitational collapse.

Outline of the Proof: the proof has three ingredients: first, the HRT prescription (2.1) combined with the fact that turning on local unitary operators in the dual theory at $\mathscr{I}$ results in causally propagating bulk perturbations; second, a theorem of [43] proving that an apparent horizon must lie between trapped surfaces and "normal" surfaces – surfaces in which ingoing light rays converge and outgoing light rays expand; and third, the holographic description of apparent horizons [44, 45] (which is not an additional ingredient but rather a construction that relies only on HRT).

First, we use the fact that $S_{\text{vN}}[\rho]$ is invariant under local unitary operations on the boundary state ρ to argue that the HRT surface X of a connected component of $\mathscr{I}$ cannot be timelike separated to any portion of $\mathscr{I}$. This is a well-established requirement often referred to as causal wedge inclusion [40, 41]: if X were timelike-separated to $\mathscr{I}$, then it would be possible for a local unitary CFT operator acting to create a bulk signal propagating causally to X, which could modify its area; this is illustrated in Fig. 2.1. Thus we would find that if X were timelike to any $p \in \mathscr{I}$ it would be possible to modify $S_{\text{vN}}[\rho]$ via local unitaries acting on ρ.[4]

Next, the theorem of [43] guarantees that a certain type of apparent horizon always lies between normal and trapped surfaces. We show that, given a trapped surface in our

[4]Causal wedge inclusion can also be shown to follow from entanglement wedge nesting [46], which states that if the size of a CFT subregion increases from R_1 to $R_2 \supset R_1$, the part of the bulk that could be reconstructed from R_1 can also be reconstructed from R_2.

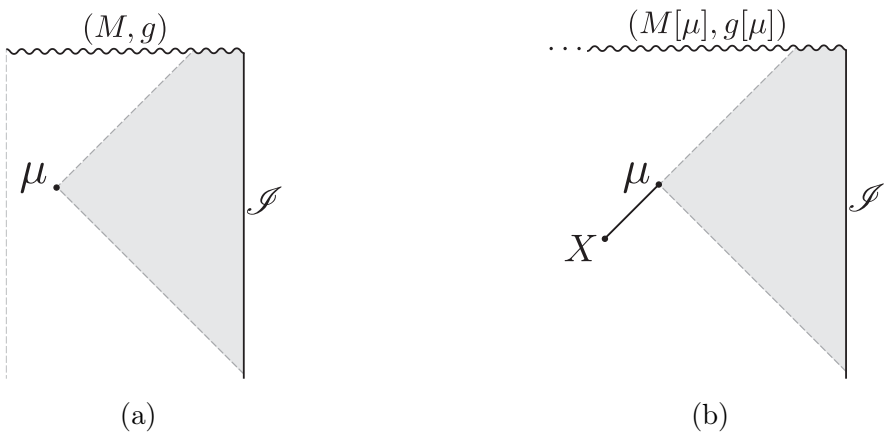


Figure 2.2: (a) An apparent horizon μ (more precisely, a minimar surface) in a spacetime (M, g). (b) The same apparent horizon in its corresponding coarse-grained spacetime $(M[\mu], g[\mu])$, where there HRT surface X with respect to $\mathscr{I}$ is null related to μ. The shaded wedge outside μ is common to both spacetimes.

setup, there are always normal surfaces outside of it, and that the type of apparent horizon guaranteed by the theorem satisfies the refinement necessary for step three.

Finally, we use the holographic construction of a dual to this type of apparent horizon [44, 45]. The construction instructs us to fix the spacetime and matter outside of the apparent horizon – its so-called outer wedge – and modify the spacetime elsewhere via a specific prescription. If naked singularities and Cauchy horizons develop as a result of this prescription, we permit any extension beyond the Cauchy horizon consistent with our relatively mild assumptions about General Relativity (here we operate under the assumption that the boundary conditions are inherited by a top-down UV completion). In this newly constructed spacetime, the HRT surface X is null-separated from the apparent horizon as illustrated in Fig. 2.2. Since by step one there are no timelike curves from the HRT surface to $\mathscr{I}$, there can be no future-directed timelike curves from the apparent horizon to $\mathscr{I}$: we find that apparent horizons must lie outside of $I^-[\mathscr{I}]$ in the coarse-grained spacetime; under the assumption that singularities do not evaporate in classical spacetimes satisfying the null energy condition, we can then deduce that apparent horizons must therefore also lie behind the event horizon in the original spacetime.[5] This immediately shows that naked singularities – should they exist in holographic spacetimes – cannot incur trapped surfaces outside of horizons. The contrapositive then yields the following statement: if a given spacetime has a trapped surface outside of (or without a) horizon, it cannot be holographic; should we adopt the perspective that $\mathrm{AdS}_{D\geq 3}$ spacetimes with a well-defined UV completion are always holographic, this then becomes a potential swampland condition on the set of spacetimes with valid UV completions.

[5]Since, to our knowledge, known solutions with evaporating singularities – a concept that we will make precise in Section 2.3 – arising from the evolution of initial data feature violations of the null energy condition, we do not find this assumption to be particularly prohibitive. More generally, violations of weak cosmic censorship are normally concerned with the formation of singularities rather than their demise. This will be discussed at greater length in Section 2.5 .

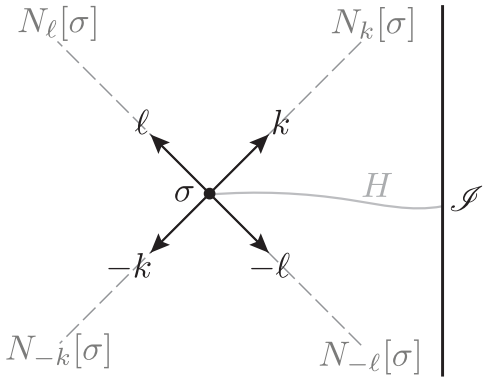


Figure 2.3: An illustration of the four null normals of a surface σ (here shown to be homologous to $\mathscr{I}$) together with the corresponding null congruences.

Relation to Prior Work: As alluded to earlier, previous attempts to prove various implications of the weak cosmic censorship conjecture have frequently encountered complications related to classifying singularities [28–32]. By considering trapped surfaces directly, and by defining evaporating singularities in terms of homology hypersurfaces (see Definition 1), we are able to avoid this complication altogether. Additionally, since cosmic censorship is known to be violated on a measure-zero set of the space of solutions to the Einstein equations [22–24], a genericity assumption is necessary for there to be any chance of the usual statement weak cosmic censorship to be true. The first half of our proof requires no assumptions about any genericity condition, and proves definitively that spacetimes with marginally trapped surfaces outside of horizons are not holographic. The second part of the proof, that all trapped surfaces are also behind horizons does assume a mild genericity condition and a technical assumption. We expect that these assumptions likely can be relaxed.

2.1.1 Assumptions and Conventions

Assumptions: We will assume the null convergence condition, $R_{ab}k^a k^b \geq 0$ for all null vectors k^a as well as the HRT prescription for computing S_{vN} (2.1) and entanglement wedge reconstruction. In particular, this means that we work strictly within classical General Relativity. We further assume that our spacetime is time-orientable and has no closed timelike curves. We further assume that in a maximal conformal extension $(\partial M, h)$ of $\mathscr{I}$, $\mathscr{I}$ is both globally hyperbolic and geodesically complete; unless explicitly noted otherwise, by $\mathscr{I}$ we will always mean a connected component of ∂M. By contrast, ∂M will always denote the complete conformal boundary. We assume that $\mathscr{I}$ is spatially compact.

Conventions: In describing domains of dependence that include the relevant portion of the asymptotic boundary, we will include $\mathscr{I}$ in any Cauchy slices of these domains of dependence, per standard conventions (see [47] for a definition).

1. Bulk: we assume that there are no closed causal curves and that $J^+(p) \cap J^-(q)$ is compact in the conformal completion for each pair of points q, p. This is typically

referred to as the AdS version of global hyperbolicity; in a slight abuse of notation, we shall simply refer to it as global hyperbolicity. We refer to the past Cauchy horizon of D as $\partial^- D$ and the future Cauchy horizon as $\partial^+ D$. For a set $A \subset M$, the boundary of A is denoted ∂A. We take $\hat{A}$ to be the closure of A in the conformal completion $M \cup \partial M$. If an intersection between a subset of M and $\mathscr{I}$ is taken, it is always implicitly assumed that the closure in the conformal completion is taken first. Examples of the sets described above are shown in Fig. 2.4.

By a surface, we will always mean a spacelike (achronal) codimension-two embedded submanifold (without boundary). A surface σ is said to be homologous to $\mathscr{I}$ if there is an achronal *homology* hypersurface H between σ and C; i.e. if there is an achronal hypersurface H such that (1) $\partial H = \sigma \cup C$ where C is a Cauchy slice of $\mathscr{I}$ and (2) H is compact in the conformal completion of (M, g). Any surface σ (homologous to $\mathscr{I}$ or otherwise) has two linearly independent normals, which we may pick to be null. We will denote the future-directed null normals k^a and ℓ^a; if σ is additionally homologous to $\mathscr{I}$, then at least one of these null normals points towards $\mathscr{I}$ – i.e. is "outwards-pointing"; we will name that vector k^a.[6] The four null geodesic congruences obtained by firing null geodesics along $\pm k^a$ and $\pm \ell^a$ from σ and terminating at conjugate points and geodesic intersections are denoted $N_{\pm k}[\sigma]$ and $N_{\pm \ell}[\sigma]$. See Fig. 2.3 for an illustration of these concepts. We will generally be interested in the expansion of a null congruence in the $n^a \in \{\ell^a, k^a\}$ direction, defined as

$$\theta_n = n^a \nabla_a \left(\ln \text{Area}[\sigma] \right) . \tag{2.2}$$

A marginally trapped surface satisfies $\theta_k = 0$ and $\theta_\ell < 0$ (whereas if θ_ℓ is unconstrained it is simply marginally *outer* trapped). A trapped surface has negative expansion along both future expansions.

Finally, as described in the introduction, we will work with what we shall refer to *permissible extensions* of a Cauchy horizon as an extension of a spacetime beyond a Cauchy horizon that is consistent with General Relativity and all our global assumptions listed above, together with one more condition: we assume that we never extend spacetimes beyond the "holographic region", should singularities terminating CFT evolution arise (see discussion below). All other conventions are as in [26].

2. Boundary: We use the letter C to refer to Cauchy slices of any maximal conformal extension of $\mathscr{I}$, and by i^+ we mean future timelike infinity of M.[7]. Since the utility of our proof is in the provision of a UV complete description of the gravitating system via the CFT, we are also concerned with the evolution of the boundary theory. That is, if at any point the CFT (in the large-N limit) becomes sick (e.g. if the stress tensor becomes divergent [48], Hamiltonian becomes unbounded from below, etc. [49]) in finite boundary time in a maximal conformal extension, we must conclude that this similarly puts an end to bulk evolution. Thus if the CFT evolution is well-defined (for $N \to \infty$)

[6] If ∂M consists of multiple connected components $\mathscr{I}$ and σ is homologous to multiple components, it is possible for both k^a and ℓ^a to be outwards-pointing, and it will be either clear from context or immaterial which one is which.

[7] Even though the commonly drawn conformal diagram of AdS does not show i^+, it nonetheless exists!

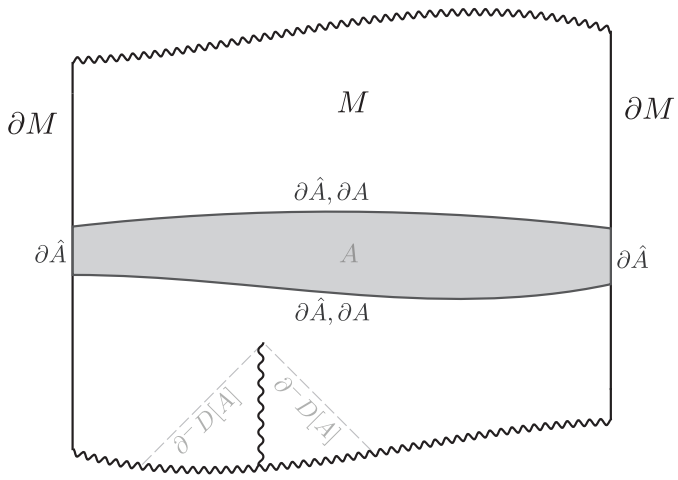


Figure 2.4: An example of an asymptotically AdS spacetime M with conformal boundary ∂M, and with set $A \subset M$ marked. The boundaries of A, ∂A in M and $\partial\hat{A}$ in $M \cup \partial M$, are highlighted. Furthermore, the past Cauchy horizon of $D[A] = D[\partial A]$ is illustrated.

only between two (potentially empty) boundary time slices C^- and C^+ of the maximal conformal extension $(\partial M, h)$, C^+ being to the future of C^-, then we excise $J^+[C^+]$ and $J^-[C^-]$ from (M, g). This is because these regions are not encoded in the CFT state between C^- and C^+, and so the spacetime without the excision is not completely encoded in the CFT and thus not holographic.[8] To facilitate terminology here, we define $\widetilde{i^+}$ as the futuremost endpoint of the CFT evolution. If the CFT evolution exists for all time in the maximal conformal extension, then $i^+ = \widetilde{i^+}$; otherwise, if the CFT evolution ends prematurely at some finite time, $\widetilde{i^+}$ becomes the effective timelike infinity: any signals that travel from the bulk to the boundary and arrive in the future of $\widetilde{i^+}$ are non-holographic.

2.2 Apparent Horizons in Holography

As discussed in Section 2.1, an apparent horizon is intuitively the boundary to the very strong gravity region on a given spatial slice; more precisely, it is the boundary between normal and trapped surfaces on a spatial slice. In this article we will be concerned with a slight refinement of apparent horizons called "minimar surfaces" [44].

A compact, connected surface μ is called a *future minimar* surface if it satisfies the following:

1. μ is marginally trapped;

[8] An example of a scenario like this is Choptuik critical collapse in AdS [48], where the CFT stress tensor becomes singular when the Cauchy horizon caused by the naked singularity reaches $\mathscr{I}$. Any potential extension of the spacetime to the future of this time is not holographically encoded in the CFT.

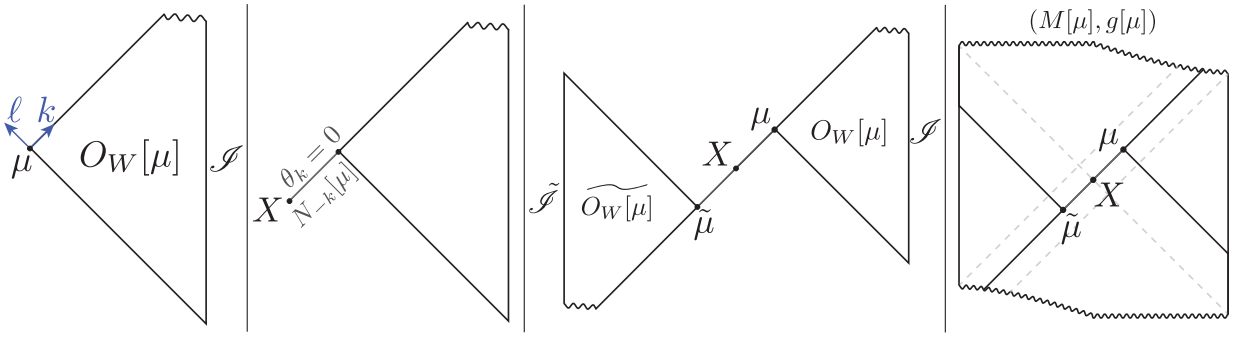


Figure 2.5: The coarse-grained spacetime construction of [44, 45]. The first panel shows the outer wedge of a minimar μ. The coarse grained spacetime elsewhere is specified by a piecewise null stationary initial data hypersurface $N_{-k}[\mu]$ emanating in the $-k^a$ direction from μ. It is possible to prescribe initial data on $N_{-k}[\mu]$ so that there is an extremal X surface on it. At this extremal surface, the spacetime is CPT reflected to generate a complete Cauchy slice of $(M[\mu], g[\mu])$.

2. μ is homologous to $\mathscr{I}$;

3. μ is *strictly stable*, meaning there exists choice of k^a and ℓ^a such that $k^a \nabla_a \theta_\ell < 0$ everywhere on μ (this effectively means that there exist untrapped surfaces outside of μ and trapped surfaces inside μ);

4. μ is the surface of least area on at least one of its homology hypersurfaces.

In analogy with the entanglement wedge, the "outer wedge " of μ, denoted $O_W[\mu]$, is the domain of dependence of H. The absence of global hyperbolicity raises a potential concern that the outer wedge is dependent on the choice of homology hypersurface. In the subsequent section, we will prove that this concern is unfounded in the absence of evaporating singularities. Finally, note that it is possible to define a past analogue of a minimar by replacing $\ell \to -k$, $k \to -\ell$. All of our results will be valid for this time-reversed choice, but for brevity we will focus on future case. Consequently we often refer to future minimars just as minimars.

Our results rely heavily on the primary construction of the dual to the apparent horizon, which builds a spacetime in which the area of the HRT surface is identical to the area of μ, and which agrees with (M, g) on $O_W[\mu]$. The construction is illustrated and described in Fig. 2.5. Because the spacetime in question is generated by discarding the region that was originally behind μ, it is often referred to as the coarse-grained spacetime, denoted $(M[\mu], g[\mu])$. Here we include in $(M[\mu], g[\mu])$ both the Cauchy development of the homology slice H and any permissible extension of the Cauchy horizon. This means that $(M[\mu], g[\mu])$ is not necessarily globally hyperbolic, and so some additional work, carried out in the proof of Lemma 1 and Theorem 1, is needed to show that the extremal surface X on $N_{-k}[\mu]$ about which we CPT reflect still is the HRT surface of $\mathscr{I}$ in $(M[\mu], g[\mu])$.

For our purposes, the takeaway from this construction is that for any minimar surface, there exists a spacetime $(M[\mu], g[\mu])$ in which μ is null-separated along $-k^a$ from the HRT surface X of $\mathscr{I}$. Furthermore, since the construction of $(M[\mu], g[\mu])$ makes no assumptions about global hyperbolicity [45], we are free to apply it in the context under consideration in this article.

2.3 Apparent Horizons Lie Behind Event Horizons

In this section, we argue that minimar surfaces must lie behind event horizons. The proof hinges on the well-known requirement discussed in Section 2.1 that for consistency of the proposal (2.1), the HRT surface of $\mathscr{I}$ must be spacelike- or null-separated to $\mathscr{I}$.[9] In modifying the area of the HRT surface, we would by the holographic prescription also modify $S_{\text{vN}}[\rho_{\text{bdy}}]$. Since the latter is invariant under local unitary operations, we immediately arrive at a contradiction; thus HRT surfaces must lie behind event horizons [41].

From this requirement of causal wedge inclusion, we will show that minimar surfaces must also lie behind an event horizon in the original spacetime. To prove this second step, it is critical that strictly classical GR satisfying the null convergence condition admit no evaporating singularities. Let us now make this requirement precise:

Definition 1. *An asymptotically AdS spacetime (M, g) is said to be devoid of evaporating singularities if for every closed set $K \subset M$, when $\partial\hat{K}$ is a compact hypersurface in the conformal completion then $\hat{K}$ is compact in the conformal completion.*

This property, by Lemma 8.2.1 of [26], ensures that no inextendible curves are imprisoned in $\hat{K}$, so that as we follow an inextendible geodesic in K they either leave K through ∂K or go to the conformal boundary. In particular, this rules out geodesic incompleteness between complete hypersurfaces that do not touch singularities. In the special case where ∂K is the union of two achronal surfaces, then compactness of $\hat{K}$ implies global hyperbolicity of $\hat{K}$.

As explained in the introduction, we will assume that all *strictly classical* solutions to GR that admit a UV completion in quantum gravity are devoid of evaporating singularities. We emphasize that this is a far weaker assumption than strong asymptotic predictability [26], since spacetime can violate global hyperbolicity arbitrarily badly near the asymptotic region owing to non-evaporating singularities. Our assumption further means that no permissible extension of the Cauchy horizon results in an evaporating singularity in classical GR. See Fig. 2.6 for some examples.

From this, we prove that the choice of homology hypersurface is immaterial to the definition of $O_W[\mu]$:

Lemma 1. *Let (M, g) be devoid of evaporating singularities. If H_1 and H_2 are two homology hypersurfaces of a surface μ homologous to $\mathscr{I}$, then $D[H_1] = D[H_2]$.*

Proof. To prove this, it is sufficient to show that every inextendible timelike curve intersecting H_1 also either intersects H_2 or reaches the conformal boundary (since H_1 includes a slice of $\mathscr{I}$). Assume first that $H_1 \cap H_2 = \varnothing$, and let C_1 and C_2 be the Cauchy slices of $\mathscr{I}$ where they are anchored. By global hyperbolicity and spatial compactness of $\mathscr{I}$ there is a compact subset $I \subset \mathscr{I}$ with $\partial I = C_1 \cup C_2$. Then $\partial\hat{K} \equiv I \cup H_1 \cup H_2$ is a compact hypersurface in the conformal completion (see Fig. 2.7 for an illustration), and the region $\hat{K}$ bounded by $\partial\hat{K}$ in the conformal completion is also compact.

[9] Readers concerned about recent results involving quantum extremal surfaces [47] outside of horizons [50] should defer their concern to Section 2.5, where we discuss quantum corrections and non-standard boundary conditions.

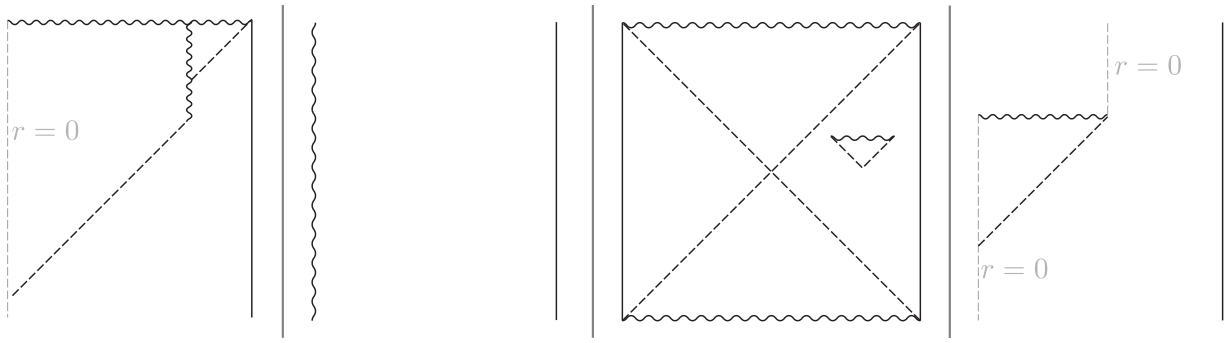


Figure 2.6: Sketched conformal diagrams of four spacetimes. The two left-most spacetimes are neither globally hyperbolic nor strongly asymptotically predictable, but they are devoid of evaporating singularities. The two right-most spacetimes contain evaporating singularities.

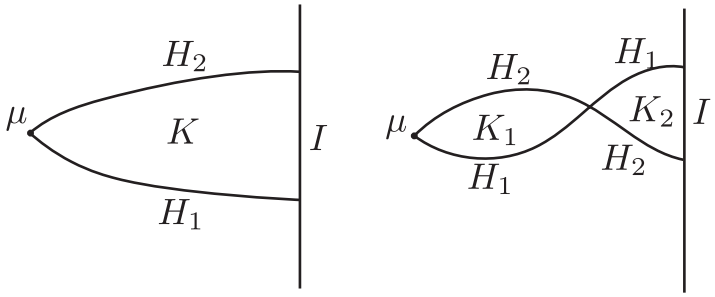


Figure 2.7: Illustration of two homology slices, which by the lack of evaporating singularities, must have the same outer wedge.

Assume an inextendible timelike curve γ enters K through H_1, and assume without loss of generality that it enters K to the future. By compactness of $\hat{K}$, γ must either reach the conformal boundary or leave K to its future. It cannot leave K through H_1 by global hyperbolicity of K. Thus if γ does not go to the conformal boundary, it intersects H_2.

In the case where H_1 and H_2 intersect we potentially get multiple compact regions $\hat{K}_1, \hat{K}_2, \ldots$ bounded by $H_1 \cup H_2 \cup I$, as shown in Fig. 2.7. Since the above argument will apply to each region we again find that γ intersects H_2 or reaches I.[10] □

For the remaining of the paper, unless stated otherwise we will make our above assumption that classical GR allows no evaporating singularities that satisfy the null convergence condition.

Theorem 1. *Let μ be a future minimar surface in (M, g). Then μ lies behind the future event horizon in (M, g).*

Proof. Let Σ_{tot} be an initial data slice of the spacetime $(M[\mu], g[\mu])$ consisting of $N_{-k}[\mu]$, H, and $\widetilde{H}$.[11] Within $D[\Sigma_{\text{tot}}]$, we know that the extremal surface $X[\mu]$ constructed via the

[10] Here we are ignoring potential issues that could arise if the intersection is a dense measure zero set; we assume that the domain of dependence of any one of these hypersurfaces is codimension-zero, and thus that we can "wiggle" the hypersurface to avoid that scenario.

[11] Here $\widetilde{A}$ refers to the CPT conjugate of a quantity $A \in O_W[\mu]$ in the coarse-grained spacetime.

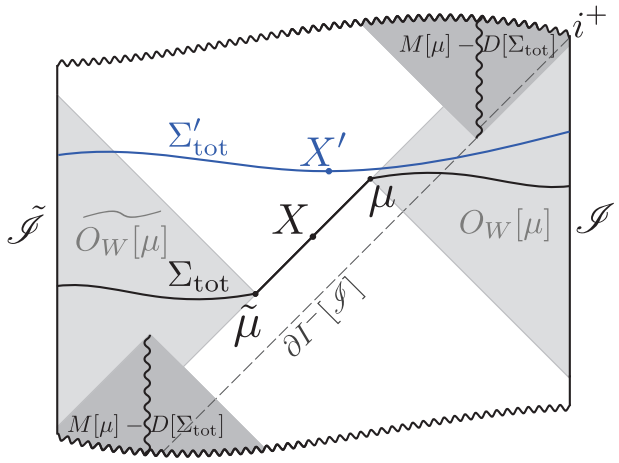


Figure 2.8: Illustration of a coarse grained spacetime extended beyond Cauchy horizons into the dark gray regions. X' is an extremal surface that is a candidate HRT surface. Since X' must be homologous to both $\mathscr{I}$ and $\tilde{\mathscr{I}}$, it will lie in $D[\Sigma_{\text{tot}}]$. μ must be spacelike to $\tilde{i}^+$ for any choice of extension $M[\mu] - D[\Sigma_{\text{tot}}]$.

above prescription is the minimal area extremal surface in $D[\Sigma_{\text{tot}}]$. If this remains true for any permissible extension of the Cauchy horizon, then $X[\mu]$ must be the HRT surface of $(M[\mu], g[\mu])$.

Suppose now that there is a permissible extension of the Cauchy horizon containing an extremal surface X' also homologous to $\mathscr{I}$. Since the total CFT state on all boundaries is pure, if X' were to be the HRT surface of $\mathscr{I}$, then it would also be the HRT surface of the complement $\widetilde{\mathscr{I}}$ by complementary recovery of classical holographic codes, and so it would have to be homologous to $\widetilde{\mathscr{I}}$. From the $\mathscr{I}$ and $\widetilde{\mathscr{I}}$ homology hypersurfaces of X' we construct the complete hypersurface Σ'_{tot} as the union (see illustration in Fig. 2.8). By the absence of evaporating singularities and the requirement that homology hypersurfaces are compact in the conformal completion, the region between Σ_{tot} and Σ'_{tot} is globally hyperbolic. But this means that $D[\Sigma'_{\text{tot}}] = D[\Sigma_{\text{tot}}]$, and so $X' \subset D[\Sigma_{\text{tot}}]$. This contradicts the assumption that X' lies in the extension beyond the Cauchy horizon. Thus the proof that $X[\mu]$ is the HRT surface from the case where $(M[\mu], g[\mu])$ is globally hyperbolic still applies, and so μ must be behind the horizon in $(M[\mu], g[\mu])$.

Consider now the subset $Z = J^+[\partial^+ O_W[\mu]] - J^+[\mu]$ of $M[\mu]$, which we can choose so that it agrees with the corresponding subset of M (see Fig. 2.9 for an example). If $\widetilde{i^+} \not\subset Z$, then CFT evolution proceeds beyond $\mathscr{I} \cap Z$, and $D[\Sigma_{\text{tot}}] \cup Z$ can be extended further. However, a maximal extension of $D[\Sigma_{\text{tot}}] \cup Z$ must put $\mathscr{I}$ in causal contact with μ, violating causal wedge inclusion, and so we conclude that $\widetilde{i^+} \subset Z$. Thus μ must lie behind the horizon in $O_W[\mu] \cup Z$, and by the absence of closed timelike curves this must remain true when completing $O_W[\mu] \cup Z$ into M. □

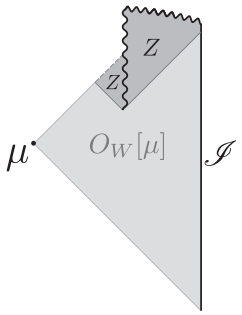


Figure 2.9: Example of a Cauchy extension Z of the outer wedge that is spacelike to μ. The fine-grained spacetime M will induce a particular choice of Z. It is possible to choose the coarse grained spacetime so that $O_W[\mu] \cup Z \subset M[\mu]$.

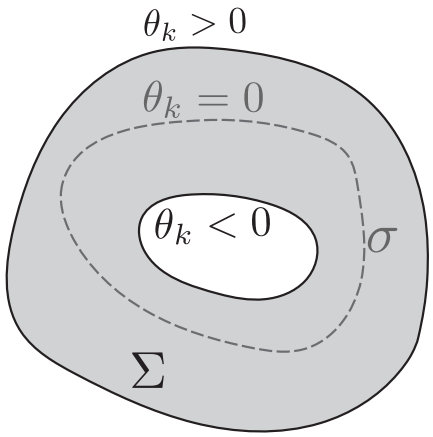


Figure 2.10: Illustration of a compact spacelike hypersurface Σ with an inner boundary that is outer trapped and an outer boundary that is outer untrapped. By Theorem 2 a marginally outer trapped surface σ in Σ is guaranteed to exist.

2.4 Trapped Surfaces Lie Behind Event Horizons

Having argued that minimar surfaces lie behind event horizons, we proceed to show the main result of this article, that *trapped* surfaces lie behind event horizons. The idea is simple: prove that between a trapped surface and the asymptotic boundary there always exists at least one minimar surface. We will make use of the following theorem:

Theorem 2 ([43, 51])**.** *Let Σ be a compact spacelike hypersurface with an inner boundary and an outer boundary. Assume that the inner boundary is outer trapped and that the outer boundary is outer untrapped. Then Σ contains a smooth stable marginally outer trapped surface σ.*

The quantities described in this theorem are illustrated in Fig. 2.10. In fact, we will use a stronger version of the above theorem that describes how σ evolves with an evolving family of spatial slices. We will assume a generic condition on trapped surfaces which is

frequently used in classical gravity proofs involved marginally trapped surfaces (see e.g. [52, 53]: $R_{ab}n^a n^b + \sigma_{ab}\sigma^{ab} > 0$, where σ_{ab} is the shear tensor and n^a is the generator of a null congruence fired orthogonally from the trapped surface. This condition ensures that every leaf in a spacelike hypersurface foliated by marginally outer trapped surfaces is strictly stable [52]. Note that in the following we will only require this condition inside of the outer wedge.

Combining now Theorem 2.1, 3.1 and 6.4 of [54] with the assumption of genericity and its implication of strict stability [52], we have the following:[12]

Theorem 3 ([54]). *Let Σ be a spacelike hypersurface and let $\partial\Sigma$ consist of two disconnected components σ_1 and σ_2. Assume further that one of these components, σ_1 is outer trapped, while the other σ_2 is outer untrapped. Then the boundary of the outer trapped region on Σ, denoted σ_t, is (1) a smooth, strictly stable, marginally outer trapped surface homologous to σ_2. Furthermore, if a spacetime region admits a foliation $\{\Sigma_t\}$ by such hypersurfaces, then the union $\mathcal{H} = \bigcup_{t\in[0,T]} \sigma_t$ is a piecewise smooth spacelike manifold with a finite number of connected components.*

To apply Theorem 3 to prove that trapped surfaces lie behind event horizon, we need σ_t to also be (1) marginally trapped and (2) homologous to $\mathscr{I}$. To achieve these conditions, we require the implementation of a mild assumption about the past of trapped surfaces ($\theta_\ell < 0$, $\theta_k < 0$). Since cosmic censorship is primarily concerned with the future of trapped surfaces, these assumptions do not heavily constrain our results.

Definition 2. *We call a surface σ past well-behaved if it is homologous to $\mathscr{I}$ and (1) $C[\sigma] = \partial^- O_W[\sigma] \cap \mathscr{I}$ is a Cauchy slice of $\mathscr{I}$ and (2) $\partial^- O_W[\sigma] \subset D$ for some interior of a domain of dependence D.*[13]

With this in place we prove a lemma essential for our main result:

Lemma 2. *Let τ be a past well-behaved trapped surface in a generic spacetime (M, g). Then there exists a spacelike manifold $H[\tau]$ in $O_W[\tau]$ which is foliated by past well-behaved strictly stable marginally outer trapped surfaces homologous $\mathscr{I}$.*

Proof. Pick a smooth one-parameter family of non-intersecting spatial slices Σ_t for $t \in [0, T]$ that are all contained in $D \cap O_W[\tau]$, where D is the interior of the domain of dependence containing $\partial^- O_W[\tau]$. Pick the hypersurfaces Σ_t so that they are anchored on Cauchy slices C_t of the conformal boundary, and pick the inner boundaries of Σ_t so they are either sufficiently close to τ or lying on $N_k[\tau]$, so that they are also trapped. By AdS asymptotics and global hyperbolicity in D, we can always choose the family so that each Σ_t has an untrapped surface near $\mathscr{I}$. Theorem 3 now guarantees the existence of a set $\mathcal{H}$ which is the union of smooth spacelike manifolds foliated by smooth strictly stable marginally outer trapped surfaces μ_t homologous to C_t (and thus $\mathscr{I}$), together with surfaces where $\mathcal{H}$ jumps. See Fig. 2.11. Since we choose our foliation so $\mathcal{H} \subset D$, the μ_t are also past well-behaved. Hence, any of the connected components of $\mathcal{H}$, after removing jump surfaces, satisfies all of our claimed properties of $H[\mu]$. □

[12] The theorem is only proven for $D = 4$ – we will assume it holds for $D \geq 3$, which appears likely given that Theorem 2 has been proven for $3 \leq D \leq 8$ [51].

[13] This is morally equivalent to demanding the existence of an open set $\mathcal{O}$ around σ that can be covered by surfaces that also fulfill requirement (1).

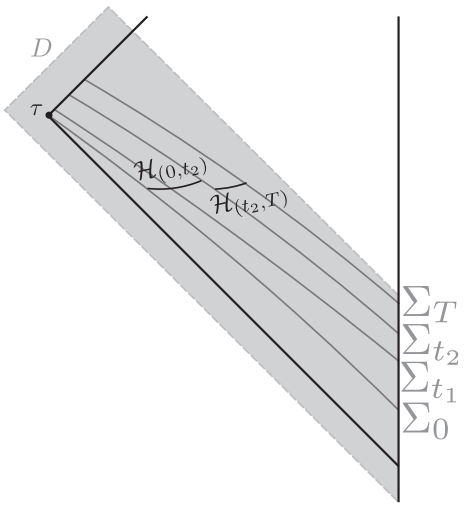


Figure 2.11: An example of the construction of the manifold $H[\tau]$ outside a past well-behaved trapped surface τ. Σ_t is a one-parameter family of spatial slices, and both $\mathcal{H}_{(0,t_2)}$ and $\mathcal{H}_{(t_2,T)}$ are spacelike manifolds foliated by marginally outer trapped surfaces. $H[\tau]$ can be taken to be either.

To find our main result, all that is missing is that at least one of the marginally outer trapped surfaces from the previous lemma is also inner trapped. By the past well-behaved assumption, these surfaces cannot be inner untrapped (so-called "anti-normal"). If the spacetime were spherically symmetric we would now be done: by past well-behavedness we can fire a spherically symmetric past horizon from $\mathscr{I}$ that hits μ_t. Choosing Σ_t (and thus μ_t) to respect the spherical symmetry, μ_t would be contained in the horizon, and so $\theta_\ell[\mu_t] < 0$, where strict inequality is guaranteed by our genericity condition. In the absence of spherical symmetry, this becomes harder to prove. We nevertheless find it to be a reasonable assumption for the following two reasons: (1) there is a continuously large amount of freedom in μ_t given that the choice of foliation Σ_t is arbitrary. (2) By past well-behavedness we can fire a one-parameter family of horizons from the boundary in D and pick each Σ_t to lie arbitrarily close to a member of this family of horizons. Thus we can construct an $H[\mu]$ where every leaf lies arbitrarily close to a strictly inner trapped surface (strictness follows from genericity). In fact, if we were allowed to use Theorem 3 directly on null foliations with a measure zero set of non-differentiable points, then we could redo the proof of Lemma 2 with Σ_t chosen to be future horizons, which would yield a proof that $\theta_\ell[\mu_t] < 0$ as well. With these justifications in mind, we will simply assume that a past-well behaved τ has at least one choice of $H[\tau]$ containing an inner trapped leaf μ_t. In light of the guaranteed existence of $H[\tau]$ by Lemma 2, this can be considered as an addition to the definition of being past well-behaved.

Finally, we can show our main result.

Theorem 4. *Let τ be a trapped surface. If τ is past well-behaved or has a trapped surface in its outer wedge which is past well behaved, then no future causal curves from τ can reach $\mathscr{I}$. In particular, τ lies behind an event horizon.*

Proof. By Lemma 2 we know there is a past well-behaved surface μ in $O_W[\tau]$ satisfying the

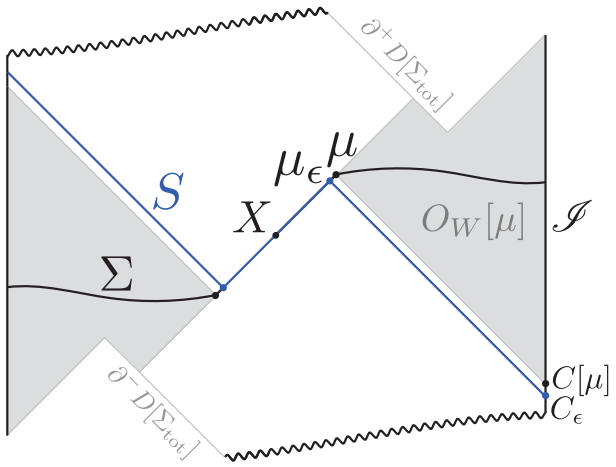


Figure 2.12: Construction of Cauchy surface S on which X is minimal in the globally hyperbolic region of the coarse grained spacetime $D[\Sigma]$.

minimar conditions (1)–(3). We now show that for a surface satisfying past well-behavedness, only properties (1)–(3) of a minimar are needed for the coarse grained spacetime to exist and having the required property that the extremal surface X on $N_{-k}[\mu]$ is the HRT surface of $\mathscr{I}$. That is, the requirement of minimality of μ on a homology slice can be exchanged for past well-behavedness, and Theorem 1 still applies to μ.

Existence of the coarse grained spacetime is immediate, since that only relies on properties (1)–(3), as shown in [45]. What remains is to show that X is HRT. The crucial piece used for proving this in [45] was that X is contained in a Cauchy surface of the coarse grained spacetime in which it is minimal. We now show that there is such a Cauchy surface S in the coarse grained spacetime even if we do not assume minimality of μ on a homology slice.

Let Σ be the Cauchy surface of the coarse grained spacetime formed by union of the homology slice of μ with respect to $\mathscr{I}$, its CPT conjugate, and $N_{-k}[\mu]$. Let us for now work in the maximal Cauchy development $D[\Sigma]$, currently refraining from extending possible Cauchy horizons.

By past well-behavedness in the original spacetime, $\partial^- O_W[\mu]$ is contained in a globally hyperbolic set there, and there are no singularities in a small neighbourhood around it. Since the coarse grained spacetime of μ shares the same data on $\partial^- O_W[\mu]$, any potential singularity in the coarse grained spacetime has to be finitely separated from $\partial^- O_W[\mu]$.[14] Thus, $C[\mu]$ is contained in the interior of $D[\Sigma]$. Generically, $C[\mu]$ will have kinks. Deform $C[\mu]$ slightly to the past into the Cauchy surface C_ϵ to smooth out the kinks, and consider the past horizon $H = \partial I^+[C_\epsilon] \cap J^-[\Sigma]$ and its intersection with the stationary null congruence, denoted $\mu_\epsilon = H \cap N_{-k}[\mu]$. See Fig. 2.12. We know that if μ_ϵ is nonempty, it must lie between X and μ on $N_{-k}[\mu]$, since X is not causally separated from $\mathscr{I}$. But since $\mu \subset I^+[C_\epsilon]$, we know that μ_ϵ is nonempty and in fact a complete slice of $N_{-k}[\mu]$, since otherwise a generator of $N_{-k}[\mu]$ going from X to μ would be a causal curve that enters $I^+[C_\epsilon]$ without intersecting $\partial I^+[C_\epsilon]$.

[14] We have not ruled out that there is a stress tensor shock along $\partial^- O_W[\mu]$, but this would not ruin global hyperbolicity.

We now take S to be piecewise null Cauchy slice given by the union of H, $\tilde{H}$ and the part of $N_{-k}[\mu]$ between μ_ϵ and its CPT conjugate. Since H is a future horizon it has $\theta_\ell < 0$, and since it ends on $N_{-k}[\mu]$, where every slice has the same area as μ, we get that any surface σ contained in H has $\text{Area}[\sigma] \geq \text{Area}[\mu_\epsilon] = \text{Area}[X]$. Clearly this also holds true for $N_{-k}[\mu]$ and $\tilde{H}$, and thus for the whole of S. Hence, we have constructed a Cauchy slice S of $D[\Sigma]$ on which X is minimal. Even though S does not intersect $O_W[\mu]$ or μ, the proof from [45] that X is the HRT surface carries through. Furthermore, we are now free to add any permissible extension of the Cauchy horizon, and the absence of evaporating singularities guarantees that X remains that HRT surface, as shown in the proof of Theorem 1. But now Theorem 1 guarantees that μ lies behind the future event horizon of $\mathscr{I}$, and since $\mu \subset O_W[\tau]$, τ lies behind the event horizon as well. □

A simple corollary of our main result follows; we will relax our assumption that classical GR is devoid of evaporating singularities now:

Corollary 1. *If there exists a past well behaved trapped surface τ in a classical asymptotically AdS spacetime (M, g) satisfying the null convergence condition, then at least one of the following holds:*

1. *(M, g) has an event horizon, and τ lies behind it;*
2. *Classical GR admits solutions with evaporating singularities (in particular, there exist solutions with $O_W[\tau]$ and evaporating singularities);*
3. *(M, g) has no holographic dual.*

2.5 Discussion

We have shown from the holographic dictionary that trapped surfaces, hallmarks of strong gravity, must be cloaked from the asymptotic boundary by event horizons. This proof comes at the heels of the holographic derivation of the Penrose Inequality in [34]; both results constitute strong evidence in favor of validity of a version of weak cosmic censorship in gravitational theories whose UV completion is well-described by AdS/CFT, even though the conjecture might be false in gravitational theories without such a UV completion. That is, these derivations suggest that there may be a formulation of weak cosmic censorship that is enforced by quantum gravity in the classical regime (which may not be valid as a statement about gravitational theories that do not exist as classical limits to quantum gravity). It is thus tempting to speculate that the location of trapped surfaces may serve as a "swampland" criterion of sorts [55]. We will not subscribe to such an interpretation here, but rather discuss the precise implications of our results in holography.

Before we do so, however, let us briefly comment on the assumption that evaporating singularities are absent in classical GR. This may seem like an odd assumption to make given that the actual singularity region is not described by the classical theory. However, the absence of of evaporating singularities is a constraint on the *predictions* of the classical theory: regardless of whether we expect quantum corrections to be important in that regime, we may still ask what purely classical gravity would predict. This does not, however, preclude an

investigation of the location of *quantum* trapped surfaces [56] and quantum cosmic censorship in (perturbative) quantum gravity where the null energy condition is violated; that is indeed a natural next step.[15]

Typicality of Black Holes: The proof that trapped surfaces imply the existence of event horizons pairs nicely with the Penrose singularity theorem [57], which states that trapped surfaces imply the existence of singularities, and so singularities are at least as typical as trapped surfaces. For an asymptotic observer dreaming of immortality, the Penrose singularity theorem can be a foreboding omen.[16] However, our result should provide some relief to the observer, since it shows that *event horizons are just as typical*, in the sense that every time the Penrose theorem is invoked to deduce the existence of a singularity in a holographic theory, we can also deduce the existence of an event horizon hiding this singularity. Since a black hole usually is defined by the existence of an event horizon, our result shows that black holes are as typical as trapped surfaces. Black holes always appear whenever gravity gets strong enough to focus both ingoing and outgoing lightrays. This is especially reassuring given the prevalence of gedankenexperiments (in holography and beyond!) whose conclusions are reliant upon the formation of event horizons as a direct consequence of typical gravitational collapse.

Trapped Surfaces, Horizons, and Complexity: In general, knowledge of the spacetime geometry at finite time is insufficient to locate event horizons: the latter are teleological with respect to $\mathscr{I}$. Our result provides a time-local constraint on the existence and location of event horizons from the behavior of trapped surfaces, whose location (and existence) can be ascertained in the neighborhood of a single moment of time. This suggests a potential application to the work starting with [58–60] on the connection between the event horizon and maximal chaos and fast scrambling in holography. Should particular moment-of-time data in the CFT be sufficient to deduce the existence of a trapped surface, this data must be also be sufficient to detect such signatures of event horizons, although the converse is false. This raises an interesting question: if trapped surfaces are a time-local guarantee for the existence of event horizons, what is the CFT dual of a trapped surface?

Some work has already been done in this direction: in [44, 45], it was conjectured that the coarse-grained spacetime associated to an apparent horizon μ is dual to a CFT quantum state ρ_{coarse} that preserves all correlators of "simple CFT operators" with simple sources turned on:

$$\text{Tr}[\rho\mathcal{O}_{\text{simple}}] = \text{Tr}\left[\rho_{\text{coarse}}\mathcal{O}_{\text{simple}}\right], \tag{2.3}$$

where by simple operators we mean CFT operators with support in $\mathscr{I} \cap O_W[\mu]$ that result in causal propagation in the bulk, and we have suppressed the sources (and ρ is the original state). On the other hand, the expectation value of highly complex operators is allowed to differ between the two states. Consistent with expectations about complexity of operators localized to the deep black hole interior [61, 62], operators localized in the spacetime behind μ should not be simply reconstructible with access to $\mathscr{I} \cap O_W[\mu]$. In particular, the simple

[15] NE thanks R. Bousso and M. Tomasevic for extensive conversations on this.

[16] Indeed, as Wald describes it [7], the formation of naked singularities would be a potential mechanism for a mad scientist to destroy the universe.

entropy proposal, if correct, would imply that correlation functions of a small number of the local CFT primaries of HKLL [63–65] restricted to $O_W[\mu] \cap \mathscr{I}$ should not contain the information about the spacetime behind the associated minimar surface μ. Our results serve as an important consistency check to this: if a trapped surface could be in the causal wedge of $\mathscr{I}$, then there would – by Theorem 3 – be a minimar surface in the causal wedge; but then parts of the region behind μ would be reconstructible from simple operators in $\mathscr{I} \cap O_W[\mu]$ by HKLL, in contradiction with the expectation that only complex operators are sensitive to the physics behind μ.

Thus, since trapped surfaces are generally behind minimar surfaces, trapped surfaces appear to be a robust time-local signal of bulk physics that cannot be reconstructed from the CFT with "simple experiments" (under assumption of the simple entropy proposal). In fact, via the simple entropy conjecture, there are protocols that employ simple operators that allow reconstruction of physics outside of minimars but inside horizons (see also upcoming work [66]). We could speculate that minimars (and their associated holographic screens [53, 67]) are the boundary inside of which simple reconstruction methods break down; since we have shown that these surfaces are not found in the asymptotic region, our results are suggestive of a time-local separation between simple and complex physics in the bulk not dissimilar from the one presented in [62].[17]

This is of course in harmony with the proposed relation between operator CFT complexity and bulk depth in [68–70]. In upcoming work [71], we give a precise covariant description in which regions deeper in the black hole are of higher complexity than their shallower counterparts.

Quantum Corrections and Black Hole Evaporation: In purely classical gravity, the assumption of the absence of evaporating singularities is quite reasonable: to our knowledge, known examples of evaporating singularities always involve some violation of the null energy condition. Thus our statement should really be viewed as a prediction of quantum gravity on the behavior of strictly classical GR with matter that is consistent with holography. What about quantum corrections?

Once quantum backreaction is included – in particular, once black holes can evaporate – it is possible for trapped surfaces to lie outside of event horizons. Interestingly, however, this effect appears at later times in the evaporation process (see e.g.[42]). Thus it is tempting to speculate that even in the perturbative quantum gravity regime, at times t much smaller than the Page time, the classical GR result remains valid: trapped surfaces lie behind the event horizon. We emphasize that this is purely speculative: our derivation does not apply in such a setting since the event horizon is teleological and takes into account the entire evaporation process.

Might we expect a quantum version of our statement to be valid? That is, do *quantum* trapped surfaces [56] lie behind event horizons? In [50], quantum extremal surfaces [47] "outside" of the horizon were found; in this case the existence of non-standard boundary conditions at the AdS boundary were crucial. Interestingly, the quantum focusing conjecture [72]

[17] The reader may prima facie suspect a contradiction here: our results apply to the nearest marginally trapped surface, whereas the Python's lunch conjecture applies to the nearest extremal surface. The difference lies in having access to the entire boundary as opposed to just $O_W[\mu] \cap \mathscr{I}$.This tension will be discussed more explicitly in [66].

in this case nevertheless enforces the absence of causal communication between the quantum extremal surface and $\mathscr{I}$. These complications illustrate the subtleties that must be accounted for in formulating a quantum version of our proof, in which evaporating singularities can no longer be ignored: the absence of causal communication from quantum trapped surfaces to $\mathscr{I}$ is not equivalent to the absence of quantum trapped surfaces in the causal wedge. This suggests that the correct generalization may actually involve an understanding of whether communication can occur *in practice* rather than whether or not it is forbidden by causal structure.

Acknowledgments

It is a pleasure to thank C. Akers, R. Bousso, R. Emparan, S. Fischetti, G. Horowitz, P. Jefferson, S. Leutheusser, and M. Tomasevic for discussions. This work is supported in part by NSF grant no. PHY-2011905 and the MIT department of physics. The work of NE is also supported in part by the U.S. Department of Energy, Office of Science, Office of High Energy Physics of U.S. Department of Energy under grant Contract Number DE-SC0012567 (High Energy Theory research). The work of ÅF is also supported in part by an Aker Scholarship.

References

[1] S. W. Hawking, *Gravitational radiation from colliding black holes*, *Phys. Rev. Lett.* **26** (1971) 1344–1346.

[2] J. D. Bekenstein, *Black holes and the second law*, *Nuovo Cim. Lett.* **4** (1972) 737–740.

[3] J. M. Bardeen, B. Carter, and S. W. Hawking, *The four laws of black hole mechanics*, *Commun. Math. Phys.* **31** (1973) 161.

[4] J. L. Friedman, K. Schleich, and D. M. Witt, *Topological censorship*, *Phys. Rev. Lett.* **71** (1993) 1486–1489, [gr-qc/9305017]. [Erratum: Phys.Rev.Lett. 75, 1872 (1995)].

[5] R. Penrose, *Gravitational collapse: The role of general relativity*, *Riv. Nuovo Cim.* **1** (1969) 252–276. [Gen. Rel. Grav.34,1141(2002)].

[6] R. P. Geroch and G. T. Horowitz, *Global structure of spacetimes*, in *General Relativity: An Einstein Centenary Survey*, pp. 212–293. 1979.

[7] R. M. Wald, *Gravitational collapse and cosmic censorship*, pp. 69–85, 10, 1997. gr-qc/9710068.

[8] J. Santos, *Connecting the weak gravity conjecture to the weak cosmic censorship*, . Talk given at Strings 2018.

[9] G. T. Horowitz, J. E. Santos, and B. Way, *Evidence for an Electrifying Violation of Cosmic Censorship*, *Class. Quant. Grav.* **33** (2016), no. 19 195007, [arXiv:1604.06465].

[10] T. Crisford and J. E. Santos, *Violating the Weak Cosmic Censorship Conjecture in Four-Dimensional Anti–de Sitter Space*, *Phys. Rev. Lett.* **118** (2017), no. 18 181101, [arXiv:1702.05490].

[11] T. Crisford, G. T. Horowitz, and J. E. Santos, *Testing the Weak Gravity - Cosmic Censorship Connection*, *Phys. Rev. D* **97** (2018), no. 6 066005, [arXiv:1709.07880].

[12] T. Crisford, G. T. Horowitz, and J. E. Santos, *Attempts at vacuum counterexamples to cosmic censorship in AdS*, *JHEP* **02** (2019) 092, [arXiv:1805.06469].

[13] G. T. Horowitz and J. E. Santos, *Further evidence for the weak gravity — cosmic censorship connection*, *JHEP* **06** (2019) 122, [arXiv:1901.11096].

[14] R. Gregory and R. Laflamme, *Black strings and p-branes are unstable*, *Phys. Rev. Lett.* **70** (1993) 2837–2840, [hep-th/9301052].

[15] R. Gregory and R. Laflamme, *The Instability of charged black strings and p-branes*, *Nucl. Phys. B* **428** (1994) 399–434, [hep-th/9404071].

[16] L. Lehner and F. Pretorius, *Black Strings, Low Viscosity Fluids, and Violation of Cosmic Censorship*, *Phys. Rev. Lett.* **105** (2010) 101102, [arXiv:1006.5960].

[17] P. Figueras, M. Kunesch, and S. Tunyasuvunakool, *End Point of Black Ring Instabilities and the Weak Cosmic Censorship Conjecture*, *Phys. Rev. Lett.* **116** (2016), no. 7 071102, [arXiv:1512.04532].

[18] P. Figueras, M. Kunesch, L. Lehner, and S. Tunyasuvunakool, *End Point of the Ultraspinning Instability and Violation of Cosmic Censorship*, *Phys. Rev. Lett.* **118** (2017), no. 15 151103, [arXiv:1702.01755].

[19] T. Andrade, R. Emparan, D. Licht, and R. Luna, *Cosmic censorship violation in black hole collisions in higher dimensions*, *JHEP* **04** (2019) 121, [arXiv:1812.05017].

[20] T. Andrade, R. Emparan, D. Licht, and R. Luna, *Black hole collisions, instabilities, and cosmic censorship violation at large D*, *JHEP* **09** (2019) 099, [arXiv:1908.03424].

[21] T. Andrade, P. Figueras, and U. Sperhake, *Violations of Weak Cosmic Censorship in Black Hole collisions*, arXiv:2011.03049.

[22] M. W. Choptuik, *Universality and scaling in gravitational collapse of a massless scalar field*, *Phys. Rev. Lett.* **70** (1993) 9–12.

[23] D. Christodoulou, *Examples of naked singularity formation in the gravitational collapse of a scalar field*, *Annals Math.* **140** (1994) 607–653.

[24] R. S. Hamade and J. M. Stewart, *The Spherically symmetric collapse of a massless scalar field*, *Class. Quant. Grav.* **13** (1996) 497–512, [gr-qc/9506044].

[25] S. W. Hawking and G. F. R. Ellis, *The large scale stucture of space-time*. Cambridge University Press, Cambridge, England, 1973.

[26] R. M. Wald, *General Relativity*. The University of Chicago Press, Chicago, 1984.

[27] N. Arkani-Hamed, L. Motl, A. Nicolis, and C. Vafa, *The string landscape, black holes and gravity as the weakest force*, *JHEP* **06** (2007) 060, [hep-th/0601001].

[28] F. J. Tipler, *Singularities in conformally flat spacetimes*, *Phys. Lett. A* **64** (1977) 8–10.

[29] C. J. S. Clarke, *Naked Singularities and Causality Violation*, in *Relativistic Astrophysics and Cosmology* (V. de Sabbata and T. M. Karade, eds.), vol. 1, p. 111, Jan., 1984.

[30] C. J. S. Clarke, *Naked Singularities and Causality Violation*, in *Relativistic Astrophysics and Cosmology* (V. de Sabbata and T. M. Karade, eds.), vol. 1, p. 111, Jan., 1984.

[31] B. C. Nolan, *Strengths of singularities in spherical symmetry*, *Phys. Rev. D* **60** (1999) 024014, [gr-qc/9902021].

[32] A. Ori, *Strength of curvature singularities*, *Phys. Rev. D* **61** (2000) 064016.

[33] R. Emparan, *Predictivity lost, predictivity regained: a Miltonian cosmic censorship conjecture*, *Int. J. Mod. Phys. D* **29** (2020), no. 14 2043021, [arXiv:2005.07389].

[34] N. Engelhardt and G. T. Horowitz, *Holographic argument for the Penrose inequality in AdS spacetimes*, *Phys. Rev. D* **99** (2019), no. 12 126009, [arXiv:1903.00555].

[35] R. Penrose, *Naked Singularities*, in *Sixth Texas Symposium on Relativistic Astrophysics* (D. J. Hegyi, ed.), vol. 224 of *Annals of the New York Academy of Sciences*, p. 125, 1973.

[36] I. Itkin and Y. Oz, *Penrose Inequality for Asymptotically AdS Spaces*, *Phys. Lett. B* **708** (2012) 307–308, [arXiv:1106.2683].

[37] V. Husain and S. Singh, *Penrose inequality in anti–de Sitter space*, *Phys. Rev. D* **96** (2017), no. 10 104055, [arXiv:1709.02395].

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

[39] V. E. Hubeny, M. Rangamani, and T. Takayanagi, *A Covariant holographic entanglement entropy proposal*, *JHEP* **0707** (2007) 062, [arXiv:0705.0016].

[40] A. C. Wall, *Maximin Surfaces, and the Strong Subadditivity of the Covariant Holographic Entanglement Entropy*, *Class.Quant.Grav.* **31** (2014), no. 22 225007, [arXiv:1211.3494].

[41] M. Headrick, V. E. Hubeny, A. Lawrence, and M. Rangamani, *Causality & holographic entanglement entropy*, *JHEP* **12** (2014) 162, [arXiv:1408.6300].

[42] R. Bousso and N. Engelhardt, *Generalized Second Law for Cosmology*, *Phys. Rev.* **D93** (2016), no. 2 024025, [arXiv:1510.02099].

[43] L. Andersson and J. Metzger, *The Area of horizons and the trapped region*, *Commun. Math. Phys.* **290** (2009) 941–972, [arXiv:0708.4252].

[44] N. Engelhardt and A. C. Wall, *Decoding the Apparent Horizon: Coarse-Grained Holographic Entropy*, *Phys. Rev. Lett.* **121** (2018), no. 21 211301, [arXiv:1706.02038].

[45] N. Engelhardt and A. C. Wall, *Coarse Graining Holographic Black Holes*, *JHEP* **05** (2019) 160, [arXiv:1806.01281].

[46] C. Akers, J. Koeller, S. Leichenauer, and A. Levine, *Geometric Constraints from Subregion Duality Beyond the Classical Regime*, arXiv:1610.08968.

[47] N. Engelhardt and A. C. Wall, *Quantum Extremal Surfaces: Holographic Entanglement Entropy beyond the Classical Regime*, *JHEP* **01** (2015) 073, [arXiv:1408.3203].

[48] P. M. Chesler and B. Way, *Holographic Signatures of Critical Collapse*, *Phys. Rev. Lett.* **122** (2019), no. 23 231101, [arXiv:1902.07218].

[49] T. Hertog and G. T. Horowitz, *Holographic description of AdS cosmologies*, *JHEP* **04** (2005) 005, [hep-th/0503071].

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

[51] L. Andersson, M. Eichmair, and J. Metzger, *Jang's equation and its applications to marginally trapped surfaces*, in *4th International Conference on Complex Analysis and Dynamical Systems*, 6, 2010. arXiv:1006.4601.

[52] A. Ashtekar and B. Krishnan, *Dynamical horizons and their properties*, *Phys.Rev.* **D68** (2003) 104030, [gr-qc/0308033].

[53] R. Bousso and N. Engelhardt, *Proof of a New Area Law in General Relativity*, *Phys. Rev.* **D92** (2015), no. 4 044031, [arXiv:1504.07660].

[54] L. Andersson, M. Mars, J. Metzger, and W. Simon, *The time evolution of marginally trapped surfaces*, *Classical and Quantum Gravity* **26** (Apr, 2009) 085018, [arXiv:0811.4721].

[55] C. Vafa, *The string landscape and the swampland*, hep-th/0509212.

[56] A. C. Wall, *The Generalized Second Law implies a Quantum Singularity Theorem*, *Class.Quant.Grav.* **30** (2013) 165003, [arXiv:1010.5513].

[57] R. Penrose, *Gravitational collapse and space-time singularities*, *Phys. Rev. Lett.* **14** (1965) 57–59.

[58] S. H. Shenker and D. Stanford, *Black holes and the butterfly effect*, *JHEP* **03** (2014) 067, [arXiv:1306.0622].

[59] D. A. Roberts, D. Stanford, and L. Susskind, *Localized shocks*, *JHEP* **03** (2015) 051, [arXiv:1409.8180].

[60] J. Maldacena, S. H. Shenker, and D. Stanford, *A bound on chaos*, *JHEP* **08** (2016) 106, [arXiv:1503.01409].

[61] D. Harlow and P. Hayden, *Quantum Computation vs. Firewalls*, *JHEP* **1306** (2013) 085, [arXiv:1301.4504].

[62] A. R. Brown, H. Gharibyan, G. Penington, and L. Susskind, *The Python's Lunch: geometric obstructions to decoding Hawking radiation*, *JHEP* **08** (2020) 121, [arXiv:1912.00228].

[63] A. Hamilton, D. N. Kabat, G. Lifschytz, and D. A. Lowe, *Local bulk operators in AdS/CFT: A Boundary view of horizons and locality*, *Phys.Rev.* **D73** (2006) 086003, [hep-th/0506118].

[64] A. Hamilton, D. N. Kabat, G. Lifschytz, and D. A. Lowe, *Holographic representation of local bulk operators*, *Phys.Rev.* **D74** (2006) 066009, [hep-th/0606141].

[65] A. Hamilton, D. N. Kabat, G. Lifschytz, and D. A. Lowe, *Local bulk operators in AdS/CFT: A Holographic description of the black hole interior*, *Phys. Rev.* **D75** (2007) 106001, [hep-th/0612053]. [Erratum: Phys. Rev.D75,129902(2007)].

[66] N. Engelhardt, G. Penington, and A. Shahbazi-Moghaddam Life without pythons would be so simple, to appear.

[67] R. Bousso and N. Engelhardt, *New Area Law in General Relativity*, *Phys. Rev. Lett.* **115** (2015), no. 8 081301, [arXiv:1504.07627].

[68] L. Susskind, *Why do Things Fall?*, arXiv:1802.01198.

[69] L. Susskind, *Complexity and Newton's Laws*, arXiv:1904.12819.

[70] L. Susskind and Y. Zhao, *Complexity and Momentum*, arXiv:2006.03019.

[71] N. Engelhardt and Å. Folkestad, "to appear.".

[72] R. Bousso, Z. Fisher, S. Leichenauer, and A. C. Wall, *Quantum focusing conjecture*, *Phys. Rev.* **D93** (2016), no. 6 064044, [arXiv:1506.02669].

Chapter 3

The Penrose Inequality as a Constraint on the Low Energy Limit of Quantum Gravity

Abstract: We construct initial data violating the Anti-de Sitter Penrose inequality using scalars with various potentials. Since a version of the Penrose inequality can be derived from AdS/CFT, we argue that it is a new swampland condition, ruling out holographic UV completion for theories that violate it. We produce exclusion plots on scalar couplings violating the inequality, and we find no violations for potentials from string theory. In the special case where the dominant energy condition holds, we use GR techniques to prove the AdS Penrose inequality in all dimensions, assuming spherical, planar, or hyperbolic symmetry. However, our violations show that this result cannot be generically true with only the null energy condition, and we give an analytic sufficient condition for violation of the Penrose inequality, constraining couplings of scalar potentials. Like the Breitenlohner-Freedman bound, this gives a necessary condition for the stability of AdS.

3.1 Introduction

Whether or not singularities are hidden behind event horizons is a longstanding open question in general relativity. In [1] Penrose showed that if (1) the answer to this is affirmative, and (2) collapsing matter settles down to Kerr, then the existence of certain special surfaces σ appearing in regions of strong gravity implies a lower bound on the spacetime mass:

$$G_N M \geq \sqrt{\frac{\text{Area}[\sigma]}{16\pi}}. \tag{3.1}$$

A proof of this inequality, named after Penrose, would amount to evidence in favor of singularities being hidden, but the inequality has not been proven except in special cases [2, 3] (see [4] for a review).

Recently, Engelhardt and Horowitz [5] gave a holographic argument for an AdS version of the Penrose inequality (PI), assuming the AdS/CFT correspondence, but not cosmic censorship nor anything about the endpoint of gravitational collapse. This suggests that hypothetical bulk matter allowing violation of the PI in AdS is incompatible with the AdS/CFT dictionary, and that the PI can serve as a new condition detecting low energy theories that cannot be UV completed in holographic quantum gravity, meaning theories that can never arise as the low-energy limit of a holographic quantum gravity theory valid at all energy scales.

In this article we construct violations of the PI for various scalar potentials, and produce exclusion plots in coupling space, delineating regions where we know that the PI is violated. Since the PI turns out to constrain neutral scalars, we find that it is distinct from the weak gravity conjecture [6]. Next, we present numerical evidence that supersymmetry is a sufficient condition for the PI. We also present an analytical sufficient condition on scalar couplings for a theory to violate the PI. Similar to the Breitenlohner-Freedman bound [7, 8], this provides a necessary condition for the stability of AdS. Next, while our work shows that general theories respecting the null energy condition (NEC) violate the PI, we are able to prove the PI in all dimensions greater than two for any theory satisfying the dominant energy condition (DEC), assuming spherical, planar, or hyperbolic symmetry.

We emphasize that while we in this work use the PI to constrain theories in the classical limit, these constraints are intimately tied to quantum gravity in the form of the AdS/CFT correspondence, which is a nonperturbative description of string theory in AdS [9–11]. This is because Penrose's original argument for his inequality is invalid for general low energy theories in AdS, since there exist theories violating cosmic censorship in AdS [12–15]. The only known way to to argue for the truth of the PI in AdS is using the full machinery of the AdS/CFT correspondence, and then taking its classical limit. Without reference to AdS/CFT, we have no principle to exclude theories violating the PI, while if we demand that our theory arises as the classical/low-energy limit of holographic quantum gravity, the PI must hold.

3.2 The Penrose Inequality in AdS/CFT

Consider an apparent horizon σ in an asymptotically AdS_{d+1} (AAdS) spacetime with mass M, meaning that the expansion of the outwards null geodesic congruence fired from σ is

vanishing, while the inwards expansion is non-positive. Assuming the holographic dictionary, Ref. [5] derived that

$$\text{Area}[\sigma] \leq A_{\text{BH}}(M), \tag{3.2}$$

where $A_{\text{BH}}(M)$ is the area of the most entropic stationary black hole of mass M in the theory. This is the AdS version of the PI that can be derived in holography, and by knowing the function $A_{\text{BH}}(M)$, Eq. (3.2) can be rewritten to give a lower bound on the mass, similar to Eq. (3.1) (see Eq. (3.4)).

The argument of [5] relied on (1) the HRT entropy formula [16–19], (2) the existence of the so-called coarse grained CFT state, whose von-Neumann entropy equals $\text{Area}[\sigma]/4G_N$ [20, 21], and (3) the fact that there exists a gravitational path integral for the microcanonical ensemble which has stationary black holes as saddles [22, 23]. The argument also makes the reasonable assumption that there is no spontaneous breaking of time translation symmetry in the CFT microcanonical ensemble, so that the microcanonical ensemble is dual to a stationary black hole.[1] Finally, σ had to satisfy two technical conditions: that it becomes a proper trapped surface when perturbed slightly inwards, and that σ is outermost minimal, meaning that there exists a spacelike or null hypersurface bounded by σ and the conformal boundary on which no other surface is smaller (see [20, 21] for precise conditions). In the special case where σ is an extremal surface, the first condition is not needed.

3.3 Constraining Scalar Potentials

Working with scalar fields and spherical symmetry in the classical limit, we will see that many scalar potentials that violate the DEC violate Eq. (3.2) as well. DEC-violating scalars are important, since they appear in known examples of AdS/CFT dualities after dimensional reduction of compact dimensions [24–27]. A generic DEC violating scalar potential will not even have a positive mass theorem (PMT) [28, 29], and in these theories the PI is automatically violated, but we will also find that theories where we are unable to construct negative mass solutions, despite extensive numerical search, will frequently violate the PI.

The theories we consider have the action

$$\int \mathrm{d}^{d+1}x\sqrt{-g}\left[\frac{1}{2}R + \frac{d(d-1)}{2L^2} - \frac{1}{2}|\nabla\phi|^2 - V(\phi)\right], \tag{3.3}$$

where L is a length scale that sets the cosmological constant, and where V is a potential satisfying $V(0) = V'(0) = 0$. To look for violations of the PI, we will construct AAdS initial data on a partial Cauchy surface Σ bounded by σ and the conformal boundary, such that (1) σ is an apparent horizon satisfying all the technical conditions relevant for Eq. (3.2), and (2) Σ can be embedded in a larger initial dataset on a complete hypersurface. This is sufficient to test if the PI holds for σ; the full spacetime is not needed.

What scalar potentials $V(\phi)$ should we consider in order to find violations of the PI? Ref. [30] proved an AdS_4 PI assuming spherical symmetry and the DEC. Assuming the ordinary gravitational mass M [31] is finite, we prove the following generalization (conjectured to be true in [32]):

[1] We thank Don Marolf for pointing this out.

Theorem 5. *Consider an asymptotically $AdS_{d+1\geq 3}$ spacetime with spherical ($k=1$), planar ($k=0$), or hyperbolic symmetry ($k=-1$), satisfying the Einstein equations $G_{ab}-\frac{d(d-1)}{2L^2}g_{ab}=8\pi G_N T_{ab}$ and the DEC: $T_{ab}u^a v^a\geq 0$ for all timelike u^a, v^a. If σ is a symmetric outermost marginally trapped surface with respect to a connected component of the conformal boundary with mass M, then*

$$\frac{16\pi G_N}{(d-1)\Omega_k}M\geq k\left(\frac{\text{Area}[\sigma]}{\Omega_k}\right)^{\frac{d-2}{d-1}}+\frac{1}{L^2}\left(\frac{\text{Area}[\sigma]}{\Omega_k}\right)^{\frac{d}{d-1}}. \tag{3.4}$$

Here Ω_k is the volume of the $(d-1)$–dimensional unit sphere, the plane, or the unit hyperbolic space (or a compactification thereof, in the latter two cases). While Ω_k might be infinite, the ratios $\text{Area}[\sigma]/\Omega_k$ and M/Ω_k are well defined. Furthermore, taking $k=1$ and $L\to\infty$ we get the PI for spherically symmetric asymptotically flat space in general dimensions. The mass is conventionally defined so $M=0$ for pure AdS (see [33] for a discussion definitions of mass in AdS). Let us now turn to the proof.

Proof: Consider an AAdS_{d+1} spacetime with spherical, planar, or hyperbolic symmetry, and consider a null gauge with coordinates (x^+,x^-,Ω^i) and metric

$$\mathrm{d}s^2=-2e^{-f(x^+,x^-)}\mathrm{d}x^+\mathrm{d}x^-+r(x^+,x^-)^2\mathrm{d}\Omega_k^2, \tag{3.5}$$

where r is a function of (x^+,x^-) and where $\mathrm{d}\Omega_k^2$ locally is the (unit) metric on the sphere, plane, or hyperbolic space. Define $k_\pm^a=(\partial_{x^\pm})^a\equiv(\partial_\pm)^a$, which has associated null expansions $\theta_\pm=(d-1)r^{-1}\partial_\pm r$. The quantity

$$\mu(x^+,x^-)=r^d\left[\frac{k}{r^2}-\frac{2\theta_+\theta_-}{k_+\cdot k_-(d-1)^2}+\frac{1}{L^2}\right], \tag{3.6}$$

can be seen to reduce to the spacetime mass at $r=\infty$, up to an overall factor: $16\pi G_N M=(d-1)\Omega_k\mu|_{r=\infty}$. The null-null components of the Einstein equations (in units with $8\pi G_N=1$) reduce to

$$\begin{aligned}\frac{rT_{\pm\pm}}{d-1}&=-\partial_\pm f\partial_\pm r-\partial_\pm^2 r,\\ \frac{rT_{+-}}{d-1}&=\partial_+\partial_- r+\frac{d^2-3d+2}{(d-1)r}\left[\frac{e^{-f}}{2}\left(k+\frac{r^2}{L^2}\right)+\partial_+ r\partial_- r\right]\\ &\quad+\frac{r}{L^2}e^{-f}\end{aligned} \tag{3.7}$$

Proceeding similarly to Ref. [34], we compute $\partial_\pm\mu$ and use Eqs. (3.7) to eliminate $\partial_\pm r,\partial_\pm^2 r,\partial_+\partial_- r$, yielding

$$\partial_\pm\mu=\frac{2e^f r^d}{(d-1)^2}\left(T_{+-}\theta_\pm-\theta_\mp T_{\pm\pm}\right). \tag{3.8}$$

The DEC implies that $T_{\pm\pm}\geq 0$ and $T_{+-}\geq 0$. Thus, $\pm\partial_\pm\mu$ is positive in an untrapped region ($\theta_+\geq 0,\theta_-\leq 0$), and so there μ is monotonically non-decreasing in an outwards spacelike direction. Evaluating μ on a marginally trapped surface that can be deformed to infinity along a untrapped spacelike path, which exists by the assumption that σ is outermost marginally trapped, gives that $kr^{d-2}+r^d L^{-2}\leq\mu|_{r=\infty}$. Converting $\mu|_{r=\infty}$ to mass gives Eq. (3.4). □

Now, the above proof applies for an apparent horizon which is outermost marginally trapped, which is not always the same as outermost minimal. However, at a moment of time-symmetry the two always coincide, since in this case we have that $\theta_\pm = \pm\mathcal{K}$ [35], where $\mathcal{K}$ is the mean curvature of σ in Σ, and minimality means that $\mathcal{K} = 0$. Thus, to look for violations of the PI in our setup, Theorem 5 shows that we need to consider theories violating the DEC, which for (3.3) means potentials that are negative somewhere.

As mentioned, DEC violating potentials arise in known $\mathrm{AdS}_{d+1}/\mathrm{CFT}_d$ dualities after dimensional reduction, but we can also see their relevance more directly. In AdS/CFT, bulk scalar fields are dual to local scalar operators $O(x)$ in the CFT that transform with scaling dimension Δ under dilatations: $O(x) \to \lambda^\Delta O(\lambda x)$. It turns out that whenever O is a relevant operator (i.e. $\Delta < d$), we must have that $m^2 \equiv \partial_\phi^2 V(\phi)|_{\phi=0} < 0$, leading to DEC violation. This follows from the standard expression for the scaling dimension Δ of O [11]: $\Delta = d/2 + \sqrt{(d/2)^2 + m^2 L^2}$.[2] $\Delta < d$ indeed means negative m^2, which is allowed as long as the Breitenlohner-Freedman (BF) bound [7, 8] is satisfied: $m^2 \geq m^2_{\mathrm{BF}} \equiv -d^2/(4L^2)$.

3.3.1 Black Hole Uniqueness, Positive Mass, and Compact Dimensions

Before constructing initial data, a few subtleties and known results should be addressed. First, the reference black hole of mass M appearing in Eq. (3.2) is the one that dominates the microcanonical ensemble at that mass, which is the one with the largest area [23]. Thus, if there exist black holes with larger area than AdS-Schwarzschild at a given mass, we seemingly have to construct these before claiming a violation. Black hole uniqueness is not established in AdS, so this seems like a difficult task. However, spherical symmetry allows significant simplification. In the static spherically symmetric case, Ref. [36] recently proved that the NEC implies $A_{\mathrm{BH}}(M) \leq A_{\mathrm{AdS-Schwarzschild}}(M)$, so AdS-Schwarzschild is the only spherical black hole that can dominate the microcanonical ensemble. Since the theories we consider here respect the NEC, we thus know that AdS-Schwarzschild is the correct black hole to compare to in Eq. (3.2), assuming we can take the reference black hole to be spherically symmetric. This is reasonable, and amounts to the assumption that the CFT microcanonical ensemble on a sphere does not break rotational symmetry spontaneously (in the bulk this is the fact that introducing spin at fixed energy tends to reduce the area, as can be seen from Kerr-AdS [37] and other known spinning black hole solutions [38–40]).

Second, it has been proven that the PMT holds even in certain theories violating the DEC. The prime example is in classical supergravity (SUGRA) theories [7, 8, 41], but in Einstein-scalar theory more general results are known. It was proved in [42, 43] that the PMT holds if the scalar potential $V(\phi)$ can be written as

$$V(\phi) = \frac{d(d-1)}{2L^2} + (d-1)W'(\phi)^2 - dW(\phi)^2 \tag{3.9}$$

for some real function $W(\phi)$ defined for all $\phi \in \mathbb{R}$ and satisfying $W'(0) = 0$ (provided we only turn on the scalar mode with fastest falloff [44–46], which is what we do here). If we considered

[2] It is possible to choose boundary conditions so that the scaling dimension of the operator dual to ϕ is $\Delta = \frac{d}{2} - \sqrt{(d/2)^2 + m^2 L^2}$ [11]. We do not do this here, as these boundary conditions require modifications of the definition of the spacetime mass.

a supersymmetric theory, W would be the so-called superpotential, but supersymmetry is not required, and W can be any function satisfying the above properties. Nevertheless, we keep referring to W as a superpotential. It is not known whether the existence of W is a necessary condition for the existence of a PMT; the proofs of [42, 43] only show that it is sufficient.

Third, suppose that an AAdS_{d+1} solution is a dimensional reduction of a higher dimensional solution with some number of compact dimensions. If the higher dimensional solution is a warped product rather than a product metric between AAdS_{d+1} and the compact space, then it is not a priori obvious that a violation of the lower dimensional PI implies a violation of the higher dimensional one. For theories stemming from higher dimensions, it could in principle be that the PI only is valid with all dimensions included, but our numerical findings argue against this, since potentials from known AdS/CFT dualities seem to respect the lower dimensional PI, as we will see. [3]

3.3.2 Constructing Initial Data

All the quantities appearing in the Penrose inequality can be located on a single timeslice, so we can test the Penrose inequality with initial datasets rather than full spacetimes. Let us now describe how we construct initial data. A spacelike initial dataset for the Einstein-Klein-Gordon system on a manifold Σ at a moment of time symmetry consists of a Riemannian metric γ_{ab} and a scalar profile ϕ on Σ that together satisfy the Einstein constraint equations. The extrinsic curvature K_{ab} and time-derivative of ϕ on Σ are both vanishing. In this case, the full constraint equations reduce to

$$\mathcal{R} + \frac{d(d-1)}{L^2} = |\nabla\phi|^2 + 2V(\phi), \tag{3.10}$$

where $\mathcal{R}$ is the Ricci scalar of γ_{ab}.

Next, we want the initial data to have finite mass and evolve to an AAdS spacetime, which constrains ϕ to fall off sufficiently fast. Furthermore, we demand σ to be outermost minimal, so that we can test Eq. (3.2). Note that $K_{ab} = 0$ implies that σ is extremal, so we need not impose the condition that σ can be perturbed inwards to a trapped surface.

To make the procedure explicit, we pick our coordinate system on Σ to be

$$\mathrm{d}s^2 = \frac{\mathrm{d}r^2}{1 + \frac{r^2}{L^2} - \frac{\omega(r)}{r^{d-2}}} + r^2\mathrm{d}\Omega^2, \quad r \in [r_0, \infty), \tag{3.11}$$

where $\omega(r)$ is a real function and $\mathrm{d}\Omega^2$ the metric of a round unit $(d-1)$–sphere. The marginally trapped surface σ is the sphere at $r = r_0 > 0$, and since we are considering a spacelike manifold, we need $\omega(r) \leq r^{d-2} + r^d L^{-2}$. As discussed in [48], the above coordinates break down only at locally stationary spheres, where the former inequality becomes an equality. Since we want σ to be outermost minimal, one coordinate system of the form (3.11) must be enough to cover Σ. In these coordinates, for a general choice of scalar profile $\phi(r)$,

[3] Furthermore, [47] found compelling evidence that $\mathrm{Area}[\sigma]/G_N$ is the same whether is computed with compact dimensions included or in the dimensional reduction, so that $\mathrm{Area}[\sigma]/(G_N M)$ is invariant under dimensional reduction, with the truth value of the PI being the same with or without compact dimensions.

the solution to the constraint reads (see for example [24, 48])

$$\omega(r) = e^{-h(r)}\left[\omega(r_0) + \int_{r_0}^{r} \mathrm{d}\rho \frac{e^{h(\rho)}\rho^{d-1}\chi(\rho)}{d-1}\right], \tag{3.12}$$
$$h(r) = \int_{r_0}^{r} \mathrm{d}\rho \frac{\rho\phi'(\rho)^2}{d-1}, \;\; \chi(r) = \left(1 + \frac{r^2}{L^2}\right)\phi'(r)^2 + 2V(\phi)\,.$$

To construct particular initial datasets, we must provide the profile $\phi(r)$ on $[r_0, \infty)$, together with value for r_0. The constant $\omega(r_0)$ is fixed by the condition of σ being marginally trapped, giving that $\omega(r_0) = r_0^{d-2} + r_0^d L^{-2}$. Finally, we can complete our initial dataset by gluing a second copy of the initial dataset to itself along σ [4] (possible since σ is extremal – see [20, 21] for details).

Let us now choose concrete scalar profiles. Since we are looking for counterexamples to the PI rather than a proof, we are free to consider special initial data. We consider two types of profiles, either

$$\phi(r) = \sum_{k=0}^{3} \mathrm{sign}(\eta_k)\left(\frac{|\eta_k|}{r}\right)^{\Delta+2k}, \tag{3.13}$$

or

$$\phi(r) = \begin{cases} \mu \log(r/R_0) & r_0 \le r \le R_0 \\ 0 & R_0 \le r \end{cases}, \tag{3.14}$$

for general constants $\{\eta_k\}$ and $\{\mu, R_0\}$ parametrizing the initial data. After picking numerical values of r_0 and either $\{\eta_k\}$ or $\{\mu, R_0\}$, we can compute the integrals (3.12) numerically, and we can obtain the mass as $16\pi G_N M = (d-1)\mathrm{Vol}[S^{d-1}]\omega(\infty)$. The only remaining thing to check is that $\omega(r)$ never exceeds $r^{d-2} + r^d L^{-2}$ for $r > r_0$. As long as this is true, σ satisfies the technical conditions required for the holographic derivation of Eq. (3.2).

Why do we choose the profiles (3.13) and (3.14)? By trying to minimize the mass while holding $\mathrm{Area}[\sigma] \propto r_0^{d-2}$ fixed, we are maximizing the chance of violating the PI, since smaller M means smaller $\mathrm{A}_{\mathrm{BH}}(M)$. To achieve a small mass, we want large regions of nonzero scalar field in order to accumulate negative energy through the potential, while minimizing the positive gradient contribution from $\chi(r)$. Thus, we want a scalar that falls off slowly and without unnecessary non-monotonic behavior. Furthermore, due to the factor $\exp\left[-(d-1)^{-1}\int_r^\infty \mathrm{d}\rho\rho\phi'(\rho)^2\right]$ in the integrand of Eq. (3.12) when computing $\omega(\infty) \propto M$, it is the behavior of ϕ at large r that matters (or the largest values of r where ϕ has support). Contributions to the mass from smaller r are exponentially suppressed. Now, a logarithmic profile has a slow monotonic falloff, but it requires compact support in order to have the requisite asymptotics. The profile (3.13) has the slowest possible falloff compatible with non-compact support and standard Dirichlet boundary conditions.

We now generate a particular dataset by first drawing r_0 with a uniform distribution from the range $(10^{-2}L, 20L)$, allowing both small and large black holes. For the profile (3.13), we draw the coefficients η_k from the range $(-3r_0, 3r_0)$, again with a uniform distribution. For

[4] Gluing Σ to an identical copy Σ' leads to a kink in $\phi(r)$ at σ, but we can smooth out this kink in an arbitrarily small neighbourhood $U = (r_0, r_0 + \epsilon) \subset \Sigma'$ without altering the initial data on Σ. This might produce a large $\phi''(r)$ in U, but since $\phi''(r)$ does not appear in the constraint equations the solution to the constraints on Σ' still exists for sufficiently small ϵ.

the profile (3.14) we draw $\mu \in (0, 10)$ and $R_0 - r_0 \in (0, 100L)$. The parameter ranges are chosen partly through trial and error – if we increase the parameter ranges for η_k or μ, we mostly produce invalid datasets where $\omega(r) \geq r^{d-2} + r^d L^2$ at some finite $r > r_0$. This is not surprising, since if ϕ gets a large amplitude, $\omega'(r)$ becomes large as well, causing $\omega(r)$ to overshoot $r^{d-2} + r^d L^{-2}$ near r_0.[5] Either way, the extent that our sampling of the space of profiles $\phi(r)$ is suboptimal corresponds to how much our exclusion plots below can be improved in the future.

3.3.3 Coupling Exclusion Plots

Let us first study $d = 3$ and the potential

$$V(\phi) = -\frac{9}{16}\phi^2 + 9\phi^3 + 11\phi^4, \tag{3.15}$$

which has $m^2 = \frac{1}{2}m^2_{\text{BF}}$. This theory does not have a superpotential, since solving (3.9) gives that a real $W(\phi)$ can only exist on a finite interval. However, we find no negative mass solutions after generating 10^5 initial datasets. Nevertheless, this theory violates the PI. For example, the profile

$$\phi(r) = \left(\frac{3.8}{r}\right)^{\Delta}\left[-1 + \left(\frac{3.2}{r}\right)^2 - \left(\frac{2.4}{r}\right)^4\right], \tag{3.16}$$

with $r_0 = 2.5L$ yields

$$A[\sigma]/A_{\text{AdS}-\text{Schwarzschild}}(M) \approx 1.2, \quad G_N M \approx 7L. \tag{3.17}$$

As shown by Penrose's original argument [1], the dataset (3.16) cannot settle down to a stationary black hole, so it will either collapse to a naked singularity, or we will have a Coleman-DeLuccia type decay [49],[6] where the conformal boundary terminates in finite time, and where the event horizon grows to infinite area.

Let us now repeat the analysis for multiple potentials. In Fig. 3.1 we show histograms of computed ratios $\text{Area}[\sigma]/A_{\text{AdS}-\text{Schwarzschild}}(M)$ in a large ensemble of initial datasets with potentials coming either from (1) dimensional reduction of SUGRA theories appearing in string theory and AdS/CFT, such as $D = 11$ [26, 27], Type IIB [25], or massive Type IIA [50] SUGRA, or (2) corresponding to a free tachyonic scalars with $m^2 > m^2_{\text{BF}}$. In the case of SUGRA, since we use scalar theories arising from consistent truncations, our initial datasets provide valid initial datasets in the various SUGRA theories, both in the dimensional reduction and with compact dimensions included (using the embeddings in [25–27, 50]). The specific potentials are shown in the legend of Fig. 3.1. We see that the PI holds for all our initial datasets. This does not amount to a proof that the PI holds, but it provides evidence, since for other potentials we will easily be able to produce violations while sampling from the

[5] We can check that $\omega'(r_0) \leq d r_0^{d-1} L^{-2} + (d-2) r_0^{d-3}$ is required to avoid overshooting $r^{d-2} + r^d L^{-2}$ near $r = r_0$, which gives that $\phi'(r_0)$ should be $\mathcal{O}(r_0^{-1})$, justifying the parametric dependence on r_0 chosen for μ and η_k.

[6] We thank Juan Maldacena for suggesting this possibility.

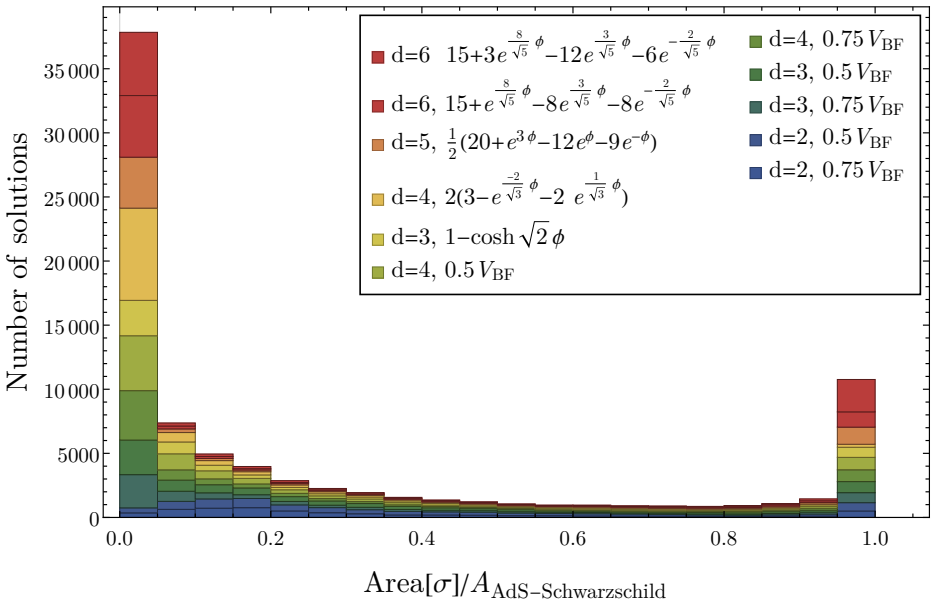


Figure 3.1: Plot of computed area ratios for various scalar potentials, with an ensemble of 10^4 initial data sets for each potential. $V_{\rm BF} \equiv \frac{1}{2}m^2_{\rm BF}\phi^2$. For the interacting theories, the $d = 6$ and $d = 3$ potentials come from S^4 [26] and S^7 [27] reduction of $D = 11$ SUGRA. The $d = 5$ potential comes from S^4 reduction of massive Type IIA SUGRA [50], and the $d = 4$ potential from S^5 reduction of Type IIB SUGRA [25].

same space of scalar profiles. This is an important consistency check on our proposal, since if the PI was violated for theories known to have a CFT dual, it presumably cannot serve as a constraint on low energy theories that can arise as the low energy limit of quantum gravity (a so-called swampland condition [6, 51, 52]).[7]

Consider now $d = 3$ and a potential with $m^2 = \frac{1}{2}m^2_{\rm BF}$, and with varying cubic and quartic couplings g_3, g_4 (see caption of Fig. 3.2). Take $g_3 \geq 0$ without loss of generality. For a given value of g_3, we can gradually lower g_4 until we find a dataset violating the PI or the PMT. In Fig. 3.2 we plot the highest value for g_4 for which we are able to find at least one violating dataset. Furthermore, we plot the region in (g_3, g_4) space in which a superpotential exists. The region of coupling space below the orange (blue) markers is ruled out by the PI (PMT). For $g_3 > 0$, the PI is a stronger condition than the PMT – at least in the space of initial data we are sampling. For reasons we do not understand, at $g_3 = 0$ where $\mathbb{Z}_2$ symmetry is restored, the PI and PMT are violated at the same time. However, $\mathbb{Z}_2$ symmetry does not appear to always guarantee coincidence, as shown in Fig. 3.3. Nevertheless, for $d = 3$ and a potential $V = \frac{1}{2}m^2\phi^2 + g_4\phi^4$, we find that the PI and PMT exclusion lines do coincide as we vary (m^2, g_4), and furthermore that exclusion line is well described by the analytical condition given below.

Note that there are no immediately obvious changes in the potential as we cross the line

[7]Note that for the $d = 4$ potential, we only consider the logarithmic profile, since the potential saturates the BF bound, so the mass formula requires modification for scalar profiles with non-compact support (see for example [53]).

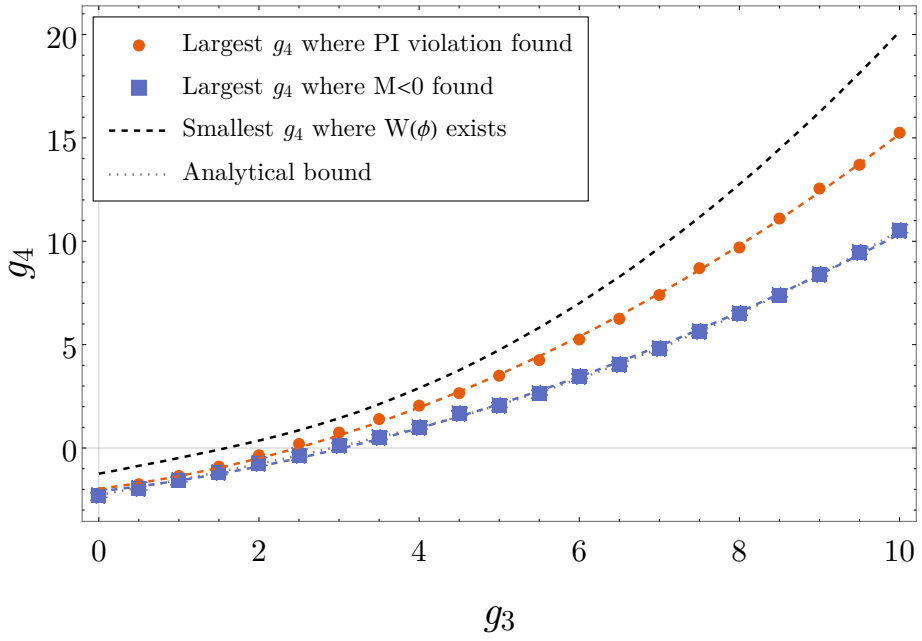


Figure 3.2: Exclusion plot on couplings for the potential $V(\phi) = \frac{1}{4}m_{\mathrm{BF}}^2\phi^2 + g_3\phi^3 + g_4\phi^4$ in $d+1 = 4$. Couplings below the circular markers are ruled out by the PI, while couplings below the squares are ruled out by positive mass. Blue and orange lines are quadratic fits, and couplings above the black dashed line give potentials which has superpotentials. Above the blue and orange markers, we have found no violations after the construction of 10^5 initial data sets using our sampling procedure. The dotted gray line, here coinciding with the blue, shows the exclusion boundary from the analytical condition (3.18).

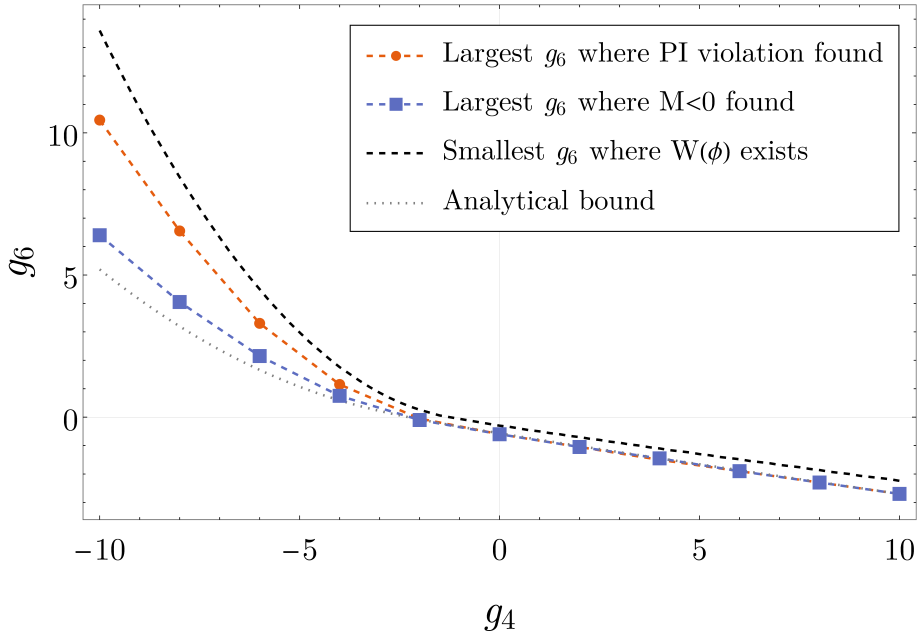


Figure 3.3: Exclusion plot on couplings for the potential $V(\phi) = \frac{1}{4}m_{\mathrm{BF}}^2\phi^2 + g_4\phi^4 + g_6\phi^6$ in $d+1 = 4$. The same description as in Fig. 3.2 applies, except the orange and blue lines are interpolations rather than fits.

into territory where we violate the PI. No new extrema develop.

3.3.4 Analytic bounds on couplings

So far we have given numerical bounds on couplings, through violation of the PI. We can also give analytical bounds, although they are somewhat weaker, and rely on violation of the PMT (implying PI violation). Consider the scalar profile (3.14), and a potential $V = \sum_{n=2}^{\infty} g_n \phi^n$. It is in fact possible to solve the integrals (3.12) analytically in terms of gamma functions, and while the solution is somewhat involved, the leading part of $\omega(\infty)$ in the limit $R_0 \to \infty$ is simple, yielding, up to $\mathcal{O}(R_0^{-2})$ corrections,

$$\frac{\lambda\omega(\infty)}{R_0^d} = \frac{\mu^2}{L^2} + 2\sum_{n=2}^{\infty} n! g_n \left[\frac{(1-d)\mu}{d(d-1)+\mu^2}\right]^n , \tag{3.18}$$

where $\lambda \equiv d(d-1)+\mu^2$, and with the dependence on r_0 contained in the $\mathcal{O}(R_0^{-2})$ terms. A sufficient condition for violation of the PMT and PI is for the RHS of (3.18) to be negative for some $\mu \in \mathbb{R}$. Thus, any theory where pure AdS is nonperturbatively stable must have a positive RHS of (3.18) for all μ. We included the exclusion line obtained from Eq. (3.18) in Figs. 3.2 and 3.3.

3.4 Discussion

There is by now a robust trend of proposing constraints on gravity theories in order for black holes to be well behaved semiclassically [6, 54], and for these constraints to later be proven in holography [55–57]. While the PI can be derived in holography, we have shown that it is generally false in GR, and argued that it serves as a new swampland [6, 51, 52] condition. As an example, we showed that it can be used to constrain scalar potentials for theories in AdS. If holography makes sense in asymptotically flat space, it is possible that the same logic can be applied there.

Acknowledgments

It is a pleasure to thank Netta Engelhardt and Gary Horowitz for comments on an earlier draft of this manuscript, and Aditya Dhumuntarao, Netta Engelhardt, Gary Horowitz, Veronika Hubeny, Marcus Khuri, Juan Maldacena, Don Marolf, and Cumrun Vafa for discussions. My research is supported in part by the John Templeton Foundation via the Black Hole Initiative, NSF grant no. PHY-2011905, and an Aker Scholarship. This research was also supported in part by the Heising-Simons Foundation, the Simons Foundation, and NSF grant no. PHY-1748958.

References

[1] R. Penrose, *Naked singularities*, *Annals of the New York Academy of Sciences* **224** (1973), no. 1 125–134.

[2] G. Huisken and T. Ilmanen, *The Inverse Mean Curvature Flow and the Riemannian Penrose Inequality*, *Journal of Differential Geometry* **59** (2001), no. 3 353 – 437.

[3] H. L. Bray, *Proof of the Riemannian Penrose Inequality Using the Positive Mass Theorem*, *Journal of Differential Geometry* **59** (2001), no. 2 177 – 267.

[4] M. Mars, *Present status of the Penrose inequality*, *Class. Quant. Grav.* **26** (2009) 193001, [arXiv:0906.5566].

[5] N. Engelhardt and G. T. Horowitz, *Holographic argument for the Penrose inequality in AdS spacetimes*, *Phys. Rev. D* **99** (2019), no. 12 126009, [arXiv:1903.00555].

[6] N. Arkani-Hamed, L. Motl, A. Nicolis, and C. Vafa, *The string landscape, black holes and gravity as the weakest force*, *JHEP* **06** (2007) 060, [hep-th/0601001].

[7] P. Breitenlohner and D. Z. Freedman, *Positive energy in anti-de Sitter backgrounds AND gauged extended supergravity*, *Phys. Lett.* **B115** (1982) 197.

[8] P. Breitenlohner and D. Z. Freedman, *Stability in Gauged Extended Supergravity*, *Annals Phys.* **144** (1982) 249.

[9] J. Maldacena, *The large N limit of superconformal field theories and supergravity*, *Adv. Theor. Math. Phys.* **2** (1998) 231, [hep-th/9711200].

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

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

[12] G. T. Horowitz, J. E. Santos, and B. Way, *Evidence for an Electrifying Violation of Cosmic Censorship*, *Class. Quant. Grav.* **33** (2016), no. 19 195007, [arXiv:1604.06465].

[13] T. Crisford and J. E. Santos, *Violating the Weak Cosmic Censorship Conjecture in Four-Dimensional Anti–de Sitter Space*, *Phys. Rev. Lett.* **118** (2017), no. 18 181101, [arXiv:1702.05490].

[14] T. Crisford, G. T. Horowitz, and J. E. Santos, *Testing the Weak Gravity - Cosmic Censorship Connection*, *Phys. Rev. D* **97** (2018), no. 6 066005, [arXiv:1709.07880].

[15] G. T. Horowitz and J. E. Santos, *Further evidence for the weak gravity — cosmic censorship connection*, *JHEP* **06** (2019) 122, [arXiv:1901.11096].

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

[17] S. Ryu and T. Takayanagi, *Aspects of Holographic Entanglement Entropy*, *JHEP* **0608** (2006) 045, [hep-th/0605073].

[18] V. E. Hubeny, M. Rangamani, and T. Takayanagi, *A Covariant holographic entanglement entropy proposal*, *JHEP* **0707** (2007) 062, [arXiv:0705.0016].

[19] A. Lewkowycz and J. Maldacena, *Generalized gravitational entropy*, *JHEP* **1308** (2013) 090, [arXiv:1304.4926].

[20] N. Engelhardt and A. C. Wall, *Decoding the Apparent Horizon: Coarse-Grained Holographic Entropy*, *Phys. Rev. Lett.* **121** (2018), no. 21 211301, [arXiv:1706.02038].

[21] N. Engelhardt and A. C. Wall, *Coarse Graining Holographic Black Holes*, *JHEP* **05** (2019) 160, [arXiv:1806.01281].

[22] J. D. Brown and J. W. York, Jr., *The Microcanonical functional integral. 1. The Gravitational field*, *Phys. Rev. D* **47** (1993) 1420–1431, [gr-qc/9209014].

[23] D. Marolf, *Microcanonical Path Integrals and the Holography of small Black Hole Interiors*, *JHEP* **09** (2018) 114, [arXiv:1808.00394].

[24] T. Hertog, G. T. Horowitz, and K. Maeda, *Negative energy density in Calabi-Yau compactifications*, *JHEP* **05** (2003) 060, [hep-th/0304199].

[25] H. Lu, C. N. Pope, and T. A. Tran, *Five-dimensional N=4, SU(2) x U(1) gauged supergravity from type IIB*, *Phys. Lett. B* **475** (2000) 261–268, [hep-th/9909203].

[26] H. Lu and C. N. Pope, *Exact embedding of N=1, D = 7 gauged supergravity in D = 11*, *Phys. Lett. B* **467** (1999) 67–72, [hep-th/9906168].

[27] M. Cvetic, M. J. Duff, P. Hoxha, J. T. Liu, H. Lu, J. X. Lu, R. Martinez-Acosta, C. N. Pope, H. Sati, and T. A. Tran, *Embedding AdS black holes in ten-dimensions and eleven-dimensions*, *Nucl. Phys. B* **558** (1999) 96–126, [hep-th/9903214].

[28] R. Schoen and S. T. Yau, *Proof of the positive mass theorem. ii*, *Comm. Math. Phys.* **79** (1981), no. 2 231–260.

[29] E. Witten, *A Simple Proof of the Positive Energy Theorem*, *Commun. Math. Phys.* **80** (1981) 381.

[30] V. Husain and S. Singh, *Penrose inequality in anti–de Sitter space*, *Phys. Rev. D* **96** (2017), no. 10 104055, [arXiv:1709.02395].

[31] A. Ashtekar and S. Das, *Asymptotically Anti-de Sitter space-times: Conserved quantities*, *Class. Quant. Grav.* **17** (2000) L17–L30, [hep-th/9911230].

[32] I. Itkin and Y. Oz, *Penrose Inequality for Asymptotically AdS Spaces*, *Phys. Lett. B* **708** (2012) 307–308, [arXiv:1106.2683].

[33] S. Hollands, A. Ishibashi, and D. Marolf, *Comparison between various notions of conserved charges in asymptotically AdS-spacetimes*, *Class. Quant. Grav.* **22** (2005) 2881–2920, [hep-th/0503045].

[34] S. A. Hayward, *Gravitational energy in spherical symmetry*, *Phys. Rev. D* **53** (1996) 1938–1949, [gr-qc/9408002].

[35] N. Engelhardt and Å. Folkestad, *Negative complexity of formation: the compact dimensions strike back*, *JHEP* **07** (2022) 031, [arXiv:2111.14897].

[36] Z.-Q. Xiao and R.-Q. Yang, *On Penrose inequality in holography*, 4, 2022.

[37] B. Carter, *Hamilton-Jacobi and Schrodinger separable solutions of Einstein's equations*, *Commun. Math. Phys.* **10** (1968), no. 4 280–310.

[38] S. S. Gubser, *Thermodynamics of spinning D3-branes*, *Nucl. Phys. B* **551** (1999) 667–684, [hep-th/9810225].

[39] M. Cvetic and S. S. Gubser, *Thermodynamic stability and phases of general spinning branes*, *JHEP* **07** (1999) 010, [hep-th/9903132].

[40] N. Altamirano, D. Kubiznak, R. B. Mann, and Z. Sherkatghanad, *Thermodynamics of rotating black holes and black rings: phase transitions and thermodynamic volume*, *Galaxies* **2** (2014) 89–159, [arXiv:1401.2586].

[41] G. W. Gibbons, C. M. Hull, and N. P. Warner, *The stability of gauged supergravity*, *Nucl. Phys.* **B218** (1983) 173.

[42] W. Boucher, *Positive energy without supersymmetry*, *Nucl. Phys.* **B242** (1984) 282.

[43] P. K. Townsend, *Positive Energy and the Scalar Potential in Higher Dimensional (Super)gravity Theories*, *Phys. Lett. B* **148** (1984) 55–59.

[44] A. J. Amsel and D. Marolf, *Energy Bounds in Designer Gravity*, *Phys. Rev. D* **74** (2006) 064006, [hep-th/0605101]. [Erratum: Phys.Rev.D 75, 029901 (2007)].

[45] A. J. Amsel, T. Hertog, S. Hollands, and D. Marolf, *A Tale of two superpotentials: Stability and instability in designer gravity*, *Phys. Rev. D* **75** (2007) 084008, [hep-th/0701038]. [Erratum: Phys.Rev.D 77, 049903 (2008)].

[46] T. Faulkner, G. T. Horowitz, and M. M. Roberts, *New stability results for Einstein scalar gravity*, *Class. Quant. Grav.* **27** (2010) 205007, [arXiv:1006.2387].

[47] P. A. R. Jones and M. Taylor, *Entanglement entropy in top-down models*, *JHEP* **08** (2016) 158, [arXiv:1602.04825].

[48] N. Engelhardt and Å. Folkestad, *General bounds on holographic complexity*, *JHEP* **01** (2022) 040, [arXiv:2109.06883].

[49] S. Coleman and F. D. Luccia, *Gravitational effects on and of vacuum decay*, *Phys. Rev. D* **21** (1980) 3305–3315.

[50] M. Cvetic, H. Lu, and C. N. Pope, *Gauged six-dimensional supergravity from massive type IIA*, *Phys. Rev. Lett.* **83** (1999) 5226–5229, [hep-th/9906221].

[51] C. Vafa, *The String landscape and the swampland*, 9, 2005.

[52] H. Ooguri and C. Vafa, *On the geometry of the string landscape and the swampland*, *Nucl. Phys.* **B766** (2007) 21–33, [hep-th/0605264].

[53] T. Hertog and G. T. Horowitz, *Towards a big crunch dual*, *JHEP* **07** (2004) 073, [hep-th/0406134].

[54] T. Banks and N. Seiberg, *Symmetries and Strings in Field Theory and Gravity*, *Phys. Rev. D* **83** (2011) 084019, [arXiv:1011.5120].

[55] D. Harlow and H. Ooguri, *Symmetries in quantum field theory and quantum gravity*, *Commun. Math. Phys.* **383** (2021), no. 3 1669–1804, [arXiv:1810.05338].

[56] D. Harlow and H. Ooguri, *Constraints on Symmetries from Holography*, *Phys. Rev. Lett.* **122** (2019), no. 19 191601, [arXiv:1810.05337].

[57] N. Engelhardt and Å. Folkestad, *Holography abhors visible trapped surfaces*, *JHEP* **07** (2021) 066, [arXiv:2012.11445].

Chapter 4

Cryptographic Censorship

Abstract: We formulate and take two large strides towards proving a quantum version of the weak cosmic censorship conjecture. We first prove "Cryptographic Censorship": a theorem showing that when the time evolution operator of a holographic CFT is approximately pseudorandom (or Haar random) on some code subspace, then there must be an event horizon in the corresponding bulk dual. This result provides a general condition that guarantees (in finite time) event horizon formation, with minimal assumptions about the global spacetime structure. Our theorem relies on an extension of a recent quantum learning no-go theorem and is proved using new techniques of pseudorandom measure concentration. To apply this result to cosmic censorship, we separate singularities into classical, semi-Planckian, and Planckian types. We illustrate that classical and semi-Planckian singularities are compatible with approximately pseudorandom CFT time evolution; thus, if such singularities are indeed approximately pseudorandom, by Cryptographic Censorship, they cannot exist in the absence of event horizons. This result provides a sufficient condition guaranteeing that seminal holographic results on quantum chaos and thermalization, whose general applicability relies on typicality of horizons, will not be invalidated by the formation of naked singularities in AdS/CFT.

4.1 Introduction

Curvature singularities are ubiquitous in General Relativity: the evolution of smooth initial data generically results in such singularities [1–3]. Observationally, however, evidence for singularity formation from generic gravitational processes (such as matter collapse) remains at best indirect. This apparent absence of directly-observable singularities is often explained by the general expectation that the formation of horizons is likewise generic: so singularities are usually cloaked behind event horizons. This hypothesis — that singularities resulting from initial data evolution should be generically hidden behind horizons — is known as the Weak Cosmic Censorship Conjecture (WCCC) [4]. It is a cornerstone assumption of almost all results in modern gravitational physics, from classical theorems such as the Hawking area law [5], to key results from the general "it from qubit" program. In particular, the expectation that CFT thermalization corresponds to black hole equilibration, and more generally that black hole statistics are reflected in CFT thermal states, crucially depends on the assumption that states with horizons are typical in some measure.[1] If the WCCC were false, naked singularities could be as typical as black holes; our understanding of quantum chaos, eigenstate thermalization, and black hole information in holography would be called into question.

Surprisingly, it turns out that naked singularities do exist in solutions to the Einstein equation satisfying various energy conditions and regular initial data, and these states do *not* constitute a measure zero set. There is by now a large repertoire of counterexamples to the WCCC: in more than four spacetime dimensions, the studies of Gregory–Laflamme type instabilities of [7] have illustrated the existence of "pinch-off" singularities as the end-stage of certain black holes (e.g., the black string [8]) — see [9] for a review; more recently, [10–13] constructed extended naked singularities in asymptotically AdS_4 spacetimes with various matter profiles; see also [14]. A similar mechanism was used by [15] to construct a likely counterexample in the four-dimensional asymptotically flat setting.

It thus appears that horizons and singularities are not related in any trivial or obvious way: curvature singularities may form outside (or in the absence) of any horizons, and horizons may form without singularities behind them [16]. There is no extant formulation of WCCC that is (1) not violated and (2) sufficient to guarantee the consistency of seminal results in AdS/CFT such as the universality of the saturation of the chaos bound and the bulk manifestation of the CFT thermal partition function.

Some progress towards understanding cosmic censorship in quantum gravity was made in [10–12], where it was noted that many (though not quite all) of the counterexamples also violate the Weak Gravity Conjecture [17], suggesting the possibility that the WCCC in four dimensions is valid in the landscape but violated in the swampland. Further evidence in favor of this conclusion was presented in [14], which constructed large sets of AdS_4 initial data satisfying the gravitational constraints and the null energy condition whose future development must include naked singularities, but these theories violated properties required of theories admitting a UV completion via holography [18]. This hypothesis would suggest that while classical General Relativity may violate the WCCC, the semiclassical limit of quantum gravity does not.

[1] This is not identical to the firewall sense of typicality of [6]. See footnote 7 on page 79.

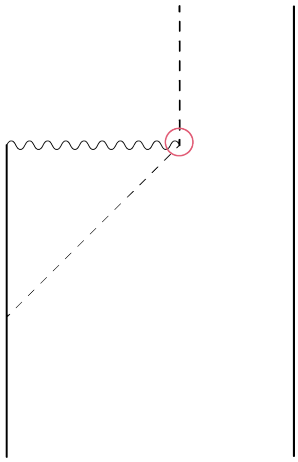


Figure 4.1: The evaporation point of a black hole is, in some sense that we make precise in Sec. 4.4.1, "small".

However, this naïve "swampland censorship" scenario is clearly incomplete. In higher dimensions, the Gregory–Laflamme instability is not excluded by any known swampland constraints; moreover the inclusion of quantum corrections is expected to lead to black hole evaporation, which also violates the WCCC. However, the absence of naked singularities in the observable universe together with the top down arguments from quantum gravity suggest that *some* version of WCCC remains valid: intuitively, a commonality between the Gregory–Laflamme instability and the evaporating black hole is that there is a heuristic sense, which in this article we will make precise, that the naked part of the singularities is "small". See Fig. 4.1. That is, naked singularities can exist, but they must only be visible to asymptotic observers for a "short" period of time. Furthermore, they can only exist in spacetimes that have horizons — so that the bulk of the singularity can remain hidden.

A formulation of such a "quantum gravitational" version of cosmic censorship – which we propose later in this article – and its subsequent proof would require as *sine qua non* both (1) a condition that can guarantee the formation of horizons in broad generality, and (2) a definition of a "small" naked singularity.

Point (1) is a notoriously difficult open problem. Event horizons are by definition teleological: an event horizon is defined with reference to future infinity. Consequently, very few extant results can guarantee the existence of such a global construct even under very restrictive assumptions (and none guarantee that a given event horizon will successfully hide the full extent of a curvature singularity). See [19–21] for a series of failed attempts to construct criteria to this effect. Addressing point (1) would also have the potential to shed light on a number of other research programs, such as thermalization in AdS/CFT (see e.g. [22] for a review) and the firewall problem [6, 23, 24]. Point (2) has seen various approaches (see [25]), but none of those have successfully made contact with the generic existence of horizons, since (1) has thus far remained unresolved.

Here we accomplish both (1) and (2) in the context of AdS/CFT: we give a general guarantee for the existence of event horizons at finite time, and criteria for the size of a singularity that make contact with the guarantee for event horizons. In particular, we classify singularities into three categories: classical, semi-Planckian and Planck-sized, and we motivate

an assumption on the behavior of singularities that would guarantee the existence of horizons at finite time — via point (1) — in spacetimes containing such classical or semi-Planckian singularities. In the rest of this section, we give a rough outline of the central ingredients of (1) and (2).

A General Guarantee for Horizon Formation

Under the standard assumption of global hyperbolicity,[2] which guarantees various consistency results such as unitary invariance of the von Neumann entropy and entanglement wedge nesting, there are various sufficient conditions that may guarantee existence of an event horizon, such as exponential reconstruction complexity or chaos. However, global hyperbolicity *implies* the WCCC as a consequence: almost all results ensuring consistency of the AdS/CFT dictionary assume cosmic censorship — even though it is known that cosmic censorship can be violated in perfectly reasonable AdS/CFT setups (e.g., the evaporating AdS black hole). This strongly suggests that while the WCCC is sufficient for consistency, it is not in fact necessary. We will argue for a more minimalistic form of censorship, which *is* necessary;[3] it remains to be seen whether this version is also sufficient.

Our goal is to find a cosmic censorship conjecture that follows directly from consistency of the AdS/CFT dictionary. Clearly, an essential part of a holographic formulation of cosmic censorship is the holographic dual of an event horizon: we are looking for the boundary dual to the process of bulk horizon formation. This specifically means that we are interested in the existence of a region outside of the causal wedge of the entire asymptotic boundary. The separation between the causal (or simple [27–29]) wedge and the entanglement wedge has been previously linked in [29–31] to the separation between polynomial and exponential quantum complexity classes in the dual theory. Our approach to finding a guarantee for the formation of event horizons will thus focus on reconstruction restrictions by complexity class.

More concretely, we ask when there is a guarantee for the existence of some operator Q in the large-N limit that cannot be reconstructed from knowledge of the causal wedge alone, given that the latter can be reconstructed efficiently.[4]

In [32], two of us provided such a guarantee at large but finite N under certain assumptions. The full statement of the relevant theorem will be reviewed at length in Sec. 4.3.1; here we provide a heuristic statement only, starting with the most significant assumption: that the action of the fundamental time evolution operator U of the system is well modeled on our code subspace by a *pseudorandom* unitary [33].[5] In AdS/CFT, this assumption is valid whenever the time evolution of the CFT on a particular choice of code subspace is well approximated by a pseudorandom unitary at any finite value of N;[6] this is not dissimilar from the assumption of e.g. [34] that the fundamental time evolution of a black hole is well

[2] We really mean AdS hyperbolicity here, which is essentially global hyperbolicity when the asymptotic boundary is included. See [26] for the precise definition.

[3] We hope that this sentence is not taken out of context.

[4] By an efficient reconstruction, we mean one that can be implemented by a quantum circuit whose size is polynomially bounded in the logarithm of the dimension of the code subspace under consideration.

[5] We use "fundamental" to refer to the full quantum description of the gravitational system, as in [31].

[6] We will eventually work at strictly infinite N, but then we will work only with time-evolved states rather than the unitary operator that generates that time evolution, since the aforementioned unitary operator is not well-defined in the strict infinite N limit.

modeled by a two-design or as in [31, 35] by a (pseudo)random unitary. Pseudorandom unitaries are operators drawn from an ensemble $\{U_k\}$ that are indistinguishable from Haar random unitaries by any efficient quantum algorithm. Our results below also hold for Haar random unitaries, but the advantage of pseudorandom unitaries over Haar random unitaries is that by definition they can be constructed by an efficient process. While existence of pseudorandom unitaries has not been explicitly demonstrated yet, candidate constructions have appeared in, e.g., [33]. We will not make use of any such explicit constructions.

Under this first assumption of pseudorandom unitary (PRU) dynamics, the theorem of [32] considers any efficient quantum algorithm $\mathcal{A}$ that:

(a) Prepares polynomially complex states $|\varphi\rangle$ and has access to the action of the time evolution operator U on the expectation values of polynomially complex observables in such states;

(b) Uses the information in (a) to attempt to learn U (or a circuit representation of U), the pseudorandom time evolution operator, thereby producing some operator $\widehat{U}$;

(c) After completing (b), is handed the actual state of the system $|\psi\rangle$ and is asked to predict U acting on $|\psi\rangle$ using the outcome of (b).

The theorem states that on average — and as we will show in Sec. 4.3.2, for typical pseudorandom states and unitaries[7] — any efficient algorithm will fail to produce an accurate guess for U acting on $|\psi\rangle$.[8] That is, an efficient algorithm's best guess $\widehat{U}$ for U will satisfy:

$$|\langle\psi|\widehat{U}^\dagger U|\psi\rangle|^2 \leq 1-\alpha\ , \tag{4.1}$$

where α is an $O(1)$ constant, independent of the size of the Hilbert space in which $|\psi\rangle$ lives. This result in particular guarantees the existence of a distinguishing operator Q whose expectation value differs between $\widehat{U}|\psi\rangle$ and $U|\psi\rangle$ by at least 2α.

We now apply this result to our gravitational setup using the fact that reconstruction is a particular type of learning. We will consider code subspaces $\mathcal{H}_{\text{code}}$ of dimension $e^{O(G_N^a)}$, where $-1 \leq a < 0$, which do not necessarily have horizons in typical states. For example, a microcanonical window of width $O(N)$; another example is a set of states where the leading order contribution to the geometry is the same for all of the states at least for a time of $O(G_N^0)$. We will show that it is easy to prepare candidates for such code subspaces without obviously incurring horizons. We next assume that the time evolution is *gravitationally* pseudorandom: we will define this term more precisely in Sec. 4.3.2, but roughly speaking it denotes an $O(G_N^a)$

[7] Caution: this is not the typical use of 'typical'. Here we will use the qualifier typical specifically for states in the code subspace to which measure concentration applies; this depends on the choice of distribution of states under consideration. For example, we may consider "typical Haar random states" or "typical pseudorandom states". We will always note the distribution under discussion. This nomenclature is not to be confused with the previously used 'generic', which refers to an open subset of the space of solutions to the Einstein equation.

[8] Note that *prima facie*, the fact that we extend Theorem 2 of [32] to any fixed, typical U and ψ appears to contradict some of the statements in the introduction of [32], i.e., that randomness is a *sine qua non* for a quantum learning no-go theorem. This appears to be the case because if $U = U_0$ is fixed, there exists a trivial $\mathcal{A}^U$ that just exactly outputs U_0. The point is that U_0 is here 'atypical', and the algorithm will fail for all other $U \neq U_0$.

amount of pseudorandomness; we find that this amount of pseudorandomness is generally compatible with a $\mathcal{H}_{\text{code}}$ of size $e^{O(G_N^a)}$, which are the code subspaces that we consider. Next, we argue for the existence of an algorithm that, as part of the learning phase, can reconstruct the causal wedge of any efficiently-preparable state in this code subspace. Under certain technical assumptions, such an algorithm is able to output a guess $\widehat{U}$ for the time evolution of the initial state that accurately reproduces the expectation values of all operators in the causal wedge. However, the theorem of [32] guarantees that the fidelity between the predicted time evolution and the actual time evolved state is low, which implies that there exists a distinguishing operator Q whose expectation values differ at $O(1)$ between the actual time evolved state and the predicted time evolution of the state.

Since there exists a learning algorithm that can accurately reconstruct the time evolution of all operators in the causal wedge but cannot accurately reconstruct the expectation value of some operator Q, which we show exists in the large-N limit, we immediately find that the dual to Q lies outside of the causal wedge. Thus, the boundary of the causal wedge is nontrivial and an event horizon must exist.

We call this result *Cryptographic Censorship*:

(Pseudo)random dynamics guarantee event horizon formation in typical (pseudo)random states.

This is the first general condition that guarantees the formation of horizons without requiring asymptotic time evolution, thus addressing the first task above: the prescription of a broadly applicable quantum complexity theoretic diagnostic of event horizon formation.[9] This result can be thought of as a quantum instability of generic asymptotically AdS spacetimes, assuming that typical pseudorandom dynamics are a good model for (sufficiently long) time evolution of generic spacetimes. We may apply this result to our two examples of code subspaces: microcanonical windows of width $O(N)$ and states with an identical leading order geometry at some fixed time. In the case of the former, we find that pseudorandom (respectively, random) states in a given microcanonical window are exponentially (respectively, doubly exponentially) likely to have horizons. In the case of the latter, we find that whenever it is possible to squeeze a gravitational amount of pseudorandomness into an $O(1)$ time band on the boundary, an event horizon must form.

Before we address the second task mentioned above — classifying the size of a singularity — let us motivate why pseudorandom time evolution is natural in strongly gravitating systems. There is a general expectation that some amount of (approximate) randomness is crucial for understanding black hole dynamics: due to scrambling, many aspects of black holes appear to be well modeled by Haar random unitary dynamics or sometimes by the weaker notion of k-designs [34]. The chaotic nature of black hole physics [36–38] and the connection between randomness and chaos [39] have strengthened this expectation. However, we also expect that time evolution in quantum gravity can be efficiently implemented. Haar random unitaries and (exact) k-designs do not have this property. We thus find pseudorandom unitaries to be a well-motivated model of black hole dynamics; indeed, as we shall see, pseudorandom unitary time evolution guarantees the existence of an event horizon (and thus, by definition, of a black hole).

[9] Here we are defining an event horizon as $\partial J^-[\mathscr{I}]$, so this includes holographic cosmologies in AdS.

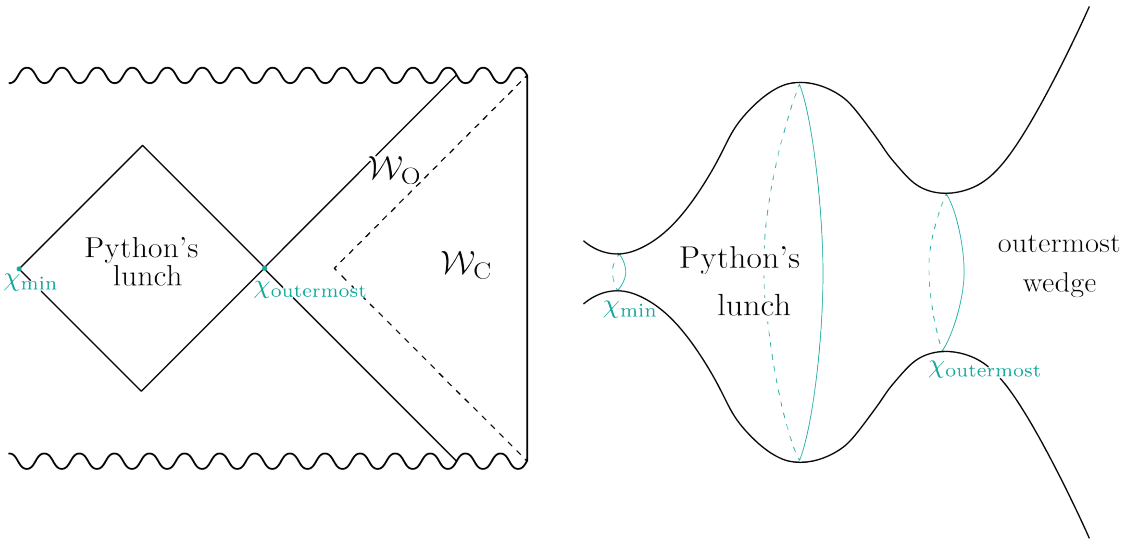


Figure 4.2: A spacetime with a Python's Lunch. In the Penrose diagram on the left, the lunch, causal wedge ($\mathcal{W}_C$) and the wedge of the outermost QES ($\mathcal{W}_O$) are indicated. The left and right green dots are the minimal and outermost QES, respectively. The right panel depicts a timeslice.

Thus far we have argued that pseudorandomness is a valid model for time evolution of strongly gravitating systems using established facts about black holes. We shall postpone motivating pseudorandomness as a model for time evolution in the context of naked singularities until we have defined what we mean by "large" and "small" naked singularities. We therefore proceed to give an overview of this second task.

Size of a Naked Singularity

Intuitively, the naked singularities that we would like quantum gravity to exclude are classically extended with divergent curvatures, and emitting large amounts of observable, causally propagating radiation as prescribed by some fundamental quantum gravity evolution.

In many cases, such large curvatures create (quantum) trapped surfaces and (quantum) extremal surfaces. In [40], unitary invariance of S_{vN} in the dual CFT was used to prove that classical extremal (and trapped) surfaces resulting from classical singularities that do not "evaporate" must lie behind horizons in AdS/CFT. As a warmup, we will prove that even upon inclusion of quantum corrections and under no restrictions on the behavior of the singularity, quantum extremal surfaces (QESs) generically lie behind horizons. The argument makes essential use of the geometrization of regions of high- and low-reconstruction complexity: the causal wedge, and more generally the wedge of the outermost QES, admits an efficient reconstruction — i.e., subexponentially complex in $\log \dim \mathcal{H}_{code}$, where $\mathcal{H}_{code}$ is as before the code subspace on which reconstruction is taking place [30], while the complementary wedge — the so-called Python's Lunch — does not [41]. See Fig. 4.2. In particular, acting on the Python's Lunch is an exponentially difficult (in $\log \dim \mathcal{H}_{code}$) task in the boundary theory. If any QES lies in causal contact of the boundary, it is possible to act on the lunch by throwing in some timelike probe, as illustrated in Fig. 4.3. This is a simple task, and thus

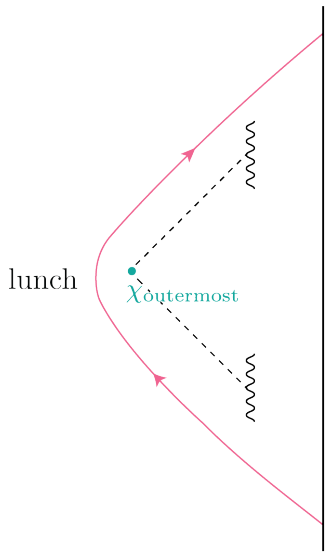


Figure 4.3: A forbidden situation: a QES in causal contact with the boundary.

cannot happen: QESs — minimal or otherwise — must therefore lie outside of the causal wedge[10] of the relevant region.[11]

However, singularities are not always associated with (quantum) trapped or (quantum) extremal surfaces. In these cases, the aforementioned result showing that extremal surfaces imply horizons will not be useful. However, if a singularity could be shown to be associated to a gravitational amount of pseudorandomness, then Cryptographic Censorship ensures that these singularities induce horizons. But how can we know which singularities are "severe enough" to accommodate gravitational pseudorandomness? As a first step towards this question, we give a precise definition of the size of a naked singularity Γ, denoted $|\Gamma|$. This quantity measures the minimal amount of CFT time evolution required to prescribe (or equivalently, obtain) all information about the singularity. For a timelike singularity, we furthermore show that this quantity also measures how many modes within the semiclassical EFT can interact with the singularity. Of course, the dynamics of the singularity itself would not be semiclassical, but we argue that in the $\ell_{\text{Pl}} \to 0$ limit, from the perspective of weakly backreacting probe fields, a timelike singularity might reasonably be conjectured to act primarily as a complicated UV-sensitive boundary condition for their evolution. Then we show that if $|\Gamma|$ does not go to zero as $\ell_{\text{Pl}} \to 0$, or if it goes to zero sufficiently slowly, then the singularity is compatible with inducing gravitational pseudorandomness in the dynamics. With this in mind, we classify singularities into three groups: classical, semi-Planckian, and Planckian. We speculate that classical naked singularities, and possibly semi-Planckian ones, are forbidden by Cryptographic Censorship. Compatibility with gravitational pseudorandomness is of course necessary but

[10]For those readers who have previously seen the Quantum Focusing Conjecture (QFC) used to justify why QESs lie behind horizons [42, 43], note that all such proofs to date have assumed cosmic censorship (in the form of global hyperbolicity). We prove this result using the holographic dictionary instead of using the QFC.

[11]Any readers confused about the islands outside the horizon phenomenon [44, 45] should recall that the horizons of [44, 45] are not the horizons of the system for which the island is quantum extremal but are rather the horizons of a smaller system.

(a) (b)

Figure 4.4: (a) A spacelike singularity with both past and future causal curves to the same conformal boundary $\mathscr{I}$. (b) A simple model of pseudorandom evolution: radiation that is incoming at the naked singularity with some initial phase α_i picks up a pseudorandom phase, such that the final phase α_f is pseudorandomized.

not sufficient to guarantee actual gravitational pseudorandomness in the dynamics. We of course cannot give a top-down derivation that the unitary dynamics of a holographic CFT are approximately pseudorandom on such code subspaces, but we shall motivate this as a reasonable assumption for generic classical timelike singularities. Cryptographic Censorship, together with this expectation, is the basis of our subsequent conjecture of Quantum Cosmic Censorship: that classical (and perhaps semi-Planckian) singularities must be hidden behind horizons; only Planckian singularities may be visible to the asymptotic observer.

Let us now motivate the assumption of pseudorandomness as a model for singularity dynamics (which we have previously motivated for black holes; see also [32, 35]). A natural source of pseudorandom dynamics associated to a singularity is due to the Planckian physics at the singularity itself — what we described above as a UV-sensitive boundary condition. Let us briefly elaborate on this point. Intuitively, we may expect that to semiclassical observers — that is, to observers who are bounded in energy and complexity [31] — the radiation emitted from a timelike singularity or a spacelike singularity (with both past and future causal curves to the same $\mathscr{I}$, see Fig. 4.4a) appears roughly random. See Fig. 4.4b for a simple model.

Of course, the radiation cannot be actually Haar random (as this cannot be efficiently generated), so we instead expect that in many cases, pseudorandomness is an accurate model for the boundary dynamics describing a singularity. When this criterion is satisfied, the theorem of [32] together with the argument outlined above guarantees that the spacetime has an event horizon.[12] It immediately follows that *whenever the fundamental dynamics of a*

[12]Note that our argument does not guarantee that the singularity lies behind the event horizon, nor should it: we expect that small singularities can exist outside of horizons. Our techniques do suggest that it may be possible to limit the extent to which a singularity can lie outside of a horizon, but we leave a detailed investigation of this question to future work.

singularity are well approximated by a pseudorandom unitary, the singularity incurs an event horizon.

It is worth mentioning another potential source of pseudorandom dynamics associated to a singularity: spacetime dynamics in the vicinity of the singularity might induce chaos in the probe fields. The paradigmatic example of chaos in classical gravity — other than event horizons via the butterfly effect — is the BKL phenomenon [46]. This phenomenon, which involves chaotically oscillating metric components, is expected to be generic in the descent to a spacelike singularity [47]. BKL behavior is however *only* established for spacelike singularities; it is possible that certain timelike singularities could induce similar chaos upon approach (see, e.g., [48]), though we remain agnostic on the plausibility of this potential source of pseudorandomness.

We now briefly summarize the paper. We begin in the next subsection by setting up notations and conventions. In Sec. 4.2, we prove that nonminimal QESs must lie behind a horizon. In Sec. 4.3 we prove that when the time evolution in the fundamental description is sufficiently pseudorandom, the bulk spacetime contains an event horizon. We then apply this to the microcanonical window and a code subspace representing states with the same (leading order) geometry for times of $O(G_N^0)$. In Sec. 4.4, we provide a definition that quantifies the size of (the naked part of) a singularity, and we show that classical (and semi-Planckian) singularities can accommodate gravitational pseudorandomness. By Cryptographic Censorship, this would then imply that such singularities must incur horizons. In Sec. 4.4, we also speculate on how Cryptographic Censorship can be used to prove a version of cosmic censorship. We end with a conjecture for a new form of cosmic censorship that is consistent with all known counterexamples to the standard WCCC. In Sec. 4.5, we end with a short discussion of our aspirations and a few possible future directions. In Appendices 4.6.1 and 4.6.2, we provide proof details for the statements of Sec. 4.3.

4.1.1 Preliminaries

Assumptions. Unlike much of the AdS/CFT literature, we will not be imposing the assumption of global hyperbolicity on the bulk spacetimes under consideration, as this is tantamount to assuming weak (and strong) cosmic censorship. This means that, even in the $G_N \to 0$ limit, QFT in curved spacetime (including linearized gravitons) cannot be utilized for time evolution. However, we will still assume that QFT in curved spacetime is valid within causal diamonds that contain no singularities. We will further assume that any spacetime under consideration admits an initial data slice Σ with an everywhere defined and nowhere vanishing continuous timelike unit normal; in particular, we require that $D[\Sigma]$ contains an open set containing Σ. Any singularity in the spacetime will be assumed to act as a UV-sensitive boundary condition that determines the relation between the algebras of different causal diamonds.[13] Finally, we will take a conformal frame which is taken to be fixed at any value of N, including in the large-N limit.

[13]The singularity only has an effect on the relation between two diamonds D_1 and D_2 when the set $[J^+(D_1) \cup J^-(D_1)] \cap [J^+(D_2) \cup J^-(D_2)]$ contains part of the singularity. If this set does not contain a part of the singularity, the EFT can be used to relate the algebra of observables in D_1 and D_2.

Notation and conventions. We now proceed to introduce and review some language and notation from the quantum computing literature, which will be important for the formulation of Cryptographic Censorship. The rest of this section contains notation that will mostly be used in Sec. 4.3. We will use $\mathcal{H}$ to denote Hilbert spaces and for a pure state $|\psi\rangle$, we will use ψ to denote the density matrix $|\psi\rangle\langle\psi|$. When we write $x \leftarrow \mathcal{X}$, we mean a variable x is drawn from a distribution $\mathcal{X}$. For example, by $U \leftarrow \mu$, we mean to draw a unitary U (from the unitary group $\mathbf{U}(\mathcal{H})$) using the Haar measure μ; we will also use μ to denote the distribution of Haar random states, e.g., when we write $|\psi\rangle \leftarrow \mu$. We use $\Pr_{x\leftarrow\mathcal{X}}$ to denote taking the probability of some outcome occurring for x drawn from $\mathcal{X}$. We will often work with a parameter k, which in the cryptographic literature is called a "key". This should be thought of as some small amount of random classical information that can be used to encode or decode data. Concretely, it is a bit string that is sampled uniformly from a *key space* $\mathcal{K}$, a set of bit strings. The dimension of $\mathcal{K}$ scales with a parameter κ, which is frequently referred to in the literature as the "security parameter". We will always have in mind that we are scaling $\log\dim\mathcal{H} = \text{poly}(\kappa)$ and will eventually take the $\kappa\to\infty$ limit. A function $f(x)$ is said to be *negligible* in its parameter x if for any constant $c>0$, $|f(x)| \leq x^{-c}$ for any sufficiently large x. We will say that a quantity $f(x)$ is $O(g(x))$, denoted by $f(x)\sim O(g(x))$, if for some positive constants c_1, c_2 and for any sufficiently large x, $c_1 g(x) \leq |f(x)| \leq c_2 g(x)$.[14] Finally, we will call an operator V *approximately norm-preserving* if it approximately preserves state norms, i.e., for any state $|\psi\rangle$, $\langle\psi|V^\dagger V|\psi\rangle$ is exponentially close to 1 for sufficiently large κ. This of course includes any unitary operators.

Our results will make heavy use of quantum algorithms, which we now define. A quantum algorithm $\mathcal{A}$ is a map from an input Hilbert space $\mathcal{H}_{\text{in}}$ to some output set $\mathcal{C}$. Note that the set $\mathcal{C}$ can be a Hilbert space, a classical space, or some other set altogether. For example, a quantum algorithm could take as input a quantum state $|\psi\rangle\in\mathcal{H}_{\text{in}}$ and output a classical bit representation of an operator $\mathcal{O}$ (e.g., the coefficients of the representation of $\mathcal{O}$ in some fixed choice of basis). A quantum algorithm $\mathcal{A}$ acting on some $\mathcal{H}_{\text{in}}$ is said to be *efficient*, *polynomially complex*, or equivalently *computationally bounded* if it can be implemented by a quantum circuit of size no greater than polynomial in $\log\dim\mathcal{H}_{\text{in}}$.[15] Here we will typically take $\mathcal{H}_{\text{in}}$ to be the code subspace under consideration. We will often drop the reference to $\log\dim\mathcal{H}_{\text{in}}$ when describing complexity or efficiency of such algorithms.

Consider now an efficient algorithm designed to take in a state $|\psi\rangle$ and output 0 if it determines that $|\psi\rangle$ was sampled from distribution μ_1 and 1 if it determines $|\psi\rangle$ came from μ_2. In this case, we say the two ensembles are computationally indistinguishable — which we define precisely for general algorithms below — if any efficient algorithm succeeds at this task with probability upper bounded by 1/2 up to negligible corrections. It should be clear that this is equivalent to saying that the quantum algorithm outputs 1 (or equivalently 0) with almost the same probability when acting on both μ_1 and μ_2.

Note that the authors of [33], whose definitions of pseudorandom unitaries and states we will use, demand that pseudorandom ensembles are indistinguishable from the Haar ensemble even if the algorithm is allowed a reasonable (polynomial) number of copies of the state.

[14] In computer science literature this is denoted by $f=\Theta(g(x))$.

[15] This involves some canonical choice of gates; however, this choice does not change the scaling in $\log\dim\mathcal{H}_{\text{in}}$.

Having multiple copies of the state is essential to allow the algorithm to probe coherent information about the quantum state.

With this intuition, we present the exact definition of computational indistinguishability of ensembles of states for general quantum algorithms:

Definition 3 (Computational Indistinguishability)**.** *Two ensembles of states, $\mathcal{S}_1$ and $\mathcal{S}_2$, in some Hilbert space $\mathcal{H}$, with size parameterized by a parameter κ, are* computationally indistinguishable *if for any efficient quantum algorithm $\mathcal{A}$ and for all $m = \text{poly}(\kappa)$,*

$$\left| \Pr_{|\psi_1\rangle \leftarrow \mathcal{S}_1} \left(\mathcal{A}(|\psi_1\rangle^{\otimes m}) = 1\right) - \Pr_{|\psi_2\rangle \leftarrow \mathcal{S}_2} \left(\mathcal{A}(|\psi_2\rangle^{\otimes m}) = 1\right) \right| = \text{negl}(\kappa) \ . \tag{4.2}$$

When we write $\mathcal{A} = 1$, we are considering that the algorithm is given some binary task (e.g., deciding whether a state is pseudorandom or not), and outputs either 0 or 1; we write $\mathcal{A} = 1$ to indicate that it outputs 1.

The specific quantum algorithms that we will consider are learning algorithms; these are efficient algorithms that take certain data as input, implement some processing with the goal of learning some outcome, and finally produce their best guess for the outcome. In our particular context, the learning algorithms will aim to learn a circuit (or algebraic) representation of a unitary operator U. During the initial learning phase, the algorithms are given "oracle access" to some unknown unitary U, denoted by $\mathcal{A}^U$: this means that the algorithms have the ability to query U a polynomial number of times on a polynomial number of (efficiently preparable) states $|\varphi_i\rangle \in \mathcal{H}$ to obtain $U|\varphi_i\rangle$ during its learning phase. During this phase, the algorithms can perform any polynomially complex computation between queries, on both the $U|\varphi_i\rangle$ from its queries thus far, and on its own registers. At the end of this phase, the algorithms' oracle access to U is terminated, and they are tasked with producing a maximally accurate representation of U or a particular aspect or feature of it: e.g., U acting on some initial state $|\psi_{\text{initial}}\rangle$ which would now be given to the algorithm.

Let us briefly make a note on the different parameters and dimensions used throughout this paper. In the context of AdS/CFT, we will be working with code subspaces of dimension $\dim \mathcal{H}_{\text{code}} = 2^\kappa$, where the security parameter $\kappa = \text{poly}(N)$. For concreteness, we will consider $G_N \sim 1/N^2$, but the relation between G_N and N can be changed to the reader's liking. We will often use N instead of κ to parameterize functions, e.g., for negligible functions we will write $\eta(N)$.

Note that theorems and proofs in this paper are well above the physics level of rigor but fall short of a mathematician's level of rigor in a few places. We identify these gaps where possible and leave a fully rigorous formulation thereof to the curious mathematician.

4.2 Warmup: QESs Lie Behind Horizons

We will show in this section that if χ is a nonempty QES homologous to $\mathscr{I}$ in a spacetime (M, g), then it lies behind both past and future event horizons. The quantum focusing conjecture (QFC) [49] with reflecting boundary conditions at $\mathscr{I}$ guarantees by the quantum singularity theorem [50] that at least one null geodesic fired from χ is geodesically incomplete.[16]

[16] Strictly speaking, the restricted QFC [51] is sufficient to guarantee this.

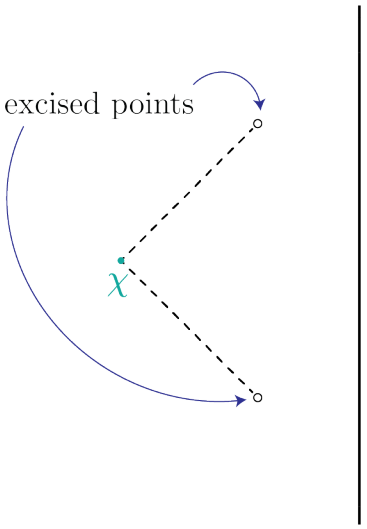


Figure 4.5: An example where null geodesics fired from χ are geodesically incomplete, but there is no horizon.

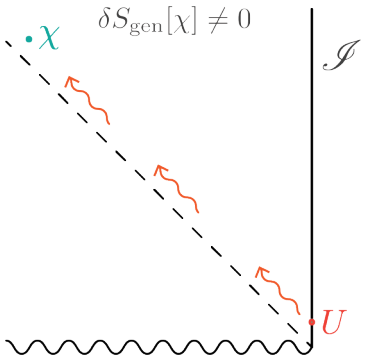


Figure 4.6: A QES outside the horizon, leading to a contradiction with the unitary invariance of von Neumann entropy: $S_{\text{vN}}[\rho_{\mathscr{I}}] = S_{\text{vN}}[U^\dagger \rho_{\mathscr{I}} U]$.

However, in the absence of a stringent causality assumption such as AdS hyperbolicity, null geodesic incompleteness does not guarantee the existence of horizons. See Fig. 4.5 for an illustration.

When χ is the minimal (and nonempty) QES for $\mathscr{I}$ in M, it immediately follows from the QES formula relating $S_{\text{gen}}[\chi]$ to $S_{\text{vN}}[\rho_{\mathscr{I}}]$ that bulk perturbations locally propagating from $\mathscr{I}$ (future- or past-directed) cannot modify $S_{\text{gen}}[\chi]$ and thus cannot reach χ: in this case, $\chi \cap I^{\pm}[\mathscr{I}] = \varnothing$.[17] See Fig. 4.6 for a review of the argument, originally presented in [42].

However, from the perspective of cosmic censorship, global minimality of a QES is of little relevance. We expect that a similar result should apply to any QES independently of its global properties. In [40] some strides were made towards this end by CPT-conjugating the spacetime around the nonminimal QES χ to generate a new spacetime in which χ was minimal. This allowed [40] to prove that a large class of extremal surfaces and trapped

[17]Note that we have excluded the future horismos $\partial I^{\pm}[\mathscr{I}]$ from this conclusion, as $i^{\pm}$ are not precluded from null communication with χ.

surfaces must lie behind horizons. These results, however, relied on an assumption that we aim to relax in this section: that there can be no "evaporating singularities" (see [40] for the precise definition).

To relax this assumption, we will instead assume the Python's Lunch proposal [29, 41, 52] in addition to the QES formula to argue that (quantum) extremal surfaces must generically be hidden behind both past and future horizons.[18] Or more precisely, if they lie outside a horizon, this distance must shrink to zero in the $G_N \to 0$ limit. Let us briefly review some facts about the Python's Lunch proposal. A Python's Lunch L with respect to a conformal boundary $\mathscr{I}$ is a region of spacetime defined by (at least) three mutually spacelike separated QESs homologous to $\mathscr{I}$: χ_{min}, χ_{bulge}, and χ_{aptz}. The first is the globally minimal QES, while χ_{aptz} is a candidate QES, but having larger generalized entropy than χ_{min}. The surfaces χ_{min} and χ_{aptz} are so-called throats: QESs whose generalized entropy is roughly locally minimal in space and locally maximal in time (see [52] for details). The bulge surface χ_{bulge} is also a QES, but it is locally maximal in time and space (see again [52] for details). The region L is defined as the domain of dependence of any spacelike slice bounded by χ_{min} and χ_{aptz}, and the Python's Lunch proposal conjectures that the CFT computational complexity required to implement any operator in L scales as

$$\exp\left[\frac{S_{\text{gen}}[\chi_{\text{bulge}}] - S_{\text{gen}}[\chi_{\text{aptz}}]}{2}\right]. \tag{4.3}$$

We will only identify a Python's Lunch in a portion of spacetime that is globally hyperbolic, and we will make the reasonable assumption that the presence of potential spacetime singularities away from L does not make operator reconstruction any easier.

Besides the Python's Lunch proposal, our assumptions in this section (and this section only) are as follows.

Assumptions and Definitions. We assume the following in addition to our general assumptions stated at the beginning of Sec. 4.1.1:

- (M, g) has reflecting boundary conditions at each connected boundary component $\mathscr{I}$.
- Smooth initial data: (M, g) has a smooth inextendible timeslice Σ_0 containing two QESs χ_{min} and χ homologous to $\mathscr{I}$, and where χ_{min} is the QES of minimal generalized entropy that is homologous to $\mathscr{I}$.
- A generic condition: if χ' is a bulge QES and χ'' a throat QES, then their associated generalized entropies satisfy $S_{\text{gen}}[\chi'] - S_{\text{gen}}[\chi''] \sim O(G_N^{-1})$.
- In any globally hyperbolic region of spacetime that we consider, the (restricted) QFC holds, and the quantum maximin [26, 53] and maximinimax [41] prescriptions converge.
- We denote by $W_O[\chi]$ the outer wedge of a QES χ [28, 54].
- We define $t = 0$ at $\mathscr{I}$ as $\Sigma_0 \cap \mathscr{I}$.

[18] Note that the assumption of the Python's Lunch proposal applies to this section only; it will not be used in the proof of Cryptographic Censorship in Sec. 4.3.3.

- If a point p in the bulk is such that any future (past) directed causal curve from p to $\mathscr{I}$ arrives at $\mathscr{I}$ at a boundary time that diverges to $+\infty$ ($-\infty$) in the $G_N \to 0$ limit, or if no curve to $\mathscr{I}$ exists, we say that p is behind an "effective future (past) horizon".

Theorem 6. *Let (M, g) be a spacetime as above. Assuming the Python's Lunch conjecture for complexity of operator reconstruction, the QES χ lies behind past and future effective horizons.*

Proof. Consider first working strictly within the globally hyperbolic spacetime given by the domain of dependence of Σ_0 within M, denoted as $D[\Sigma_0]$. By Proposition 1 of [30], there always exists an outermost minimal QES χ_{aptz} that lies in $W_O[\chi] \cap W_O[\chi_{\text{min}}]$.[19] This surface may or may not coincide with χ itself. Let now H be a spatial slice with $\partial H = \chi_{\text{aptz}} \cup \chi_{\text{min}}$, which exists since $\chi_{\text{aptz}} \subset W_O[\chi_{\text{min}}]$, and let $L = D[H]$. See Fig. 4.7 for an illustration. By the maximinimax construction in the appendix of [41], there is another QES χ_{bulge} in L that is of the bulge type, and so L must be a Python's Lunch ($\chi_{\text{bulge}} = \chi$ is allowed). L must contain χ, since there exists a choice of H that contains χ_{min}, χ, and χ_{aptz}. If χ is equal to χ_{aptz} or χ_{bulge}, this is trivial. If χ is neither of these surfaces, then we construct H by gluing together two pieces. The first piece is the subset of Σ_0 lying between χ_{min} and χ. The second is any spacelike slice between χ and χ_{aptz} which exist by global hyperbolicity of $D[\Sigma_0]$ and the fact that χ_{aptz} lies strictly in the outer wedge of χ.

Let us now return to the full spacetime M. Define a code subspace $\mathcal{H}_{\text{code}}$ consisting of all low energy perturbative quantum states that backreact on Σ_0 at subleading order in G_N. By the genericity assumption, we have $S_{\text{gen}}[\chi_{\text{bulge}}] - S_{\text{gen}}[\chi_{\text{aptz}}] \sim O(G_N^{-1})$ so by the Python's Lunch proposal, the complexity of reconstructing operators in $\mathcal{H}_{\text{code}}$ with support in L scales as $\exp\left(cG_N^{-1}\right)$ for some constant $c > 0$. Assume now for contradiction that χ lies outside the effective past horizon. Then there exists a future-directed causal curve γ from $\mathscr{I}$ to χ that is fired from $\mathscr{I}$ at a time $t \sim O(G_N^0)$. Then we can find a simple operator Φ that modifies the quantum fields at χ. Just let Φ be an operator that injects a perturbative amount of matter with rocket boosters near $\mathscr{I}$, so that the matter travels along the curve γ. Because this matter reaches χ in a time that does not scale with G_N, this matter backreacts on Σ_0 only at subleading order, and so it acts within the code subspace. Φ can be constructed in two steps: an injection of matter near the boundary using HKLL [55–57] followed by an $O(G_N^0)$ amount of CFT time evolution. Neither is exponentially hard, and so we reach a contradiction with reconstruction complexity of operators with support in L. The argument excluding χ from intersecting the exterior of the effective future horizon proceeds *mutatis mutandis*. □

Note that there are setups where islands [44, 45], and thus QESs χ, lie outside of event horizons. A universal aspect of these constructions is that the QES in question is a QES of a different subsystem than the one whose horizon is under consideration. We may also comment more specifically on each example: first, Ref. [44] found an island outside of the black hole horizon when the CFT at $\mathscr{I}$ was coupled to a reservoir. In this case boundary conditions were not reflecting, which was assumed in our result. Furthermore, by virtue of the coupling, time evolution of the black hole system alone is no longer unitary. Thus, while

[19] Outermost minimal means that there is no surface in the outer wedge $W_O[\chi_{\text{min}}]$ homologous to χ_{min} with less S_{gen}.

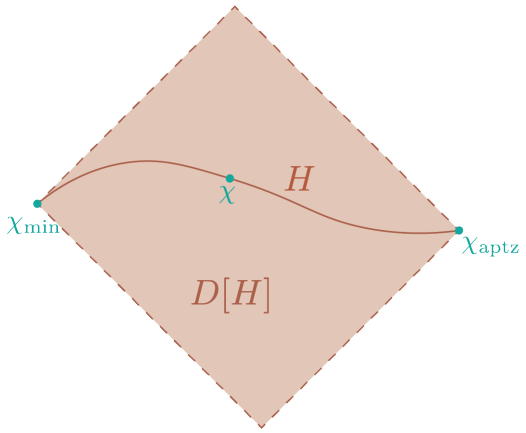


Figure 4.7: The setup in the proof of Theorem 6 where χ is nonminimal, and where H is chosen to contain χ.

a signal sent from $\mathscr{I}$ initially can be turned near the boundary with a unitary, propagating it to χ requires a non-unitary operation on the CFT, and so the entropy is expected to change. Additionally, decoupling the CFT from the auxiliary system produces a shockwave that alters the spacetime at leading order, which does push the island behind the horizon. Next, in [45] an island was found lying a distance of $O(\sqrt{\ell_{\text{Pl}} r_{\text{hor}}})$ outside the horizon. In this case the island was associated to angular momentum modes in the bulk, but qualitatively this construction is the same as Ref. [44] as far as it relates to the above theorem.

4.3 Crytographic Censorship

In this section we introduce and prove our main result: that on any sufficiently large code subspace, (pseudo)random dynamics guarantee event horizon formation in typical (pseudo)random states with a semiclassical bulk dual. The precise statement (Theorem 10: Cryptographic Censorship) is given in Sec. 4.3.3. In presenting and proving this theorem, we will be agnostic about the types of code subspaces that in a gravitational theory could correspond to this degree of pseudorandomness; in Sec. 4.3.4 we will give two examples of code subspaces in gravity that are compatible with sufficiently pseudorandom dynamics.

Intuitively, pseudorandom unitary ensembles are ensembles of unitaries that are *computationally indistinguishable* from the Haar ensemble. As an example, upon evolution for some time longer than a scrambling time, a chaotic system is generally expected to become approximately pseudorandom. Our argument for the existence of a horizon hinges upon the assumption that the fundamental time evolution is well modeled by a unitary that is sufficiently pseudorandom on the code subspace. In a sense described below, we will require the unitary to be pseudorandom by an amount that scales inversely with G_N, which we will call *gravitationally* pseudorandom.

How is this assumption of gravitational pseudorandomness sufficient to guarantee the existence of event horizons? We will outline the logic here while omitting many of the technicalities (which we leave to Apps. 4.6.1 and 4.6.2); the rest of this section is devoted to the technical points necessary for the proof of Cryptographic Censorship and its application

to gravitational code subspaces.

Recall that observables in the causal — and more generally, simple [28, 30, 54] — wedge of the boundary are computationally simple to reconstruct. This is a non-trivial aspect of bulk reconstruction in AdS/CFT [28, 58].[20] We argue that, given some reasonable technical assumptions outlined in Sec. 4.3.3, there exists a learning algorithm $\mathcal{A}_{CW}$ that can reconstruct the causal wedge of any geometric state in a code subspace of size $e^{O(G_N^a)}$ for some $-1 \leq a < 0$. However, a theorem of [32] (see Theorem 7 below for its precise statement) combined with results from measure concentration guarantees (roughly) that all efficient quantum algorithms must fail at predicting the expectation value of some operator Q, which we show exists in the large-N limit, after pseudorandom time evolution. It then immediately follows that for the algorithm $\mathcal{A}_{CW}$, there exists an operator Q that $\mathcal{A}_{CW}$ cannot reconstruct and therefore must lie outside of the causal wedge; in other words, an exterior to the causal wedge exists — by extension, a boundary to the causal wedge exists. Since the boundary of the causal wedge of $\mathscr{I}$ coincides with the event horizon, an event horizon must exist. In Sec. 4.3.3, we will formalize this rough argument into a proof.

We start by briefly reviewing pseudorandom unitaries, as well as the appropriate learning result in [32] that we will use, in Sec. 4.3.1. We then explain the required alterations and extensions necessary to apply the learning result to the holographic, $N \to \infty$ setting in Sec. 4.3.2. In Sec. 4.3.3 we state and prove our theorem of Cryptographic Censorship. Finally, in Sec. 4.3.4 we comment on two code subspaces to which our theorem applies.

4.3.1 Review: pseudorandomness and complexity of learning

Pseudorandom states and unitaries will play an important role in what follows, so we will briefly review these concepts. Within the context of quantum cryptography, pseudorandom states were first defined in [33]. Informally, an ensemble of quantum states is pseudorandom if the states are efficiently constructible, but indistinguishable from Haar random by any polynomially bounded quantum algorithm — even if that algorithm is given polynomially many copies of the state. A pseudorandom state generator G is a map from some space $\mathcal{K}$ into a Hilbert space $\mathcal{H}$. For each element $k \in \mathcal{K}$, it outputs $G(k) = |\psi_k\rangle \in \mathcal{H}$. The intuitive goal of the definition is to say that the action of feeding a random sample from $\mathcal{K}$ to G to get a state $|\psi_k\rangle$ is indistinguishable from sampling the Haar random distribution of states (in the computational sense defined in Sec. 4.1.1). The technical definition is as follows:

Definition 4 (Pseudorandom State Ensemble (PRS); Definition 2 in [33])**.** *Consider an ensemble of states $\Psi = \{|\psi_k\rangle \in \mathcal{H}\}_{k\in\mathcal{K}}$, labeled by a parameter k drawn from a key space $\mathcal{K}$, where both $\mathcal{H}$ and $\mathcal{K}$ are parameterized by a security parameter κ. Then the ensemble Ψ is a pseudorandom state ensemble if*

- *The states can be efficiently generated: there is an efficient quantum algorithm G that generates the state $|\psi_k\rangle$ given k as input. That is, for all $k \in \mathcal{K}$, $G(k) = |\psi_k\rangle$.*

[20] For the likely limited number of readers who believe in the Quantum Extended Church-Turing (QECT) thesis outside of horizons but are skeptical of arguments relying on AdS/CFT, we note that the QECT thesis by itself also implies that causal observables must be simple.

- *The states are computationally indistinguishable from Haar random: for any polynomially many copies of $|\psi_k\rangle$ with the same $k \in \mathcal{K}$ and any efficient quantum algorithm $\mathcal{A}$, there exists a negligible function* $\mathrm{negl}(\kappa)$ *such that*

$$\left| \Pr_{k\leftarrow\mathcal{K}} \left[\mathcal{A}\left(|\psi_k\rangle^{\otimes \mathrm{poly}(\kappa)}\right) = 1\right] - \Pr_{|\psi\rangle\leftarrow\mu} \left[\mathcal{A}\left(|\psi\rangle^{\otimes \mathrm{poly}(\kappa)}\right) = 1\right] \right| = \mathrm{negl}(\kappa) \ , \tag{4.4}$$

  where μ is the Haar measure on $\mathcal{H}$.

We can define pseudorandom unitaries as in [33] in a similar fashion:

Definition 5 (Pseudorandom Unitary Ensemble (PRU); Definition 5 in [33])**.** *Consider an ensemble of unitaries, $\mathcal{U} = \{U_k \in \mathbf{U}(\mathcal{H})\}_{k\in\mathcal{K}}$, labeled by a parameter k drawn from a key space $\mathcal{K}$, where both $\mathcal{K}$ and $\mathcal{H}$ are parameterized by a security parameter κ. We say $\mathcal{U}$ is a pseudorandom unitary ensemble if*

- *The unitaries are efficiently representable: there is an efficient quantum algorithm Q such that for every $k \in \mathcal{K}$ and any $|\psi\rangle \in \mathcal{H}$, $Q(k, |\psi\rangle) = U_k |\psi\rangle$.*
- *The unitaries are computationally indistinguishable from Haar random: for any efficient quantum algorithm $\mathcal{A}^U$ given oracle access to some U, there exists a negligible function* $\mathrm{negl}(\kappa)$ *such that*

$$\left| \Pr_{k\leftarrow\mathcal{K}} \left[\mathcal{A}^{U_k} = 1\right] - \Pr_{U\leftarrow\mu} \left[\mathcal{A}^{U} = 1\right] \right| = \mathrm{negl}(\kappa) \ . \tag{4.5}$$

We note that technically, PRSs and PRUs are actually sequences of ensembles, $\Psi = \{\Psi_\kappa\}_\kappa$ and $\mathcal{U} = \{\mathcal{U}_\kappa\}_\kappa$, respectively, each parametrized by $\kappa \in \mathbb{N}$, and the guarantees in the above definitions are for large κ. For ease of notation, we will instead simply write Ψ and $\mathcal{U}$, as in the above definitions, where the key k should be understood to be drawn from a key space $\mathcal{K}_\kappa$ where κ is sufficiently large. We will make an exception to this rule in some parts of the proof in App. 4.6.2, where it will be convenient to make the dependence on the security parameter explicit to take the $\kappa \to \infty$ limit.

Finally, we remark that while explicit constructions of PRSs exist (and were given first in [33]), there are currently no known constructions of PRUs, although candidate constructions were proposed in [33]. We will not require any explicit constructions of either PRSs or PRUs in this work.

Pseudorandom states vs. k-designs. For the reader familiar with the related notion of a k-design, we briefly comment on the two main differences with PRSs. First, the associated notion of indistinguishability is much stronger for k-designs: the states need to be indistinguishable from Haar random even to computationally *unbounded* observers, e.g., algorithms of exponential complexity. However, k-designs only require that the moments of these ensembles are indistinguishable up to a certain value (k), whereas PRS ensembles are computationally indistinguishable for any $\mathrm{poly}(\kappa)$ number of moments. In this sense, PRSs are comparable to a high-order k-design but with a parametrically weaker notion of

indistinguishability. Note that k-designs are less relevant for our purposes as they apply to unbounded observers, whereas our focus here is on bounded computations.

We now state the learning result of [32], which was briefly reviewed at a lower level of detail in Sec. 4.1. This result (and our Cryptographic Censorship result) also holds for Haar random U and $|\psi\rangle$, but we will focus on pseudorandom states and unitaries as we believe they present a more realistic characterization of the time evolution of physical systems.

Theorem 7 (Complexity of Learning Pseudorandom Unitaries, Theorem 2 of [32])**.** *Let $\kappa \in \mathbb{N}$ be the security parameter, and let $\mathcal{U}$ be any pseudorandom unitary ensemble and Ψ be any pseudorandom state ensemble in a Hilbert space $\mathcal{H}$ where $|\mathcal{H}| = 2^\kappa$. For any quantum algorithm $\mathcal{A}$ of $\mathrm{poly}(\kappa)$ complexity that, given oracle access to the pseudorandom unitary $U \leftarrow \mathcal{U}$, produces (a circuit representation of) an approximately norm-preserving operator[21] $\widehat{U}$ as its best guess for U, there exists a constant $\alpha > 0$ independent of κ such that for any $\kappa > \kappa_0$, the following quantity is bounded:*

$$\operatorname*{avg}_{\substack{U\leftarrow\mathcal{U}\\ |\psi\rangle\leftarrow\Psi\\ \widehat{U}\leftarrow\mathcal{A}^U}} \left[F\left(\widehat{U}\,|\psi\rangle\,, U\,|\psi\rangle\right)\right] \leq 1-\alpha\ .$$

Here, $F(\widehat{U}\,|\psi\rangle\,, U\,|\psi\rangle) = |\left\langle \psi|\widehat{U}^\dagger U|\psi \right| \left. \psi|\widehat{U}^\dagger U|\psi \right\rangle|^2$ which is the fidelity when $\widehat{U}$ is unitary. In this context, it is approximately the fidelity since we are working with approximately norm-preserving operators.

This theorem shows that it is impossible for any efficient quantum learning algorithm to use oracle access to a pseudorandom U to produce a polynomially complex circuit or (succinct) algebraic expression that accurately implements U on a pseudorandom state $|\psi\rangle$. As we will explain in more detail in the next section, we will be interested in applying this result to a sequence of code subspaces $\{\mathcal{H}^{(\kappa)}_{\text{code}}\}_\kappa$ each of dimension $|\mathcal{H}^{(\kappa)}_{\text{code}}| = 2^\kappa$, where we will take $\kappa = \mathrm{poly}(N)$.

We now comment on two aspects of this theorem that will appear extensively in the subsequent section. First, Theorem 7 implies that, *on average*, there exists a distinguishing operator that the learning algorithm fails to predict (see Appendix B in [32] for an explanation).[22] In particular, when the fidelity between $\widehat{U}\,|\psi\rangle$ and $U\,|\psi\rangle$ is $1 - O(1)$, then there exists an operator Q such that[23]

$$\operatorname*{avg}_{\widehat{U}\leftarrow\mathcal{A}^U} \left|\mathrm{tr}\Big(Q\widehat{U}\psi(\widehat{U})^\dagger\Big) - \mathrm{tr}\big(QU\psi(U)^\dagger\big)\right| \geq O(1)\ . \tag{4.6}$$

[21] We note that [32] showed this result for a wider class of algorithms — those producing any operator $\widehat{U}$ with column norms bounded above by ≈ 1 — which includes the algorithms here. We will not need this more general result in this paper.

[22] As a difference of language, [32] writes this as the distinguishing "POVM" and lower bounds the distance between measurement outcomes obtained via this POVM.

[23] Note that the quantum algorithm $\mathcal{A}^U$, as l result of, e.g., intrinsic randomness in the outcome of measurements during the learning process, generally outputs a $\widehat{U}$ that is not a deterministic function of U. Therefore, we average over the algorithm's output $\widehat{U} \leftarrow \mathcal{A}^U$.

In the next section, we will use measure concentration to show that for *nearly all* of the unitaries in the ensemble $\mathcal{U}$, such a fidelity bound holds and thus there is a distinguishing operator.[24] This operator will play a crucial role in the proof of Cryptographic Censorship.

Second, it immediately follows from Theorem 7 that

$$\sup_{\kappa\to\infty} \underset{\substack{U\leftarrow\mathcal{U}\\ |\psi\rangle\leftarrow\Psi\\ \widehat{U}\leftarrow\mathcal{A}^U}}{\text{avg}} \left[F\left(\widehat{U}\,|\psi\rangle\,, U\,|\psi\rangle\right)\right] \leq 1-\alpha\ , \tag{4.7}$$

which crucially depends on the upper bound of $1-\alpha$ being independent of κ. This will be important in extending and applying this result to holography, where we will take the large κ (or equivalently, large N) limit, and want the guarantee that the large N limits of these states also have fidelity at most $1-\alpha$, and thus a distinguishing operator exists.

4.3.2 Complexity of learning in holography

We now turn to the application of this learning result to holography. Since our goal is to constrain semiclassical gravitational phenomena, our focus will be on the large-N limit in general and on quantities that scale with N in this limit in particular. For concreteness, and without loss of generality, we use $G_N \sim 1/N^2$, and consider code subspaces $\mathcal{H}_{\text{code}}$ of dimension $e^{O(G_N^a)}$, where $-1 \leq a < 0$. To this end, we will take the security parameter $\kappa = \text{poly}(N)$ such that our code subspace $\mathcal{H}_{\text{code}}$ is of dimension 2^κ. Unitary operators that are pseudorandom on such code subspaces at any large but finite N will be called gravitationally pseudorandom (defined precisely below); we will also resolve various technical challenges that follow from taking the large-N limit (e.g., with regards to the existence of certain quantities). We begin with a formal definition of gravitational pseudorandomness.

Definition 6 (Gravitationally Pseudorandom Unitary Ensemble (GPRU))**.** *Let* $\mathcal{U} = \{U_k \in \mathbf{U}(\mathcal{H})\}_{k\in\mathcal{K}}$ *be a pseudorandom unitary ensemble, with key space* $\mathcal{K}$ *parameterized by a security parameter* κ*, where* $\mathcal{H}$ *is the Hilbert space of a large* N*, holographic quantum system where* $N \sim G_N^{-1/2}$*, with* G_N *the bulk Newton's constant. We say that* $\mathcal{U}$ *is a* gravitationally pseudorandom unitary ensemble (GPRU) *if* $\kappa = \text{poly}(N)$*.*

As noted previously, at strictly infinite N there is no notion of pseudorandom unitaries. When discussing operators we shall be primarily interested in asymptotic scaling at large N rather than the actual operators at infinite N. When working in the actual infinite N regime, we will only discuss limits of pseudorandom *states*.

Recall that we aim to apply the results of [32] to the holographic setup. The careful reader may have by now noticed two potential points of concern: first, Theorem 7 was proved on average for a PRU ensemble rather than for some typical instance; second, it is not obvious that in the large-N limit, for a typical PRU, a distinguishing operator exists. Let us briefly elaborate on each concern before proceeding to state and prove our generalization of Theorem 7 that addresses both.

[24] The result on average in [32] only implies that a constant fraction of the unitaries in $\mathcal{U}$ have such an operator, whereas we will show this holds for a fraction of the unitaries that goes to 1 as $\kappa \to \infty$.

On the first concern, the learning result of [32] is a statement about the *average* fidelity between the evolution of a pseudorandom state (PRS) evolved with the actual PRU and the same state evolved with the algorithm's best guess. The averaging is over the unitary drawn from any PRU ensemble $\mathcal{U}$, the state drawn from any PRS ensemble Ψ, and over the output of any quantum learning algorithm $\mathcal{A}$.

While much of the literature (see, e.g., for a few instances [34, 59, 60]) does model gravitational systems via randomly sampled states or dynamics, this approach is in some tension with the fact that there is just *one* (fixed) unitary operator that describes the time evolution of $\mathcal{N} = 4$ SYM. We therefore generalize the bound from an average to a *typical* pseudorandom unitary and state by bounding the fluctuations of typical members of the ensemble, first for Haar random unitaries and states and then for pseudorandom unitaries and states. We emphasize that this means a learning bound applies to any *single fixed* unitary and state that are both typical, as we state below in Theorem 8.

The second concern, involving the existence of a "non-simple observable" — i.e., the distinguishing observable that the algorithm fails to predict — for typical instances of the ensemble at large-N, is a bit more subtle. As a corollary to our theorem for learning typical pseudorandom unitaries at large but finite N, we prove, under a reasonable assumption on the limit of states as $N \to \infty$, that our theorem also bounds the fidelity at $N \to \infty$. Consequently, our theorem guarantees that for any simple boundary algorithm, there exists a *semiclassical* observable that cannot be reconstructed by the algorithm.

We now state the holographic version of Theorem 7, incorporating the generalizations advertised above. We will subsequently use this version in Sec. 4.3.3 to prove Cryptographic Censorship.

Consider a CFT with a large but finite UV energy cutoff, so that when N is finite the dimension of the Hilbert space is finite.[25] We prove the theorem below for any large but finite value of N, under the assumption that the algorithm's output as a function of U is (randomized) K-Lipschitz for some fixed constant K. We will describe this assumption in greater length in App. 4.6.1, but it boils down to a certain notion of continuity controlled by a parameter K of the algorithm as a function of U.[26] We subsequently state and prove a corollary of this theorem for infinite-dimensional Hilbert spaces, under a class of assumptions. To increase readability, we postpone both of the proofs to Apps. 4.6.1 and 4.6.2.

Theorem 8 (Complexity of Learning for Typical Pseudorandom Unitaries)**.** *Let $\kappa = \text{poly}(N)$ and consider a code subspace of dimension 2^κ. Consider any (gravitationally) pseudorandom unitary ensemble $\mathcal{U}$, any pseudorandom state ensemble Ψ, and any constant $\epsilon > 0$. For any (randomized) K-Lipschitz quantum algorithm $\mathcal{A}$ of $\text{poly}(N)$ complexity that has oracle access to $U \leftarrow \mathcal{U}$ and produces an approximately norm-preserving operator $\widehat{U}$ as its best guess for U, and for any sufficiently large N, the following holds: for a fraction $\geq 1 - e^{-\epsilon^2 2^\kappa/\left(48(1+K)^2\right)} - e^{-\epsilon^2 2^\kappa/192} - \text{negl}'(N)$ of $U \in \mathcal{U}$ and $|\psi\rangle \in \Psi$, we have that:*

[25] We are free to consider other ways of regulating the dimension of the Hilbert space if needed.

[26] It is possible that this assumption actually follows directly from our other assumptions; we leave determination of this question to future work.

1. *the following quantity is bounded*

$$\operatorname*{avg}_{\widehat{U}\leftarrow\mathcal{A}^U}\left[F\left(U\left|\psi\right\rangle,\widehat{U}\left|\psi\right\rangle\right)\right]\leq 1-\alpha+2\epsilon+\mathrm{negl}(N)\ , \tag{4.8}$$

where α is a constant independent of N, and

2. *for any $\widehat{U}\leftarrow\mathcal{A}^U$, there exists an operator Q (which may depend on U, $|\psi\rangle$, and $\widehat{U}$) such that*

$$\operatorname*{avg}_{\widehat{U}\leftarrow\mathcal{A}^U}\left|\mathrm{tr}\Big(Q\widehat{U}\psi(\widehat{U})^\dagger\Big)-\mathrm{tr}\big(QU\psi(U)^\dagger\big)\right|\geq 2\alpha-4\epsilon-\mathrm{negl}(N)\ . \tag{4.9}$$

Proof. See App. 4.6.1. □

We now extend this theorem to infinite-dimensional Hilbert spaces. For this purpose, in what follows, we reinstate the sequences of PRSs and PRUs to make the N-dependence explicit.

We now assume that for every N we have a code subspace $\mathcal{H}^{(N)}$ on which the time evolution operator $e^{-iH^{(N)}t}$ of the CFT, for some t (possibly scaling with N), has the same matrix elements on the code subspace as another unitary operator $U^{(N)}:\mathcal{H}^{(N)}\to\mathcal{H}^{(N)}$, up to exponential corrections. Furthermore, we assume there exists a (gravitationally) PRU ensemble $\{\mathcal{U}^{(N)}\}_N$ and some N_0 such that for all $N\geq N_0$, the ensemble $\mathcal{U}^{(N)}$ contains $U^{(N)}$ as a typical member. We also assume that the code subspaces $\mathcal{H}^{(N)}$ can all be embedded into an infinite-dimensional Hilbert space $\mathcal{H}^{(\infty)}$, which describes the physics at infinite N.

Next, we assume there exists a sequence of states $\{\psi^{(N)}\in\mathcal{H}^{(N)}\}_N$ that have a well-defined large-N limit in addition to being pseudorandom. Let us make this precise. We assume there exists a PRS ensemble $\{\Psi^{(N)}\}_N$ such that for all $N\geq N_0$, $\Psi^{(N)}$ contains $\psi^{(N)}$ as a typical member and $\{\psi^{(N)}\}_N$ forms a Cauchy sequence in $\mathcal{H}^{(\infty)}$ with respect to the one-norm on $\mathcal{H}^{(\infty)}$, and converges to $\psi^{(\infty)}\in\mathcal{H}^{(\infty)}$. Additionally, $\{U^{(N)}\psi^{(N)}(U^{(N)})^\dagger\}_N$ is a Cauchy sequence converging to some $\xi^{(\infty)}\in\mathcal{H}^{(\infty)}$. Finally, letting the learning algorithm's best guess for $U^{(N)}$ be $\widehat{U}^{(N)}$, we assume that $\{\widehat{U}^{(N)}\psi^{(N)}(\widehat{U}^{(N)})^\dagger\}_N$ is also a Cauchy sequence that converges to some $\widehat{\xi}^{(\infty)}\in\mathcal{H}^{(\infty)}$. We also assume that the subset of states at finite N that are the fraction of states that are typical, converge to a subset of states in the infinite-N Hilbert space that is of measure 1 (as the fraction of typical states limits to 1 as $N\to\infty$ in Theorem 8).

[Distinguishing Operator at $N\to\infty$] For any two sequences of states $\{U^{(N)}\left|\psi^{(N)}\right\rangle\}_N$ and $\{\widehat{U}^{(N)}\left|\psi^{(N)}\right\rangle\}_N$ that have well-defined limits $\left|\xi^{(\infty)}\right\rangle$ and $\left|\widehat{\xi}^{(\infty)}\right\rangle$, respectively, as $N\to\infty$ and that satisfy the assumptions stated above, both 1. and 2. in Theorem 8 also hold for the states $\left|\xi^{(\infty)}\right\rangle$ and $\left|\widehat{\xi}^{(\infty)}\right\rangle$. In particular, by taking the constant ϵ to be $<\alpha/2$, we have for any $\left|\widehat{\xi}^{(\infty)}\right\rangle$, the existence of an operator $Q^{(\infty)}$ such that,

$$\operatorname*{avg}_{\widehat{\xi}^{(\infty)}}\left|\mathrm{tr}\Big(Q^{(\infty)}\widehat{\xi}^{(\infty)}\Big)-\mathrm{tr}\big(Q^{(\infty)}\xi^{(\infty)}\big)\right|\geq O(1)\ .$$

Proof. See App. 4.6.2.

□

4.3.3 Pseudorandomness implies horizons

We are now ready to argue that when time evolution in the boundary theory is sufficiently (i.e., gravitationally) pseudorandom, the bulk has an event horizon. Recall that the logic in this argument is that for a typical pseudorandom time evolution operator U, (a) efficient quantum learning algorithms cannot accurately compute the time evolution of $\langle Q \rangle$ for some operator Q that may depend on U, and (b) there exists an efficient quantum algorithm that can accurately compute the time evolution of all local operators — i.e., operators in the causal wedge — under U. Point (a) follows directly from Corollary 4.3.2.

Point (b) is in a sense obvious holographically given the developments on complexity of reconstruction in AdS/CFT. The general expectation following from [30] is that a simple quantum algorithm should be able to reconstruct the causal wedge. However, as our presentation thus far has been at a higher level of rigor, we aim to illustrate more precisely that under a mild technical assumption, there exists a quantum learning algorithm satisfying the assumptions of Theorem 7 that, given oracle access to U, can output an operator V where $V|\psi\rangle$ can accurately reproduce the expectation values of all simple operators in the state $U|\psi\rangle$. Applicability to our strengthened version, Theorem 8, does not immediately follow and would require techniques from measure concentration. This falls outside the scope of the current work, but we expect that this may go through if the algorithm in [61] satisfies certain assumptions. Readers willing to accept the existence of the relevant algorithm — i.e., an efficient algorithm for reconstructing the causal wedge — can skip directly to the statement of Cryptographic Censorship, Theorem 10 below.

Efficiently learning the causal wedge. The starting point for our argument for (b) is the learning algorithm of [61], which we review below, together with the assumption referenced above that allows us to bridge the gap between [61]'s work and our Theorem 7.

Given an n-qubit system with n large, the authors of [61] showed that there exists an efficient learning algorithm that can learn a *bounded degree* representation of any unknown operator $\mathcal{O}^{(\text{unk})}$. The bounded degree representation $\mathcal{O}^{(k)}$ takes the form of

$$\mathcal{O}^{(k)} = \sum_{P:|P|\leq k} \alpha_P P \ , \tag{4.10}$$

where P is a Pauli operator given by a string of Pauli operators X, Y, Z, or I at each site, and k is a constant (independent of n). By $|P|$ we mean the number of qubits that are acted on non-trivially by P. A bounded-degree operator is then just a sum of individual Pauli operators with $|P|$ not scaling with n in any way.

The algorithm of [61] works by forming a guess for the coefficients of the unknown operator $\mathcal{O}^{(\text{unk})}$ in the Pauli basis, using a polynomial or subexponential number of measurements involving the operator. The authors show that this guess $\mathcal{O}^{(k)}$ for the operator reproduces the correct expectation values in certain distributions of quantum states. Quantitatively:

Theorem 9 (Theorem 13 of [61])**.** *There exists an algorithm that, for any $n, \varepsilon, \delta > 0$, any unknown n-qubit observable $\mathcal{O}^{(\text{unk})}$ and any n-qubit state distribution $\mathcal{D}$ that is invariant under single qubit gates, given training data $\{\rho_\ell, \text{tr}\left(\mathcal{O}^{(\text{unk})}\rho_\ell\right)\}_{\ell=1}^{M}$ of size*

$$M = \log\left(\frac{n}{\delta}\right) 2^{O\left(\log\left(\frac{1}{\varepsilon}\right)\log(n)\right)} \ , \tag{4.11}$$

the algorithm is of $\mathrm{poly}(M)$ *complexity and can learn a function* $h(\rho)$ *that achieves a prediction error*

$$\operatorname*{avg}_{\rho \leftarrow \mathcal{D}} \left| h(\rho) - \mathrm{tr}\left(\mathcal{O}^{(\mathrm{unk})}\rho\right)\right|^2 \leq \varepsilon \left|\left|\mathcal{O}^{(\mathrm{unk})}\right|\right|^2, \tag{4.12}$$

with probability at least $1-\delta$*. Here,* $\left|\left|\mathcal{O}^{(\mathrm{unk})}\right|\right|$ *is the operator norm of* $\mathcal{O}^{(\mathrm{unk})}$*. In particular, the algorithm produces a bounded-degree operator* $\mathcal{O}^{(k)} = \sum_{P:|P|\leq k} \alpha_P P$ *and thereby learns* $h(\rho) := \mathrm{tr}\big(\mathcal{O}^{(k)}\rho\big)$*.*

One example of such a distribution $\mathcal{D}$ is the distribution of Haar random states, which is of particular relevance to our setting.

In general, a bounded-degree operator is any operator of the form in (4.10). In the context of AdS/CFT, we expect that bounded-degree operators correspond to local operators in the CFT. In particular, expectation values of these operators fully determine the causal wedge of the entire boundary in the large-N limit. Recall that our aim is the illustration of an algorithm that falls under the purview of Theorem 7 and can learn a V that approximates the time-evolved expectation values of bounded-degree operators under U. Taking $\mathcal{O}^{(\mathrm{unk})} = U$, the algorithm in the above theorem (given the specified training data) learns an operator $U^{(k)}$ such that

$$\mathrm{tr}\big(U^{(k)}\rho\big) \approx \mathrm{tr}(U\rho) \tag{4.13}$$

(on average over ρ). Note that this algorithm requires access to training data comprised of expectation values of the unknown operator $\mathcal{O}^{(\mathrm{unk})} = U$ in a collection of states $\{\rho_\ell\}$. We expect that there exists an efficient algorithm to estimate the expectation values of U in this collection of efficiently preparable states given oracle access to U so that the learning algorithm in Theorem 9 then falls under the purview of Theorem 7.

In order to obtain a guarantee that this $U^{(k)}$ is such a V, we will make a mild assumption that (4.13) holds when multiple operators are concatenated. In general, we may expect that for any operators $\mathcal{O}_1$, $\mathcal{O}_2$, $\mathcal{O}_3$, using the learning algorithm in the theorem above to produce $\mathcal{O}_1^{(k)}$, $\mathcal{O}_2^{(k)}$, $\mathcal{O}_3^{(k)}$, the following holds:

$$\mathrm{tr}\left(\mathcal{O}_1^{(k)}\mathcal{O}_2^{(k)}\mathcal{O}_3^{(k)}\rho\right) \approx \mathrm{tr}(\mathcal{O}_1\mathcal{O}_2\mathcal{O}_3\rho)\ .$$

However, to bridge the gap to Theorem 7, it suffices that the $U^{(k)}$ produced by the algorithm satisfies the following for any bounded-degree operator $\mathcal{O}$:

$$\mathrm{tr}\big((U^{(k)})^\dagger \mathcal{O} U^{(k)}\rho\big) \approx \mathrm{tr}\big(U^\dagger \mathcal{O} U\rho\big)\ , \tag{4.14}$$

where the error scales with $1/N$ if the algorithm complexity is subexponential in N, which is a complexity to which Theorem 7 (and its strengthening Theorem 8) can be applied.[27] The upshot is that under the assumption (4.14) above, there is an efficient algorithm that obeys the assumptions of Theorem 7[28] and outputs a guess $V = U^{(k)}$ for the time evolution

[27] Note that although in previous sections we have considered polynomial complexity algorithms, we expect our results there to also apply to algorithms of subexponential complexity.

[28] Note that strictly speaking, Theorem 7 only applies to this algorithm if V is approximately norm-preserving. By taking e.g. $O = I$ in (4.14) (which holds for most ρ in the Haar distribution), we expect that the guess V should satisfy this condition.

unitary U, which correctly predicts expectation values of local operators in the CFT, up to polynomial corrections in $1/N$.

To summarize: we use the above (and expectations from holography) to motivate that there exists an algorithm, which we will call $\mathcal{A}_{CW}$, that can always learn time evolution within the causal wedge but falls under the purview of Theorem 7, which guarantees that for pseudorandom time evolution, this algorithm cannot learn the time evolution of *all* operators. This forms the basis of our theorem of Cryptographic Censorship. Under the assumptions above, which simply boil down to the existence of $\mathcal{A}_{CW}$, together with the assumptions of Section 1.1, we have the following theorem:

Theorem 10 (Cryptographic Censorship)**.** *Let U_{fund} be the time evolution operator in the large-N limit, and assume that U_{fund} is well approximated on a code subspace $\mathcal{H}_{\text{code}}$ by a unitary U which is typical in a gravitationally pseudorandom unitary ensemble, $\mathcal{U}$. Assume furthermore (as motivated above) that there exists an efficient algorithm for causal wedge reconstruction that is (randomized) K-Lipschitz for some $K > 0$, and which additionally satisfies the criteria assumed in Theorem 8. Let $|\psi(t)\rangle$ be a state at any time t whose complete time evolution is dual to a strongly causal geometric[29] bulk (M, g). Under the convergence assumptions above, if there exists a single time t such that $U|\psi(t)\rangle$ is typical in the pseudorandom state ensemble $\Psi = \mathcal{U}|\psi(t)\rangle$,[30] then there is an event horizon in (M, g).*

Note that when we say that U_{fund} is well approximated by its pseudorandom doppelganger U, we mean that the fidelity between U_{fund} and U acting on a state in $\mathcal{H}_{\text{code}}$ is at worst case exponentially close to 1.

Proof. Let U be the pseudorandom unitary approximating $e^{-iH\Delta t}$ over a time Δt. Let the fundamental quantum state of the system be $|\psi(t)\rangle$ at some initial time t, and $|\psi(t')\rangle = U|\psi(t)\rangle$ at some time, $t' = t + \Delta t$. See Fig. 4.8. By assumption, we can find a t such that $|\psi(t')\rangle$ is typical in the PRS ensemble $\Psi = \mathcal{U}|\psi(t)\rangle$.

Consider a particular (possibly randomized) algorithm, $\mathcal{A}^U_{CW}$, which by the reasoning and assumptions above based on Theorem 9, we can choose to correctly predict the values of all operators $\mathcal{O}$ that are in the causal wedge for any output $\widehat{U} \leftarrow \mathcal{A}^U_{CW}$. By Corollary 4.3.2, on average over the $\widehat{U}$ there exists a distinguishing operator. Thus there must exist at least one $\widehat{U}$ for which the corresponding Q_* distinguishes between $\widehat{U}|\psi(t')\rangle$ and $U|\psi(t')\rangle$. The operator Q_* — whose expectation value differs by $O(1)$ between these states — exists in the $N \to \infty$ algebra and so must have have a well-defined bulk dual according to the work of [62, 63]. By construction, $\mathcal{A}^U_{CW}$ can correctly predict all causal wedge operators and so the operator Q_* must have support in a bulk spacetime region that is outside the causal wedge. Thus a horizon exists. □

We end this section by pointing out that we in fact get a more general result: the proof of Theorem 10 goes through for any time evolution operator U that cannot be predicted (in the sense of Theorem 8) by efficient algorithms. In particular, the diagnostic does not have to be pseudorandomness: our theorem holds for any U that is exponentially complex to learn.

[29] By geometric we mean that fluctuations of the geometry are suppressed parametrically in N in the large-N limit.

[30] Recall that the typical states constitute a $\lim_{N\to\infty} 1 - e^{-\epsilon^2 2^\kappa/(48(1+K)^2)} - e^{-\epsilon^2 2^\kappa/192} - \text{negl}'(N) = 1$ (recall $\kappa = \text{poly}(N)$) fraction of states in Ψ, and similarly for typical $U \in \mathcal{U}$.

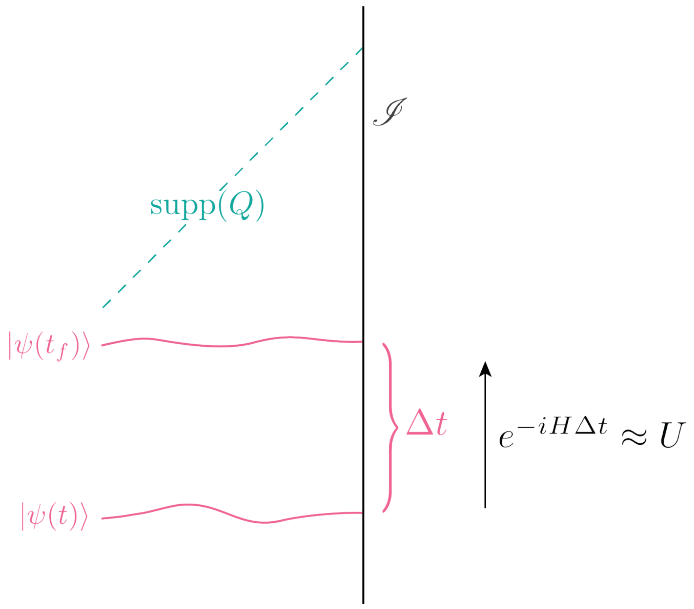


Figure 4.8: By assumption, the time evolution operator $U_{\text{fund}} = e^{-iH\Delta t}$ on $\mathcal{H}_{\text{code}}$ over a time Δt is well approximated by some $U \leftarrow \mathcal{U}$, such that $|\psi(t')\rangle$ is typical in the PRS ensemble $\Psi = \mathcal{U}|\psi(t)\rangle$. In this case, Theorem 10 guarantees the existence of a distinguishing operator Q that has support outside the causal wedge: there must be a horizon (indicated by the green dotted line).

4.3.4 Possible subspaces with pseudorandom time evolution

We now discuss possible code subspaces which accommodate gravitationally pseudorandom dynamics. We will outline two cases of interest. First, we consider a microcanonical window about some energy above the black hole threshold. Second, we will argue that there can exist a code subspace where all the states have the same classical geometry, with low energy modes on top of it, and for which it is possible to obtain gravitational pseudorandomness within an $O(1)$ time on the boundary.

The microcanonical window

An obvious candidate for a subspace in a holographic quantum system which admits pseudorandom time evolution is a microcanonical window about some energy $E_0 \sim N^2$ with energy width $\Delta E \sim O(N)$. The microcanonical ensemble defined by this window is known to have a stationary two-sided black hole dual [64]. Here we are interested in applying our result to any state in this energy window with a semiclassical geometric bulk description.

For holographic theories that exhibit maximal chaos, it is an interesting question how large T needs to be for $U(T) = e^{-iHT}$ to be well approximated by a pseudorandom unitary ensemble. A related (but different) question is how large T needs to be in a maximally chaotic system for the unitary to be well approximated by a unitary k-design. For $k = 2$, this question was answered in [65] by directly relating out-of-time-order (OTO) correlators to the question of how close a given ensemble is to a unitary 2-design. The authors found that if the (square of the) 2-point OTO correlator is small when averaged over all possible

operators, then the time evolution is well modeled by a 2-design.[31]

As discussed above, k-designs are a different but related notion from pseudorandom unitary ensembles. That said, this intuition suggests that the time T at which $U(T)$ becomes well modeled by a pseudorandom unitary is at least the scrambling time for chaotic quantum systems, since this is the time when the OTO is expected to match those of a 2-design. Note that if $U(T)$ is ever going to be modeled by a pseudorandom unitary, it better be at times T sub-exponential in the entropy, otherwise U will not be efficiently generated. Thus, the expectation is that $U(T)$ should become pseudorandom on the microcanonical subspace for sufficiently large T larger than the scrambling time. It would of course be nice to know what exactly this timescale is in certain systems.

Applying Theorem 10 to this subspace, we then see that for any state in the microcanonical window, $|\psi_0\rangle$, that is both dual to a geometric state and is also typical in the pseudorandom state ensemble, then *with a probability that is negligibly (in N) close to one there must exist a horizon in the bulk dual to $|\psi_0\rangle$.* It is natural to ask what *fraction* of the states in the microcanonical window this corresponds to. We cannot give a definitive answer to this question in the case of pseudorandom ensembles; let us just note that for Haar random unitaries and states, this would include all but a doubly-exponentially small fraction of the states. However, since it is not clear whether typical Haar random states can have semiclassical geometric duals,[32] we prefer to use pseudorandom ensembles. It is not implausible that a similar fraction of the states in some pseudorandom ensemble would obey this property, but we shall remain agnostic on this point in the absence of a proof of this statement.

A code subspace for $O(1)$ time bands on the boundary

We will now consider a different subspace, where we restrict to $O(1)$ time bands on the boundary in order to preserve the geometry of states in our code subspace within exponentially suppressed errors. We will start by making some assumptions that will apply to what follows, as well as in Sec. 4.4. These assumptions will describe how we construct a code subspace to describe time evolution of probe fields for $O(1)$ times around geometries with naked singularities.

Consider a CFT labeled by a parameter N, such that the theory in the $N \to \infty$ limit is dual to Einstein gravity coupled to matter.[33] Let $\hat{\mathcal{H}}^{(N)}$ be its full Hilbert space. We are now going to consider a finite-dimensional subspace $\mathcal{H}^{(N)}$ of $\hat{\mathcal{H}}^{(N)}$ that we will assume to have certain properties. This will formalize assumptions often implicitly made in the study of AdS/CFT, but since we will (1) study somewhat exotic setups, and (2) apply certain rigorous theorems, we are careful to state our assumptions.

Take $\left|\psi_i^{(N)}\right\rangle$ to be a basis of $\mathcal{H}^{(N)}$ labeled by i. We will assume that there exists a

[31] The authors proved this statement for an n-qubit system and the average is over all Pauli operators on this n-qubit Hilbert space, which provide a basis for all operators.

[32] Note that if the fraction of geometric states which violate our theorem is much larger than doubly exponentially small, then there would need to be a conspiracy between having a geometry and the ability to reconstruct all operators. Since we expect spacetimes with horizons to be much more "numerous" than those without, this would be rather surprising.

[33] We always assume that any effective 't Hooft coupling in the theory has been taken to infinity.

one-parameter family of unitary operators

$$U^{(N)}(t) : \mathcal{H}^{(N)} \to \mathcal{H}^{(N)} \tag{4.15}$$

such that for all $t \sim O(N^0)$,

$$|\psi_i(t)\rangle \equiv \lim_{N\to\infty} U^{(N)}(t) \left|\psi_i^{(N)}\right\rangle \tag{4.16}$$

exists in an infinite-dimensional Hilbert space $\mathcal{H}$ describing QFT in curved spacetime with perturbative backreaction. Furthermore, for each i, we assume that the (one-parameter family of) states $|\psi_i(t)\rangle$ is dual to a spacetime that is the same for each i, but with different perturbative quantum fields on top of this background. Next, we will assume that if $e^{-iH^{(N)}t}$ is the true time evolution operator in the full CFT at finite N, then for any fixed $t \sim O(N^0)$ and any state $|\varphi\rangle \in \hat{\mathcal{H}}^{(N)}$,

$$\langle\varphi| e^{-iH^{(N)}t} \left|\psi_i^{(N)}\right\rangle = \langle\varphi| U^{(N)}(t) \left|\psi_i^{(N)}\right\rangle + O(e^{-cG^a}) \tag{4.17}$$

for some $c > 0$, $a < 0$. For $|\varphi\rangle$ in the orthogonal complement of $\mathcal{H}^{(N)}$, the first term is identically zero, so our statement implies that at large N, CFT time evolution for an $O(1)$ amount time preserves our code subspace $\mathcal{H}^{(N)}$, up to exponentially small corrections. This is just the usual intuition that different geometries have exponentially small overlaps, so unless we work to exponential precision, we can consider time evolution to preserve our code subspace. Instead of working with e^{-iHt}, we replace this operator by a unitary that genuinely acts on $\mathcal{H}^{(N)}$.

Then, under the above assumptions, Theorem 10 implies the following. Consider an asymptotically AdS spacetime (M, g) where the boundary dynamics in an $O(1)$ timeband evolve according to a unitary $U_{\text{fund}} = e^{-iHt}$ and as in the above, can be approximated by $U^{(N)}(t)$ which is a typical gravitationally pseudorandom unitary (as defined in Def. 6). Let the state dual to (M, g) be typical in a pseudorandom state ensemble. Then there must be an event horizon in (M, g).

This is the version of Cryptographic Censorship that we will use in the next section.

4.4 From Cryptographic to Cosmic Censorship

In Section 4.3.4, we adapted our main result — Cryptographic Censorship — to code subspaces of states that have the same classical geometry and undergo gravitationally pseudorandom boundary dynamics in an $O(1)$ timeband. In this section we will argue that a potential example of such a code subspace is a spacetime with a 'large' naked singularity. Our goal is to connect Cryptographic Censorship to the more familiar notion of cosmic censorship, which is roughly the expectation that naked singularities formed from collapse must be hidden behind horizons.

Cryptographic Censorship guarantees that singularities that are well-described by gravitationally pseudorandom dynamics must incur event horizons. The central task in connecting Cryptographic Censorship to cosmic censorship is thus understanding whether — or when —

singularities are associated with (sufficiently) pseudorandom dynamics. If generic singularities formed from matter collapse do indeed require pseudorandom time evolution, then they cannot exist in horizonless spacetimes. This version of censorship of naked singularities is distinct from the original version of cosmic censorship: even when a singularity is associated with pseudorandom dynamics, our result does not imply that the singularity is completely hidden. This is not a bug but a feature, since certain naked singularities do exist: e.g., the Gregory–Laflamme (GL) instability, which occurs when a black string in $D > 4$ pinches off and fragments into multiple black holes [7, 8, 66].

In the following, we take some first steps towards connecting a version of cosmic censorship to Cryptographic Censorship, culminating in Sec. 4.4.3, where we state a cryptographically-inspired formulation of quantum cosmic censorship. We begin by defining a mathematically precise notion of the size of a singularity, and specifically identify how large the singularity must be in the $G_N \to 0$ limit to accommodate gravitationally pseudorandom fundamental dynamics. This definition usefully distinguishes between a Planckian and a classical naked singularity (and similarly for non-naked singularities of large and small sizes). This is in itself useful, as it separates naked singularities into classes that are expected to be excluded and those that we expect may be permitted in a quantum theory of gravity (e.g., evaporating black holes or Gregory–Laflamme instabilities).[34] Our second step uses this quantification to provide some evidence that large classical singularities can indeed be associated with gravitationally pseudorandom dynamics and thus incur horizons by Cryptographic Censorship. Finally, we propose a Quantum Weak Cosmic Censorship Conjecture, which formally requires large naked singularities to be atypical.

4.4.1 The size of singularities

We are interested in an unambiguous and diffeomorphism-invariant way of measuring the extent of a naked singularity. Since a naked singularity can only be distinguished from a cloaked one via access to $\mathscr{I}$, it is clear that this must be defined asymptotically. A definition on $\mathscr{I}$ would be automatically diffeomorphism-invariant and would more readily admit a straightforward dual CFT interpretation. To find this criterion, we will leverage the definition of a naked singularity — i.e., causal visibility to $\mathscr{I}$ — to define what will roughly correspond to the time evolution in the CFT that is necessary, in the Schrödinger picture, to reconstruct near-singularity physics.

To make this precise, we first use an extant mathematical construction to single out the singularity as a set, a nontrivial task given that the singularity is not part of the spacetime manifold. Second, we identify the time intervals on $\mathscr{I}$ that are causally-separated from this set, and finally find the minimal boundary time bands that are necessary for a full reconstruction of the singularity. The spacetime volume of these time bands in a fixed conformal frame is essentially our definition of the size of the singularity in that frame. There is thus a conformal frame ambiguity in our definition of singularity size. This, however, is immaterial for our desired application: we shall only care about scaling of the singularity size with G_N, which is independent of the choice of conformal frame. Using this scaling, we

[34] There is some extant literature dedicated to investigating the size of a singularity; see, e.g., [25]. These approaches differ from ours, as we are motivated by quantum computational considerations in the fundamental quantum description of the system.

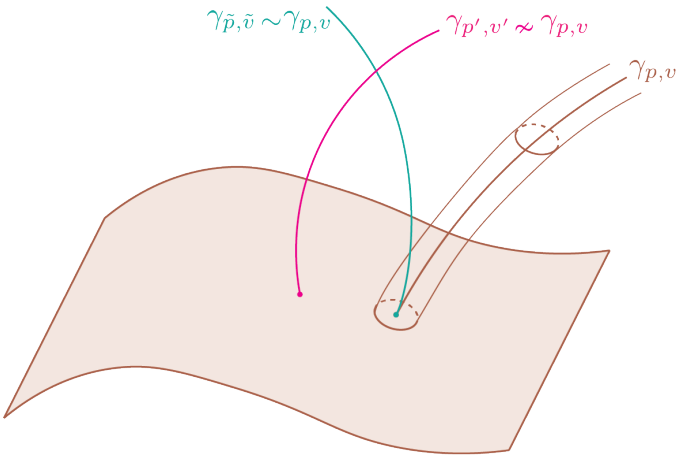


Figure 4.9: Example spacetime where the colored sheet represents a singularity. The two geodesics $\gamma_{p,v}$ and $\gamma_{\tilde{p},\tilde{v}}$ by definition end on the same singular point, since $\gamma_{\tilde{p},\tilde{v}}$ must enter any tubular neighborhood of $\gamma_{p,v}$, and vice versa.

separate the singularities into three categories: classical, semi-Planckian, and Planckian. The first type survives the $G_N \to 0$ limit, while the other two do not. These latter two sizes are distinguished via the power of G_N by which they vanish. Let us now turn to the details, starting with an identification of the set comprising the singularity.

The Singularity Set. The primary difficulty in this task is that singularities are not locations in some geodesically-incomplete manifold M but are rather associated to the endpoints of incomplete inextendible geodesics in M. However, as shown by [67], it is possible to treat singularities as if they are made up of a set of points Γ, although this set does not lie in M.[35] We will briefly review this construction, as it is instrumental to our classification of singularities. The idea is to build an equivalence class of geodesics associated to an endpoint. Consider a geodesic $\gamma_{p,v}$ in M fired from a starting point $p \in M$ along a non-vanishing tangent vector v^a at p. We will continue $\gamma_{p,v}$ so it is inextendible in the direction of v^a (with starting point p). We then identify two distinct geodesics $\gamma_{p,v} \sim \gamma_{\tilde{p},\tilde{v}}$ if $\gamma_{p,v}$ enters and remains inside every tubular neighborhood of $\gamma_{\tilde{p},\tilde{v}}$. Explicitly, if $\lambda \in [0, \lambda_0)$ is an affine parameter for $\gamma_{p,v}$, then for every given tubular neighborhood U of $\gamma_{\tilde{p},\tilde{v}}$, there exist a $0 \leq \hat{\lambda} < \lambda_0$ such that $\gamma_{p,v}(\lambda) \in U$ for all $\lambda \in (\hat{\lambda}, \lambda_0)$. See Fig. 4.9. The equivalence classes under $\sim$ then form the points in Γ, and we say that $\gamma_{p,v}, \gamma_{\tilde{p},\tilde{v}}$ are geodesics that end on the singular point $[\gamma_{p,v}] \in \Gamma$. Importantly, if $p_s \in \Gamma$, the future $I^+(p_s)$ and past $I^-(p_s)$ of p_s in M is well-defined even though p_s is not in M. Namely, we say that $p \in I^+(p_s)$ if there is a future-directed timelike curve γ through p and with past "endpoint" on p_s, where the latter means that γ, sufficiently far in its past direction, enters and remains within any tubular neighborhood of any representative geodesic in p_s. Armed with a set identified as the singularity, we now extract the part of the singularity that is naked with respect to a given conformal boundary $\mathscr{I}$ as follows:

$$\Gamma_{n(\mathscr{I})} = \{p_s \in \Gamma | I^+(p_s) \cap \mathscr{I} \neq \varnothing \text{ and } I^-(p_s) \cap \mathscr{I} \neq \varnothing\}. \tag{4.18}$$

[35] Γ also comes with a natural choice of topology [67].

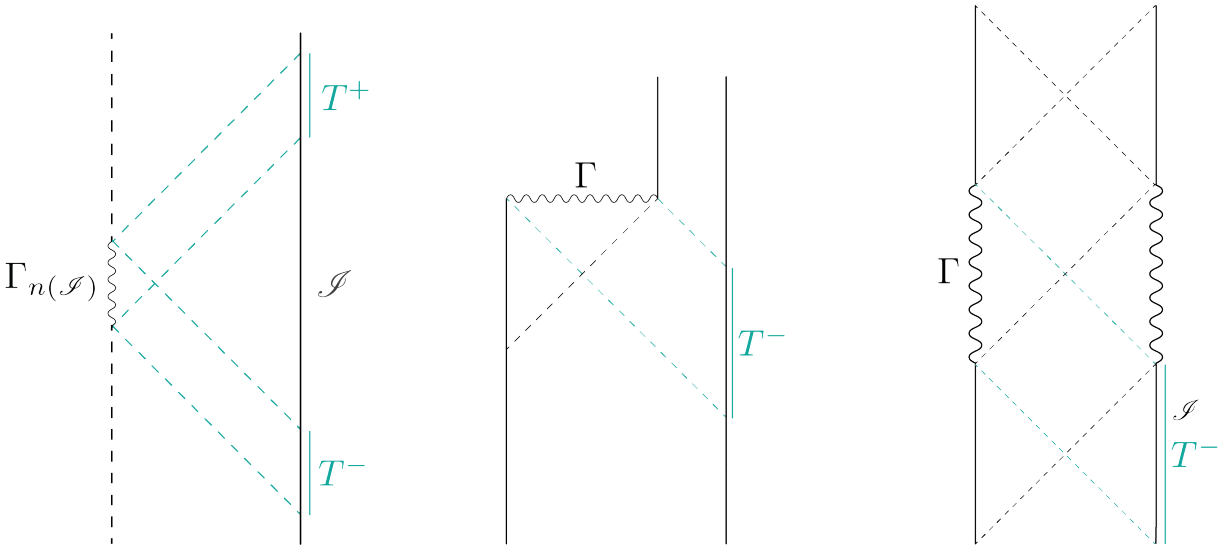


Figure 4.10: Some examples of spacetimes with (naked) singularities. In (a) the singularity is timelike and classically naked. Panel (b) depicts the evaporating black hole. The non-naked part of the singularity is classical and behind a horizon. For the naked part, see Fig. 4.12a. Panel (c) is Reissner–Nordström-AdS; here $|\Gamma|_{\mathscr{I}} = |T_-|$; note, however, that $\Gamma_{n(\mathscr{I})} = \varnothing$, as $I^{\pm} \cap \mathscr{I} = \varnothing$.

If this is nonempty, we say that Γ is naked with respect to $\mathscr{I}$.

The Singularity Size. We now proceed to classify the size of any (possibly improper) subset of a singularity. Let $\hat{\Gamma}$ be a subset of Γ; below we will mostly be interested in taking $\hat{\Gamma} = \Gamma_{n(\mathscr{I})}$, but it will be useful to take our definition to be more general. Consider the sets

$$\begin{aligned} T_+ &= \{p \in \mathscr{I} |\ \exists \text{ an achronal past-directed null geodesic from } p \text{ to } \hat{\Gamma}\} \ , \\ T_- &= \{p \in \mathscr{I} |\ \exists \text{ an achronal future-directed null geodesic from } p \text{ to } \hat{\Gamma}\} \ . \end{aligned} \tag{4.19}$$

Fig. 4.10 illustrates some examples of these sets for both naked singularities and ordinary singularities behind horizons.

We will now define the size of $\hat{\Gamma}$ according to properties of T_+, T_-. We assume that (M, g) is conformal to the Einstein static universe on each conformal boundary, and we choose the unique conformal frame where the boundary metric is identical to the Einstein static universe in which the spatial spheres have radius equal to the AdS radius.[36] This fixes a unique Lorentzian metric on $\mathscr{I}$, and when T_+ is nonempty, we define $|T_+|$ as the absolute value of the spacetime volume of T_+ in this metric; similarly for T_-. As examples, if T_+ is a timeslice or a union of a finite number of points, then $|T_+| = 0$, since these sets are measure

[36] For asymptotically locally AdS spacetimes where $\mathscr{I}$ has another conformal class, we can also define $|\hat{\Gamma}|$, once we pick some preferred conformal frame.

zero. We can now define the size $|\hat{\Gamma}|_{\mathscr{I}}$ of $\hat{\Gamma}$ with respect to $\mathscr{I}$ as follows: first, if T_+, T_- are both empty, then $|\hat{\Gamma}|_{\mathscr{I}} = 0$. Second, if only one of T_+, T_- is nonempty, say T_+, then we define $|\hat{\Gamma}|_{\mathscr{I}} = |T_+|$. Third, if both $T_\pm$ are nonempty, we define $|\hat{\Gamma}|_{\mathscr{I}} = \min(|T_+|, |T_-|)$. This finally leads us to our definition of classical and Planckian singularities, as well as an in-between regime, which we term semi-Planckian:

Definition 7. *Let (M, g) be a geodesically incomplete spacetime of dimension $D = d+1$ with a singularity $\hat{\Gamma}$, and let*

$$|\hat{\Gamma}|_{\mathscr{I}} \sim O(\ell_{\text{Pl}}^a). \tag{4.20}$$

If $a \leq 0$, we say that $\hat{\Gamma}$ is a classical singularity, while if $0 < a < d-1$, we say that $\hat{\Gamma}$ is a semi-Planckian singularity. Otherwise, it is a Planckian singularity.

Singularity Size from CFT Time Evolution. If $|\psi(t)\rangle$ is the state whose complete past and future history is dual to (M, g), then without access to the time evolution operator, we can only reconstruct operators at points in spacetime that are not causally related to the boundary time slice on which $|\psi(t)\rangle$ lives — i.e., the Wheeler–de Witt patch W_t.[37] Let us now specialize to a spacetime with a naked singularity Γ. For some fixed t, generally only a (possibly empty) subset of Γ will be contained in W_t: to learn everything about Γ (for example by reconstruction of all near-singularity correlators), a certain amount of boundary time evolution is necessary. More precisely, if T_- is nonempty, and we are evolving forwards from the past, then the past-most boundary of T_- corresponds to the earliest time that some part of Γ is encoded in the Wheeler–de Witt patch W_t. As we evolve forward through T_-, new pieces of Γ enter W_t and are encoded in the time-evolved state. Finally, once we evolve past the futuremost boundary of T_-, no new parts of Γ enter W_t that we have not already seen, and so further forward time evolution reveals nothing new about the singularity. Thus, everything about the naked singularity can be learned by carrying out CFT time evolution through the band T_- (plus bulk reconstruction not using the time-evolution operator). For a singularity where T_+ is also nonempty, everything about the singularity could alternatively have been learned by evolving through T_+. So as advertised, $|\Gamma|_{\mathscr{I}} = \min\{|T_-|, |T_+|\}$ measures the minimal amount of CFT time evolution required to learn everything about the naked singularity. Note that $|\Gamma|_{\mathscr{I}}$ has units of spacetime volume rather than time. This is a feature: if all information about the singularity is available upon evolution of the state forwards in a small spatial subregion of a boundary Cauchy slice, then this should be accounted for in the smallness of a singularity. See Fig. 4.11. Note also that if $|\Gamma|_{\mathscr{I}}$ is large, this always is caused by a large amount of time evolution, since timeslices have bounded volume.

Let us now make a few remarks on our definition, starting with a question. Under what circumstances do Planckian or semi-Planckian singularities exist? The most natural case is that we have $\lim_{\ell_{\text{Pl}} \to 0} |\hat{\Gamma}|_{\mathscr{I}} = 0$ because the CFT state acquires all information about the singularity in a CFT timestep that shrinks to zero as $\ell_{\text{Pl}} \to 0$. An example would be the singularity in the evaporating black hole, illustrated in Fig. 4.12a, or a naked timelike singularity $\hat{\Gamma}$ lying behind another naked timelike singularity $\tilde{\Gamma}$, like in Fig. 4.12b. In this

[37] This assumes for simplicity that there is only one asymptotic boundary, and that the state of the CFT on that boundary is pure — so the entanglement wedge of $|\psi\rangle$ is the entire bulk. If there are multiple boundaries or the state on one boundary is mixed for a different reason, we simply take the intersection of W_t with the entanglement wedge of $|\psi\rangle$.

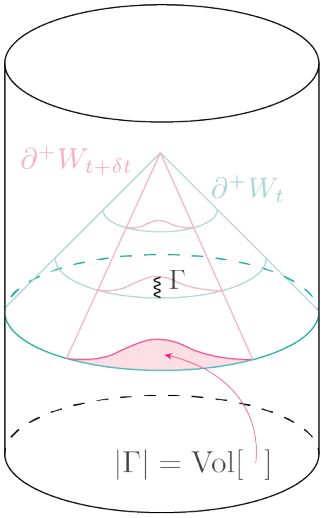


Figure 4.11: To get all information about the small singularity Γ, we only need to evolve the state forwards in a small spatial subregion of the boundary Cauchy slice, such that the Wheeler–de Witt patch now includes the singularity. Note that, as our artistic capabilities are limited, we here only indicate the future boundaries of the Wheeler–de Witt patches for times t and $t + \delta t$.

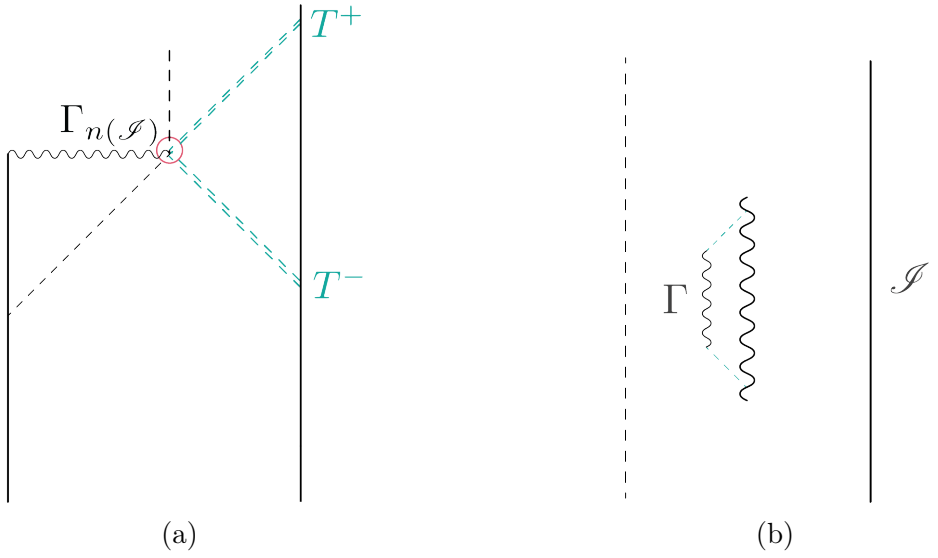


Figure 4.12: Two examples of $\lim_{\ell_{\mathrm{Pl}} \to 0} |\hat{\Gamma}|_{\mathscr{I}} = 0$: (a) For the naked part of the singularity at the endpoint of black hole evaporation, we expect that $T_\pm$ are just timeslices in the $G_N \to 0$ limit. (b) A 'hidden' naked timelike singularity; here, $T_\pm = \varnothing$.

latter case, $\hat{\Gamma}$ is such that $T_+ = T_- = \varnothing$, since all causal geodesics from the boundary to $\hat{\Gamma}$ are necessarily chronal. This latter example shows that we probably should not assign much meaning to the size of distinct connected components, and that we instead just should look at the size of the full naked singularity $\Gamma_{n(\mathscr{I})}$, which by definition contains all connected components.

4.4.2 Accommodating gravitational pseudorandomness

Our motivation for classifying a classical singularity via the scaling $a \leq 0$ is simple: such singularities are visible for a boundary time that diverges or remains $O(1)$ in the $G_N \to 0$ limit. For the remaining range of a, we have chosen to separate semi-Planckian singularities (with $0 < a < d-1$) from Planckian singularities (with $a \geq d-1$). In both cases, $|\hat{\Gamma}|_{\mathscr{I}} \to 0$ as $G_N \to 0$, but as we will show below, singularities with $a < d-1$ are compatible with an EFT code subspace which is sufficiently large that it can accommodate gravitational pseudorandomness, while singularities with $a \geq d-1$ are not.

We will now show that singularities with $a < d-1$ are compatible with gravitational pseudorandomness. Consider an asymptotically AdS spacetime with a timelike naked singularity $\Gamma = \Gamma_{n(\mathscr{I})}$. We will work with large (but possibly finite) N and demand only that effective field theory as defined by a bulk UV cutoff be valid in the asymptotic region. At finite N we will consider the failure of EFT due to large UV corrections as the diagnostic of a singularity.[38]

We will decompose our bulk state into modes with energy below the cutoff, and we will focus on modes localized near the boundary, some of which fall into the naked singularity. By unitarity, there must also be outgoing occupied modes as well, which can be traced back to the singularity. However, the relation between the modes falling into the singularity and the modes coming out is dependent on UV physics: it is inaccessible within low energy EFT. So while the low energy modes in the asymptotic region could be semiclassical as $G_N \to 0$, the time evolution operator that relates them is not known in the bulk EFT — it cannot be expressed in terms of the low-energy operators. See Fig. 4.13. From the low-energy perspective, there is a boundary condition assigned at the singularity whose precise nature is unknown.[39] Now, consider a set of ingoing wavepackets $|i\rangle_{\text{in}}$ that eventually fall into the singularity. We might expect that these modes scatter off the singularity in an apparently random manner. One toy model of such dynamics turns ingoing modes into outgoing modes of seemingly random phases θ_{ij}, schematically,

$$|i\rangle_{\text{in}} \xrightarrow{\delta t} \frac{1}{\sqrt{\mathcal{N}}} \sum_j e^{i\theta_{ij}} |j\rangle_{\text{out}} \,, \tag{4.21}$$

where $\mathcal{N}$ is a normalization constant. The phases θ_{ij} could in principle correspond to pseudorandom phases, which would be a consequence of pseudorandom dynamics in the

[38] Here we are taking EFT to be resulting solely from a UV cutoff rather than including the more complete version of an EFT proposed by [31], which also includes a complexity restriction.

[39] Since quantum gravity is expected to lack global symmetries [68–70], any approximate symmetries emergent in the semiclassical EFT should be badly violated by this boundary condition.

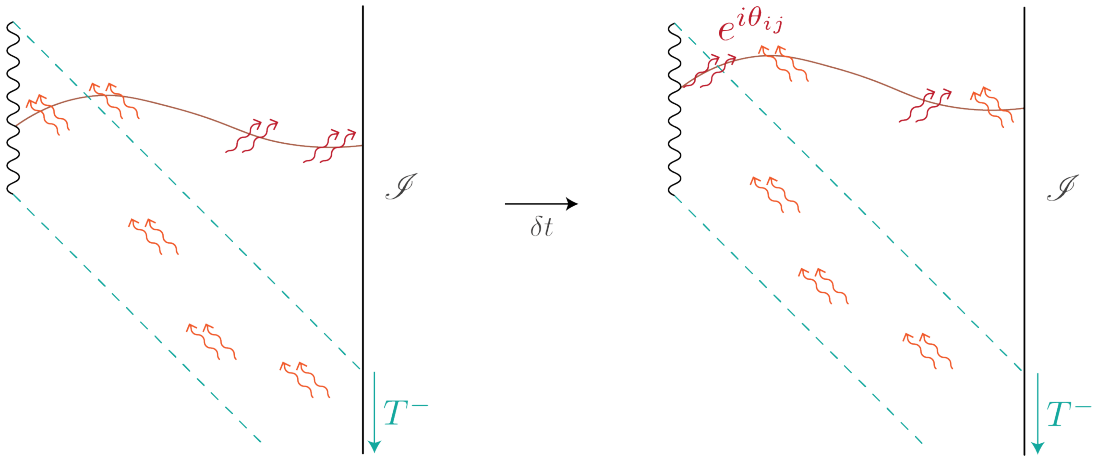


Figure 4.13: From the low-energy perspective, there is an unknown, UV-sensitive boundary condition assigned at the singularity; for example, modes that fall into the singularity are turned into a set of outgoing modes with random phases θ_{ij}.

boundary description[40] (though of course determining what θ_{ij} actually is requires access to UV physics). See, e.g., [33] for explicit constructions of pseudorandom states using pseudorandom phases. Our approach instead will be to show that once the singularity is of classical or semi-Planckian size, then there are enough EFT modes in a code subspace in which all of the states have the same leading order geometry to accommodate gravitational pseudorandomness. On the other hand, if the singularity is Planckian, we show that we do not expect this. As our analysis is asymptotic, we do not require a sharply defined geometry near the singularity.

We now give a parametric estimate of the number of modes within semiclassical EFT that can potentially get scrambled by the naked timelike singularity. That is, we want to know the number of EFT modes that are affected by the highly UV-sensitive boundary condition. Assume that T_+, T_- are nonempty, and assume without loss of generality $|\Gamma_{n(\mathscr{I})}|_{\mathscr{I}} = |T_-|$. Near $\mathscr{I}$, the leading part of the metric can be written

$$ds^2 = \frac{-dt^2 + d\rho^2 + \sin^2\rho d\Omega^2}{\cos^2\rho} + \dots \,, \tag{4.22}$$

where we work in units where the AdS radius is unity, and where $\mathscr{I}$ lies at $\rho = \frac{\pi}{2}$. For pedagogical simplicity we take the matter in our theory to be a conformal massless scalar and assume that the system is spherically symmetric, so that T_- can be taken to be a timeband $[-\Delta t, 0] \times S^{d-1}$ without loss of generality. We can now count roughly how many ingoing single-particle modes with energy below the cutoff can be sent from T_- into the singularity. As is clear from Fig. 4.14, this simply corresponds to the number of orthogonal modes in the strip given by $\rho \in \left[\frac{\pi}{2} - \Delta t, \frac{\pi}{2}\right]$ at $t = 0$.[41] As long as Δt is less than roughly $G_N M \sim O(1)$

[40] There are constraints from overall conservation of ADM energy and gauge charge. But there will be many collections of wavepackets that have the same energy, so that (4.21) would be consistent with charge conservation.

[41] Greybody factors cause some modes to scatter off background curvature before they get close to the singularity, but this should not change the parametric counting.

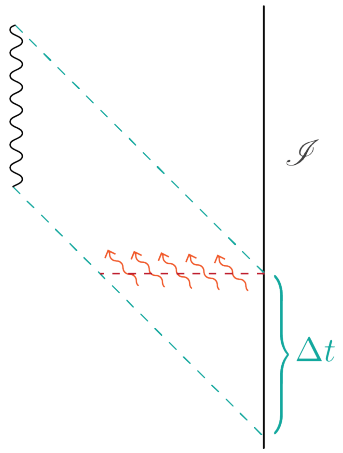


Figure 4.14: To count the ingoing modes sent from T_-, which here is the timeband $[-\Delta t, 0] \times S^{d-1}$, we can simple count the number of modes in the strip indicated by the red dotted line.

where M is the spacetime mass, we are always restricted to an asymptotic range of ρ where the metric is approximately pure AdS, so the number of modes can be counted using the pure AdS metric. We always take G_N small with $G_N M =$ fixed, and we are interested in a parametric counting that can only change if $\Delta t \to 0$ sufficiently fast in G_N so we can without loss of generality restrict to this regime. Assume now that $\tilde{u}_i$ is a collection of orthonormal flat-space positive frequency ingoing wavepackets with mean frequency ω_i in the metric $-dt^2 + d\rho^2 + \sin^2\rho d\Omega^2$; furthermore, we will assume that they are well localized within $\rho \in (\frac{\pi}{2} - \Delta t, \frac{\pi}{2})$ at some central $\rho = \rho_0$. These have total energy approximately equal to ω_i, as long as

$$\omega_i \gg \Delta\omega \geq \frac{1}{\Delta t} \ , \tag{4.23}$$

where $\Delta\omega$ is the frequency spread of each wavepacket. Now, since we have a conformally coupled scalar, $u_i = (\cos\rho)^{d-1}\tilde{u}_i$ is also an orthonormal set of positive frequency mode functions in AdS.[42] Each wavepacket has a total energy approximately equal to ω_i, also in AdS.[43] Imposing the cutoff $\omega_i \leq \epsilon^{-1}$,[44] we want to count the orthogonal wavepackets. This is equivalent to counting orthogonal flat-space modes; this is of the order

$$\frac{\Delta t \cdot \mathrm{Vol}(S^{d-1})}{\epsilon^d} \ . \tag{4.24}$$

The numerator is $|\Gamma_{n(\mathscr{I})}|_{\mathscr{I}}$, and in the absence of spherical symmetry it is clear that $|\Gamma_{n(\mathscr{I})}|_{\mathscr{I}}$ is the correct replacement for $\Delta t \cdot \mathrm{Vol}(S^{d-1})$. So for a general naked timelike singularity, we find that an order

$$\frac{|\Gamma_{n(\mathscr{I})}|_{\mathscr{I}}}{\epsilon^d} \tag{4.25}$$

[42]The scaling dimension of the scalar is $\frac{d-1}{2}$, and the conformal factor is $\cos^2\rho$.

[43]There might be an overall conformal anomaly, so that $H_{\mathrm{AdS}} = H_{\mathrm{Mink}} + H_{\mathrm{anomaly}}$, but this is irrelevant for our purposes.

[44]This corresponds to a proper length cutoff in AdS of $\epsilon_{\mathrm{proper}} \sim \epsilon/\cos\rho_0$. We do not scale ρ_0 with ℓ_{Pl}, so $\epsilon, \epsilon_{\mathrm{proper}}$ scale the same way.

number of distinct EFT modes can propagate into the singularity from T_-. Next, we want to take the UV cutoff to scale as $\epsilon = \ell_{\rm Pl}^{1-\delta}$ for some $\delta \in (0,1)$ so that our modes are parametrically below the Planck scale as $\ell_{\rm Pl} \to 0$. To be cautious, we demand that $\delta > 1/d$, which ensures that the naïve QFT entropy does not exceed the Bekenstein–Hawking entropy. To see this, note that if V is the spatial volume occupied by our wavepackets in AdS, then we need

$$\frac{V}{\epsilon_{\rm proper}^d} = \cos^d \rho_0 \frac{V}{\epsilon^d} \lesssim \frac{A}{\ell_{\rm Pl}^{d-1}} \ . \tag{4.26}$$

Remembering that volume and surface area of a large shell near the boundary of AdS scale the same way as we approach the conformal boundary, we thus need

$$\cos^d \rho_0 < O(1) \frac{\epsilon^d}{\ell_{\rm Pl}^{d-1}} = O(1) \ell_{\rm Pl}^{1-d\delta}. \tag{4.27}$$

If we take $\delta < \frac{1}{d}$, then wavepackets with mean frequency near the cutoff would give a naïve QFT entropy exceeding the Bekenstein–Hawking entropy for any $\rho_0 < \pi/2$. For $\delta = \frac{1}{d}$ we get an $O(1)$ bound on ρ_0 that we want to avoid, and so we require $\delta > \frac{1}{d}$. Thus, assuming finally that $|\Gamma_{n(\mathscr{I})}|_{\mathscr{I}} \sim O(\ell_{\rm Pl}^a)$, we get that the number of distinct semiclassical modes that can probe the naked singularity during time evolution through T_- is

$$\frac{|\Gamma_{n(\mathscr{I})}|_{\mathscr{I}}}{\epsilon^d} \sim \ell_{\rm Pl}^{a-d-d\delta}. \tag{4.28}$$

If $a < d-1$, there is always some $\delta > 1/d$ so that this diverges in the $\ell_{\rm Pl} \to 0$ limit. Hence, in the regime $0 < a < d-1$, even when the singularity shrinks to zero size as $\ell_{\rm Pl} \to 0$, the singularity is in principle compatible with influencing an infinite number of EFT modes in the $G_N \to 0$ limit: it is compatible with a gravitational amount of pseudorandomness in the time evolution.

Let us now comment on a few well known cases of naked singularities, some of which were briefly mentioned in Sec. 4.1: the singularity at the endpoint of black hole evaporation, the Gregory–Laflamme (GL) instability [7, 66], which occurs as the pinch-off endpoint of the evolution of higher dimensional extended black strings and other black objects, and Choptuik's critical collapse [71, 72] resulting from a collapsing massless scalar field which is tuned to lie exactly at the threshold of black hole formation. In the case of the endpoint of black hole evaporation, the naked part of the singularity is not extended in the $G_N \to 0$ limit. See Fig. 4.12a. Thus the naked part of the singularity is measure zero in this limit, so this naked singularity is either Planckian or semi-Planckian. In the cases of the Gregory–Laflamme instability and critical collapse, it is impossible to evolve past the pinch-off using classical gravity, so in the absence of a quantum gravitational description of the system, we cannot ascertain the nature of the naked singularity. However, it is quite plausible that these singularities end up being Planckian or semi-Planckian. The critical collapse singularity is in a sense a zero mass black hole, which within the resolution of semiclassical gravity is indistinguishable from a Planck mass black hole. So if we could carry out critical collapse at finite G_N in full nonperturbative quantum gravity, and then afterwards take the $G_N \to 0$ limit, it is reasonable to expect that the singularity becomes a Planckian or semi-Planckian singularity. The Gregory–Laflamme singularity could plausibly behave similarly: it is highly

analogous to the fluid pinch-off phenomenon, where only a minimal amount of nonperturbative evolution is necessary to re-enter the EFT regime. It is tempting to speculate that naked singularities in a holographic theory of quantum gravity can never be classical, and possibly not even semi-Planckian. We will formalize a conjecture to this effect in the next section.

Finally, let us briefly tie this discussion back to our primary theorem of Cryptographic Censorship. The fact that classical singularities are always compatible (by our definition) with gravitational pseudorandomness provides additional motivation for the intuition that the dynamics of strongly gravitating systems in the $G_N \to 0$ limit may admit a description in terms of pseudorandom unitaries. Under this supposition, Cryptographic Censorship guarantees that any such systems — including those with naked singularities — must also have event horizons. We leave the quantification of the subset of the singularity that is necessarily behind a horizon to future work.

We wish to make clear that we do not claim that all classical singularities must act as sources of pseudorandomness; a clear counterexample is given by orbifold singularities in compact dimensions. It is the ones that weak cosmic censorship is supposed to rule out — naked classical singularities resulting from the dynamical evolution of generic, smooth initial data and matter collapse — that we expect to have, to a large extent, gravitationally pseudorandom time evolution. If the singularity is static, the mode counting argument in the previous section becomes misleading. In this case, measuring later modes scattered off the singularity does not provide information we could not have already gained from scattering earlier modes, unlike in the case where the singularity is evolving. Thus, there might be a significantly smaller amount of data that is required to describe scattering off the singularity than what the naïve mode counting argument above would suggest, making static singularities less likely candidate sources of pseudorandom dynamics.

4.4.3 Quantum cosmic censorship

The combination of Cryptographic Censorship and our results above regarding compatibility of naked singularities with gravitational pseudorandomness suggest a particular reformulation of the Weak Cosmic Censorship Conjecture as a *quantum gravity* conjecture rather than a classical gravity conjecture. This point bears emphasizing: there is no *ab initio* reason that classical gravity must obey cosmic censorship. Even if there had been no known counterexamples, one of the best motivations for the Weak Cosmic Censorship Conjecture is empirical: the apparent absence of naked singularities in the night sky. Cosmic censorship is not required by consistency of known theoretical structures within General Relativity. By contrast, if naked singularities were typical in quantum gravity, then foundational results in AdS/CFT — such as the behavior of the CFT thermal state OTOC being precisely attributed to bulk event horizon dynamics — would be some remarkable conspiracy: either naked singularities engineer themselves to mimic black holes precisely in the OTOC signature, or black holes engineer themselves to mimic naked singularities. A similar coincidence would be required for various results on CFT thermalization and its bulk interpretation.[45] We therefore find that a quantum gravity formulation of cosmic censorship is *both* theoretically and empirically well-motivated; our results relating horizons to pseudorandomness provide

[45]We thank D. Harlow for extensive discussions on this point.

a clear first step towards formulating weak cosmic censorship as an emergent phenomenon in terms of fundamental quantities in quantum gravity. (For example, this would require replacing the classical gravity notion of genericity — an open set in the space of solutions to the Einstein equation — with typicality.)

For completeness, we first state the classical gravity proposal for weak cosmic censorship in AdS, which is now known to be violated in classical gravity. (It is also trivially violated in quantum gravity due to, e.g., evaporating black holes, although this requires violations of classical energy conditions.)

Conjecture 1 (Classical Weak Cosmic Censorship – false)**.** *[73, 74] Consider a geodesically complete, asymptotically AdS initial data set satisfying the dominant*[46] *energy condition. The maximal causal development of this initial data under the Einstein equation, together with boundary conditions at the AdS boundary, is generically a strongly asymptotically predictable asymptotically AdS spacetime.*

Reformulating this proposal in terms of a fundamental quantum description of the theory requires a complete revamping of the statement and the scope of its applicability. Motivated by our results in this paper, the existing set of counterexamples to cosmic censorship, and the top down motivation from thermalization and chaos in AdS/CFT, we propose the following formulation of the Weak Cosmic Censorship Conjecture (below we work in the Heisenberg picture, so a given state is its entire history):[47]

Conjecture 2 (Quantum Cosmic Censorship)**.** *Typical states with an emergent geometry in quantum gravity do not have classical naked singularities.*

4.5 Discussion

Cryptographic Censorship is the first precise and general result identifying concrete CFT behavior with bulk horizon formation. (While the *vox populi* associates CFT thermalization to black hole equilibration,[48] this hypothetical link is at best established in certain highly symmetric and nearly static black holes.) Our first application of the theorem guarantees that in the microcanonical window, under assumption of pseudorandomness, states without horizons are exponentially suppressed in probability (within their pseudorandom ensemble).

As an immediate consequence of Cryptographic Censorship, we obtain a partial resolution to the black hole teleology problem for AdS spacetimes with a CFT dual. Event horizons can be diagnosed, using Cryptographic Censorship, in finite time. The implications are significant: first, black hole mechanics in AdS/CFT may admit a legitimate thermodynamic interpretation in highly dynamical states. Second, as our primary focus in this paper, we establish a typicality result for event horizons, which is highly suggestive of a form of Quantum

[46] This is the weakest reasonable statement: we could assume only the null energy condition, but counterexamples exist even when we demand the dominant energy condition.

[47] This conjecture differs slightly from the published version of this paper, in that [75] also conjectures that semi-Planckian singularities are hidden.

[48] Note that black hole equilibration was only recently established in four-dimensional vacuum asymptotically flat space [76] for a certain range of spins and remains an open problem in AdS.

Cosmic Censorship. The fact that naïve models of code subspaces containing classical (or semi-Planckian) timelike singularities are compatible with a sufficient amount of gravitational pseudorandomness serves as strong motivation for our conjecture that typical states do not have naked singularities that survive the classical limit. We emphasize that our proof of Cryptographic Censorship in fact provides a guarantee for event horizon formation whenever the CFT time evolution is exponentially hard to learn; our focus in this article was on complexity of learning as a result of pseudorandomness, but any such sufficient source of complexity would result in an equivalent theorem under the same proof. Below we will elaborate on potential applications of our results and discuss some aspirational expansions thereof.

Towards a Converse

Since information escapes from the black hole interior under inclusion of quantum corrections, there is no obvious definition of a quantum event horizon (since all information eventually reaches $\mathscr{I}$ and causal structures are only emergent in the classical limit). It is tempting to use our results to *define* quantum event horizons — and by extension quantum black holes (i.e., at finite N): states undergoing gravitationally pseudorandom time evolution with a semiclassical bulk dual in the large-N limit. A definition, however, would require a proof of a converse to Cryptographic Censorship: a statement that CFT states with a semiclassical emergent bulk featuring a horizon typically exhibit gravitational pseudorandomness. A proof of such a converse would additionally facilitate a complete resolution of the teleology problem in AdS. The tentative link between chaos, thermalization, and pseudolandomness [34, 65, 77–80] could then potentially be used to understand the horizon thermodynamics[49] of highly dynamical black holes and the statistical interpretation of the Hawking area theorem, which remains elusive [27, 82].

Implications for the Bulk

There is an immediate implication of Cryptographic Censorship for the relationship between complexity and black hole formation: bulk processes that source a large amount of pseudorandomness (e.g., a bulk quantum computer with the power to do gravitationally pseudorandom computations) must typically collapse themselves into black holes. Prior descriptions of black hole formation typically involve energetics; to our knowledge, this is the first proof establishing a causal relation between bulk quantum computations and black hole formation. Interestingly, this appears to bear some relation to the proposal of [31] that effective field theory requires both an energy and a complexity cutoff: a bulk process sourcing large amounts of pseudorandomness must result in high reconstruction complexity. While pseudorandomness is itself not particularly complex to implement, reconstruction of pseudorandom quantities is by definition of high complexity class. In this case, Cryptographic Censorship may act to protect a computationally bounded asymptotic observer from detecting large violations of effective field theory due to computational complexity by forcing formation of an event horizon, depending on where the event horizon forms.

[49] As opposed to the thermodynamics of the apparent horizon, which are by now well-understood [28, 54, 81].

(a) (b)

Figure 4.15: The AdS-big-crunch cosmology (a) versus the flat big-bang cosmology (b). In the asymptotically flat case, there is no notion of unitary time evolution and no horizon. In the AdS case, we have both.

Stringy and Cosmological Applications

It would be interesting to consider stringy effects and ask if (and how) our theorem might apply to, e.g., the D1-D5 system, or orbifold singularities, which are horizonless (see, e.g., [83]). Exact BPS states have trivial dynamics: we expect that to produce chaos and pseudorandom dynamics we would need to sufficiently excite these states. If it is indeed the case that chaotic, pseudorandom time evolution is produced as a result of heavily exciting such states, then our theorem suggests that a horizon will appear. More generally, it is not clear that horizonless geometries have any relation to chaos, and furthermore such states are probably not typical in the measure concentration sense that we are using in this work.

Let us also briefly comment on non-AdS cosmological spacetimes. In particular, it is interesting to consider why our results fail for flat cosmologies, which do not have horizons. An obvious complaint is that our AdS/CFT arguments should not necessarily apply. They do not apply, however, in an interesting manner: the relevant feature of AdS/CFT that fails here is the lack of a notion of unitary time evolution on the conformal boundary (which in this case is null). Our results crucially rely on a fundamental description of the system which is a *unitary* quantum field theory. This in all likelihood does not apply to flat cosmologies. It does, however, apply to AdS big-crunch cosmologies, and in those, we do indeed have a horizon. See Fig. 4.15 for a comparison.

What about de Sitter? de Sitter cosmologies do have cosmological horizons, although those are not unambiguous in the same sense as AdS ones. There have been some suggestions (most recently [84, 85]) that de Sitter time evolution — insofar as such a notion can be defined — is well-modeled by an isometry rather than a unitary. A generalization from pseudorandom unitaries to pseudorandom isometries was recently proposed by [86]. It would be interesting to see if our results can be generalized to accommodate this notion and potentially have some bearing on de Sitter black holes and cosmologies; in particular, if there is any sense in which the relaxation to isometries could be mapped to the observer-dependence of de Sitter horizons.

Some Open Questions

Cryptographic Censorship paves the way for several interesting new investigations. We have expanded on a few of those above. We briefly enumerate a few others here:

1. The fact that gravitational pseudorandomness associated to a quantum computer in the bulk spacetime must result in a horizon suggests a relationship between backreaction and computational power. Does such a relation between energy and computation exist, perhaps analogous to a gravitational version of Lloyd's bound [87]?

2. Are singularities in general in fact associated with large amounts of pseudorandomness?[50] This idea features prominently in our models of timelike naked singularities, but has traditionally appeared in the context of postselection models of black hole singularities and dynamics.

3. Our definition of the size of a singularity admits a middle class of singularities that disappear in the strict larle-N limit but which could accommodate gravitational pseudorandomness in code subspaces constructed from $O(1)$ time bands. It would be interesting to understand the significance, if any, and frequency of semi-Planckian naked singularity portions.

4. Explicit constructions of pseudorandom unitary ensembles to date have not been found. Still, as assumed in this work, we expect explicit examples of AdS/CFT to be well modeled by pseudorandom dynamics. Can we use complexity theoretic conjectures and aspects of AdS/CFT to demonstrate that holographic quantum systems like the Sachdev–Ye–Kitaev (SYK) model actually generate a pseudorandom unitary ensemble at long enough times?

5. Is there anything we can infer about code subspace-dependent cryptographic conditions/constructions necessary for pseudorandomness from the holographic manifestation of our results? That is, does the existence of subspaces without horizons (e.g., the microcanonical window around the vacuum) have any meaningful implication for cryptography?

Acknowledgements

It is a pleasure to thank Jordan Cotler, Jordan Docter, Roberto Emparan, Daniel Harlow, Gary Horowitz, Hsin-Yuan (Robert) Huang, Andrea Puhm, Arvin Shahbazi-Moghaddam, Natalie Paquette, Jon Sorce, Edward Witten, and Michelle Xu for valuable discussions. This work is supported in part by the Department of Energy under Early Career Award DE-SC0021886 (NE, ÅF, and LY) and QuantISED DE-SC0020360 contract no. 578218 (NE), by the Heising-Simons Foundation under grant no. 2023-4430 (NE and AL), the Sloan Foundation (NE), the Templeton Foundation via the Black Hole Initiative (NE and EV), the Gordon and Betty Moore Foundation (EV), the Packard Foundation (AL), and the MIT

[50] See [88] for a unrelated attempt to connect singularities to holographic CA/CV-complexity conjectures.

department of physics (NE). The opinions expressed in this publication are those of the authors and do not necessarily reflect the views of the John Templeton or Moore Foundations.

4.6 Appendix

4.6.1 Proof of Theorem 8

In this appendix, we prove Theorem 8, introduced in Sec. 4.3.2. We begin with a definition of randomized K-Lipschitz, which we will require our algorithms to satisfy.

Definition 8 (Randomized K-Lipschitz)**.** *An algorithm $\mathcal{A}$ is said to be randomized K-Lipschitz if it satisfies the following condition: there exists a constant $K > 0$ such that*

$$\operatorname*{avg}_{\substack{\widehat{U}_1 \leftarrow \mathcal{A}^{U_1} \\ \widehat{U}_2 \leftarrow \mathcal{A}^{U_2}}} ||\widehat{U}_1 - \widehat{U}_2||_\infty \leq K||U_1 - U_2||_\infty \ . \tag{4.29}$$

When the algorithm is deterministic, i.e., it maps a given unitary U to a single output $\widehat{U}$, we will just refer to this property as K-Lipschitz.

In addition, here we will work with operators that are exactly norm-preserving, i.e., $||\widehat{U}||_\infty = 1$. We leave it to the mathematician to confirm that this result applies equally well to approximately norm-preserving operators.

We now repeat Theorem 8 for the reader's convenience.

Theorem 8 (Complexity of Learning for Typical Pseudorandom Unitaries)**.** *Let $\kappa = \mathrm{poly}(N)$ and consider a code subspace of dimension 2^κ. Consider any (gravitationally) pseudorandom unitary ensemble $\mathcal{U}$, any pseudorandom state ensemble Ψ, and any constant $\epsilon > 0$. For any (randomized) K-Lipschitz quantum algorithm $\mathcal{A}$ of $\mathrm{poly}(N)$ complexity that has oracle access to $U \leftarrow \mathcal{U}$ and produces an approximately norm-preserving operator $\widehat{U}$ as its best guess for U, and for any sufficiently large N, the following holds: for a fraction $\geq 1 - e^{-\epsilon^2 2^\kappa/\left(48(1+K)^2\right)} - e^{-\epsilon^2 2^\kappa/192} - \mathrm{negl}'(N)$ of $U \in \mathcal{U}$ and $|\psi\rangle \in \Psi$, we have that:*

1. *the following quantity is bounded*

$$\operatorname*{avg}_{\widehat{U} \leftarrow \mathcal{A}^U} \left[F\left(U|\psi\rangle, \widehat{U}|\psi\rangle\right)\right] \leq 1 - \alpha + 2\epsilon + \mathrm{negl}(N) \ , \tag{4.8}$$

   where α is a constant independent of N, and

2. *for any $\widehat{U} \leftarrow \mathcal{A}^U$, there exists an operator Q (which may depend on U, $|\psi\rangle$, and $\widehat{U}$) such that*

$$\operatorname*{avg}_{\widehat{U} \leftarrow \mathcal{A}^U} \left|\mathrm{tr}\left(Q\widehat{U}\psi(\widehat{U})^\dagger\right) - \mathrm{tr}\left(QU\psi(U)^\dagger\right)\right| \geq 2\alpha - 4\epsilon - \mathrm{negl}(N) \ . \tag{4.9}$$

Proof. We prove 1. and 2. consecutively for Hilbert spaces of any finite (but sufficiently large) dimension d, where we always have in mind that d scales with κ. Note that in the theorem above, and in the main body of the paper, we specialized to a code subspace of dimension $d = 2^\kappa$ with $\kappa = \mathrm{poly}(N)$; here we will leave d general.

1. Fidelity for typical pseudorandom unitaries. We start by proving that the bound on the averaged fidelity in Theorem 7 also holds for *typical* Haar random unitaries and states, and subsequently prove this for pseudorandom unitaries and states. To do so, we employ techniques from measure concentration. An introduction to the theory of measure concentration, including numerous references, is given in Appendix B of [31]. Here we only need the following lemma for measure concentration on the unitary group.

Lemma 3 (Meckes (Corollary 17 in [89]))**.** *Let $f : U(d) \to \mathbb{R}$ be K-Lipschitz in the sense that $|f(U_1) - f(U_2)| \leq K||U_1 - U_2||_2$, with $||X||_2 \equiv \sqrt{\mathrm{tr}\, X^\dagger X}$. Then in the Haar measure on $U(d)$ we have the following concentration inequalities for any $\epsilon > 0$:*

$$\begin{aligned} \Pr[f \geq \langle f|f\rangle + \epsilon] &\leq \exp\left\{\left(-\frac{\epsilon^2 d}{12K^2}\right)\right\}, \\ \Pr[f \leq \langle f|f\rangle - \epsilon] &\leq \exp\left\{\left(-\frac{\epsilon^2 d}{12K^2}\right)\right\}, \\ \Pr[|f - \langle f|f\rangle\,| \geq \epsilon] &\leq 2\exp\left\{\left(-\frac{\epsilon^2 d}{12K^2}\right)\right\}. \end{aligned} \tag{4.30}$$

For the purposes of Theorem 8, we are interested in applying this to the function $F = F(U|\psi\rangle, \widehat{U}|\psi\rangle) = |\left\langle \psi|U^\dagger \widehat{U}|\psi \middle| \psi|U^\dagger \widehat{U}|\psi \right\rangle|^2$, where we first consider Haar random U and $|\psi\rangle$, i.e., $U \leftarrow \mu$ and $|\psi\rangle \leftarrow \mu$, and furthermore $\widehat{U}$ is the guess that the algorithm $\mathcal{A}^U$ outputs. Thus, our task is to show that the function F obeys the requirements for Lemma 3 to hold, such that fluctuations around the average are exponentially suppressed. That is, we need to check that F is K-Lipschitz for some fixed K.

Note that the algorithm's output $\widehat{U}$ depends on U. This dependence is complicated: the quantum algorithm $\mathcal{A}^U$, as a result of, e.g., intrinsic randomness in the outcome of measurements during the learning process, generally outputs a $\widehat{U}$ that is not a deterministic function of U. In the following, we will start by considering instead a deterministic algorithm, i.e., the algorithm is some function $\mathcal{A}(U)$ that maps every unitary U to some norm-preserving operator $\widehat{U}$.

We then prove the following lemma:

Lemma 4. *Let $\mathcal{A}(U)$ denote any deterministic K-Lipschitz function that maps every unitary U to some norm-preserving operator $\widehat{U}$. Let $|\psi\rangle$ and $|\varphi\rangle$ be any two fixed states in the Hilbert space $\mathcal{H}$. Then the function $F(U) \equiv F(U|\psi\rangle, \mathcal{A}(U)|\varphi\rangle)$ is $(2+2K)$-Lipschitz in U.*

Proof. Our aim is to show that

$$|F(U_1) - F(U_2)| \leq (2+2K)||U_1 - U_2||_2 . \tag{4.31}$$

To show (4.31), first, notice that we can rewrite

$$\begin{aligned} F(U) = F(U|\psi\rangle, \mathcal{A}(U)|\varphi\rangle) &= \mathrm{tr}\left[\mathcal{A}(U)|\varphi\rangle\langle\varphi|\mathcal{A}(U)^\dagger U|\psi\rangle\langle\psi|U^\dagger\right] \\ &= \mathrm{tr}\left[(\mathcal{A}(U)^\dagger U)^\dagger \varphi\, \mathcal{A}(U)^\dagger U\, \psi\right] \\ &= \mathrm{tr}\left[\left((\mathcal{A}(U)^\dagger U)^\dagger \otimes \mathcal{A}(U)^\dagger U\right)(\varphi\otimes\psi)\mathsf{S}\right] . \end{aligned} \tag{4.32}$$

Here we used $\varphi = |\varphi\rangle\langle\varphi|$ and $\psi = |\psi\rangle\langle\psi|$. In the third line we introduced a second copy of the Hilbert space; S is the swap operator. For readability, in what follows, we will abbreviate $\mathcal{A}(U_i) = \mathcal{A}_i$. Then, using Hölder's inequality,

$$\begin{aligned} |F(U_1) - F(U_2)| &= \left|\mathrm{tr}\left[\left((\mathcal{A}_1^\dagger U_1)^\dagger \otimes \mathcal{A}_1^\dagger U_1 - (\mathcal{A}_2^\dagger U_2)^\dagger \otimes \mathcal{A}_2^\dagger U_2\right)(\varphi\otimes\psi)\mathsf{S}\right]\right| \\ &\leq ||\varphi\otimes\psi||_1 ||\mathsf{S}\left((\mathcal{A}_1^\dagger U_1)^\dagger \otimes \mathcal{A}_1^\dagger U_1 - (\mathcal{A}_2^\dagger U_2)^\dagger \otimes \mathcal{A}_2^\dagger U_2\right)||_\infty \ . \end{aligned} \tag{4.33}$$

Since $\varphi\otimes\psi$ is a state, its trace norm is 1. Now we can combine submultiplicativity of the operator norm with the fact that S is unitary, such that

$$\begin{aligned} &||\mathsf{S}\big((\mathcal{A}_1^\dagger U_1)^\dagger \otimes \mathcal{A}_1^\dagger U_1 - (\mathcal{A}_2^\dagger U_2)^\dagger \otimes \mathcal{A}_2^\dagger U_2\big)||_\infty \\ &\qquad \leq ||\mathsf{S}||_\infty ||\big((\mathcal{A}_1^\dagger U_1)^\dagger \otimes \mathcal{A}_1^\dagger U_1 - (\mathcal{A}_2^\dagger U_2)^\dagger \otimes \mathcal{A}_2^\dagger U_2\big)||_\infty \\ &\qquad = ||(\mathcal{A}_1^\dagger U_1)^\dagger \otimes \mathcal{A}_1^\dagger U_1 - (\mathcal{A}_2^\dagger U_2)^\dagger \otimes \mathcal{A}_2^\dagger U_2||_\infty \ . \end{aligned} \tag{4.34}$$

Recall that we want to show that this is upper bounded by some $K'||U_1 - U_2||_2$. Rewrite the last line as

$$\begin{aligned} &||(\mathcal{A}_1^\dagger U_1)^\dagger \otimes \mathcal{A}_1^\dagger U_1 - (\mathcal{A}_1^\dagger U_1)^\dagger \otimes \mathcal{A}_2^\dagger U_2 + (\mathcal{A}_1^\dagger U_1)^\dagger \otimes \mathcal{A}_2^\dagger U_2 - (\mathcal{A}_2^\dagger U_2)^\dagger \otimes \mathcal{A}_2^\dagger U_2||_\infty \\ &\quad \leq ||(\mathcal{A}_1^\dagger U_1)^\dagger \otimes \mathcal{A}_1^\dagger U_1 - (\mathcal{A}_1^\dagger U_1)^\dagger \otimes \mathcal{A}_2^\dagger U_2||_\infty + ||(\mathcal{A}_1^\dagger U_1)^\dagger \otimes \mathcal{A}_2^\dagger U_2 - (\mathcal{A}_2^\dagger U_2)^\dagger \otimes \mathcal{A}_2^\dagger U_2||_\infty \\ &\quad = 2||\mathcal{A}_1^\dagger U_1 - \mathcal{A}_2^\dagger U_2||_\infty \ . \end{aligned} \tag{4.35}$$

In the second line we used the triangle inequality; in the third we used $||A\otimes B|| = ||A||\,||B||$, and for unitaries $||U||_\infty = 1$, and we are we are working with exactly norm-preserving operators $\widehat{U}$ such that also $||\widehat{U}||_\infty = 1$. We now again use the triangle inequality and submultiplicativity of the operator norm to rewrite the final line as

$$\begin{aligned} 2||\mathcal{A}_1^\dagger U_1 - \mathcal{A}_2^\dagger U_2||_\infty &= 2||\mathcal{A}_1^\dagger U_1 - \mathcal{A}_2^\dagger U_1 + \mathcal{A}_2^\dagger U_1 - \mathcal{A}_2^\dagger U_2||_\infty \\ &\leq 2||\mathcal{A}_1^\dagger - \mathcal{A}_2^\dagger||_\infty ||U_1||_\infty + 2||\mathcal{A}_2^\dagger||_\infty ||U_1 - U_2||_\infty \\ &\leq (2+2K)||U_1 - U_2||_\infty \\ &\leq (2+2K)||U_1 - U_2||_2 \ . \end{aligned} \tag{4.36}$$

In the third line we used that $\mathcal{A}(U)$ is K-Lipschitz for some K, and finally we used $||\cdot||_\infty \leq ||\cdot||_2$. This completes the proof that the function $F(U) \equiv F(U|\psi\rangle, \mathcal{A}(U)|\varphi\rangle)$ is $(2+2K)$-Lipschitz:

$$|F(U_1) - F(U_2)| \leq (2+2K)||U_1 - U_2||_2 \ . \tag{4.37}$$

□

Then, by Lemma 3 and Lemma 4, we have the following measure concentration result: that for fraction $\geq 1 - e^{-\epsilon^2 d/(48(1+K)^2)}$ of Haar random unitaries,

$$F(U|\psi\rangle, \mathcal{A}(U)|\varphi\rangle) - \operatorname*{avg}_{V\leftarrow\mu}\left[F(V|\psi\rangle, \mathcal{A}(V)|\varphi\rangle)\right] < \epsilon \ . \tag{4.38}$$

We now wish to extend this to include measure concentration over Haar random states. To this end, we prove the following lemma.

Lemma 5. *Let $\mathcal{A}(U)$ denote any deterministic K-Lipschitz function that maps every unitary U to some norm-preserving operator $\widehat{U}$. For fraction $\geq 1 - e^{-\epsilon^2 d/(48(1+K)^2)}$ of Haar random unitaries $U \leftarrow \mu$ and fraction $\geq 1 - e^{-\epsilon^2 d/192}$ of Haar random states $|\psi\rangle \leftarrow \mu$, we have*

$$F(U|\psi\rangle, \mathcal{A}(U)|\psi\rangle) - \operatorname*{avg}_{\substack{V\leftarrow\mu \\ |\varphi\rangle\leftarrow\mu}} [F(V|\varphi\rangle, \mathcal{A}(V)|\varphi\rangle)] < 2\epsilon \ . \tag{4.39}$$

Proof. Consider first fixing some U and $\mathcal{A}(U)$. Let V be a Haar random unitary and $|\psi_0\rangle$ some basis state, such that $V|\psi_0\rangle$ is a Haar random state. We show that the function

$$F(V) \equiv F(UV|\psi_0\rangle, \mathcal{A}(U)V|\psi_0\rangle) = F(V^\dagger \mathcal{A}(U)^\dagger UV|\psi_0\rangle, |\psi_0\rangle) \tag{4.40}$$

is 4-Lipschitz. Following the same methods as in the proof of Lemma 4, we find the analogue of (4.35):

$$|F(V_1) - F(V_2)| \leq 2||V_1^\dagger \mathcal{A}(U)^\dagger UV_1 - V_2^\dagger \mathcal{A}(U)^\dagger UV_2||_\infty \ . \tag{4.41}$$

Then, by the triangle inequality,

$$\begin{aligned} |F(V_1) - F(V_2)| &\leq 2||V_1^\dagger \mathcal{A}(U)^\dagger UV_1 - V_2^\dagger \mathcal{A}(U)^\dagger UV_1 + V_2^\dagger \mathcal{A}(U)^\dagger UV_1 - V_2^\dagger \mathcal{A}(U)^\dagger UV_2||_\infty \\ &\leq 2||V_1^\dagger - V_2^\dagger||_\infty ||\mathcal{A}(U)^\dagger UV_1||_\infty + 2||V_2^\dagger \mathcal{A}(U)^\dagger U||_\infty ||V_1 - V_2||_\infty \\ &\leq 4||V_1 - V_2||_2 \ . \end{aligned} \tag{4.42}$$

Note that in the last line we used again that $\mathcal{A}(U)$ is exactly norm-preserving, such that $||\mathcal{A}(U)^\dagger UV_i||_\infty = 1$, as well as $||\cdot||_\infty \leq ||\cdot||_2$.

By Lemma 3, for fraction $\geq 1 - e^{-\epsilon^2 d/192}$ of $|\psi\rangle \equiv V|\psi_0\rangle \leftarrow \mu$,

$$F(U|\psi\rangle, \mathcal{A}(U)|\psi\rangle) - \operatorname*{avg}_{|\varphi\rangle\leftarrow\mu} [F(U|\varphi\rangle, \mathcal{A}(U)|\varphi\rangle)] < \epsilon \ . \tag{4.43}$$

We now combine this with measure concentration for Haar random unitaries. Starting from (4.38) for two (fixed) states $|\varphi\rangle$, and then averaging over $\varphi \leftarrow \mu$, we have for fraction $\geq 1 - e^{-\epsilon^2 d/48(1+K)^2}$ of $U \leftarrow \mu$,

$$\operatorname*{avg}_{|\varphi\rangle\leftarrow\mu} [F(U|\varphi\rangle, \mathcal{A}(U)|\varphi\rangle] - \operatorname*{avg}_{\substack{V\leftarrow\mu \\ |\varphi\rangle\leftarrow\mu}} [F(V|\varphi\rangle, \mathcal{A}(U)|\varphi\rangle)] < \epsilon \ . \tag{4.44}$$

Now add and subtract $F(U|\psi\rangle, \mathcal{A}(U)|\psi\rangle)$. This gives

$$\begin{aligned} F(U|\psi\rangle, \mathcal{A}(U)|\psi\rangle) - \operatorname*{avg}_{\substack{V\leftarrow\mu \\ |\varphi\rangle\leftarrow\mu}} [F(V|\varphi\rangle, \mathcal{A}(U)|\varphi\rangle)] \\ < \epsilon + F(U|\psi\rangle, \mathcal{A}(U)|\psi\rangle) - \operatorname*{avg}_{|\varphi\rangle\leftarrow\mu} [F(U|\varphi\rangle, \mathcal{A}(U)|\varphi\rangle)] \ . \end{aligned} \tag{4.45}$$

Now we can use (4.43) to bound the right hand side to 2ϵ for fraction $\geq 1 - e^{-\epsilon^2 d/192}$ of $|\psi\rangle \leftarrow \mu$. This concludes the proof of Lemma 5.

□

To prove Theorem 8, it remains to extend the above results to (1) non-deterministic functions $\mathcal{A}(U)$, e.g., quantum algorithms $\mathcal{A}^U$, and (2) to pseudorandom unitaries and states. We first briefly address point (1). We can run the proofs of Lemma 4 and Lemma 5 analogously for $\widehat{U}_i \leftarrow \mathcal{A}^{U_i}$ until we need a smoothness property of $\widehat{U}_i$. There, instead of K-Lipschitz, we use the randomized K-Lipschitz property in Definition 8. Furthermore, using the inequality

$$|\operatorname{avg} F(U_1) - \operatorname{avg} F(U_2)| \leq \operatorname{avg} |F(U_1) - F(U_2)| \,, \tag{4.46}$$

we have that the function $F(V) = \operatorname*{avg}_{\widehat{V}\leftarrow\mathcal{A}^V} \left[F(V|\psi\rangle, \widehat{V}|\psi\rangle)\right]$ is itself $(2+2K)$-Lipschitz. This allows us to extend our measure concentration results to this function $F(V)$ where previously we proved them for $F(U) = F(U|\psi\rangle, \widehat{U}|\psi\rangle)$.

Then we arrive at the following result: that for fraction $\geq 1 - e^{-\epsilon^2 d/(48(1+K)^2)}$ of $U \leftarrow \mu$, and for fraction $\geq 1 - e^{-\epsilon^2 d/192}$ of $|\psi\rangle \leftarrow \mu$ (alternatively, we can think of this as $\geq 1 - e^{-\epsilon^2 d/192} - e^{-\epsilon^2 d/(48(1+K)^2)}$ fraction of $U \leftarrow \mu$ and $|\psi\rangle \leftarrow \mu$), we have the following concentration bound:

$$\operatorname*{avg}_{\widehat{U}\leftarrow\mathcal{A}^U} \left[F(U|\psi\rangle, \widehat{U}|\psi\rangle)\right] - \operatorname*{avg}_{\substack{V\leftarrow\mu \\ |\varphi\rangle\leftarrow\mu}} \left[\operatorname*{avg}_{\widehat{V}\leftarrow\mathcal{A}^V} \left[F(V|\varphi\rangle, \widehat{V}|\varphi\rangle)\right]\right] < 2\epsilon \,. \tag{4.47}$$

Finally, we address point (2) above, and show that to preserve indistinguishability, the above result must also hold for *pseudorandom* unitaries and states. Intuitively, if the fluctuations around the Haar random averaged fidelity are exponentially suppressed, but the fluctuations around the pseudorandom averaged fidelity are not, then one could measure these fluctuations and thereby distinguish pseudorandom unitaries and states from Haar random ones, breaking the pseudorandomness.

We proceed to formally show this below.

Lemma 6. *Let $\kappa = \operatorname{poly}(N)$ be the security parameter. Let $\mathcal{U}$ be any pseudorandom unitary (PRU) ensemble and let Ψ be any pseudorandom state (PRS) ensemble. Let $\widehat{U}$ be the output of the learning algorithm $\mathcal{A}^U$, which we assume to be (randomized) K-Lipschitz, and where $U \leftarrow \mathcal{U}$. Then the following concentration inequality holds, in the measure induced on $\mathcal{U}$ and Ψ, for any $\epsilon > 0$, with probability $\geq 1 - e^{-\epsilon^2 d/48(1+K)^2} - e^{-\epsilon^2 d/192} - \operatorname{negl}'(N)$ over $U \leftarrow \mathcal{U}$ and $|\psi\rangle \leftarrow \Psi$,*

$$\operatorname*{avg}_{\widehat{U}\leftarrow\mathcal{A}^U} F\big(U|\psi\rangle, \widehat{U}|\psi\rangle\big) - \operatorname*{avg}_{\substack{V\leftarrow\mathcal{U} \\ |\varphi\rangle\leftarrow\Psi \\ \widehat{V}\leftarrow\mathcal{A}^V}} \big[F\big(V|\varphi\rangle, \widehat{V}|\varphi\rangle\big)\big] < 2\epsilon + \operatorname{negl}(N) \,. \tag{4.48}$$

Proof. To prove Lemma 6, we first extend (4.47) to pseudorandom unitaries, and then to pseudorandom states.

First, let Ψ be any ensemble of states. For ease of notation, we write

$$\beta := \operatorname*{avg}_{\substack{U\leftarrow\mu \\ |\psi\rangle\leftarrow\Psi \\ \widehat{U}\leftarrow\mathcal{A}^U}} \left[F(U|\psi\rangle, \widehat{U}|\psi\rangle)\right] \tag{4.49}$$

and

$$\beta' := \operatorname*{avg}_{\substack{U\leftarrow\mathcal{U}\\ |\psi\rangle\leftarrow\Psi\\ \widehat{U}\leftarrow\mathcal{A}^U}} \left[F(U|\psi\rangle, \widehat{U}|\psi\rangle)\right] . \tag{4.50}$$

By Lemma 7 of [32], $|\beta - \beta'| \leq \eta(N)$ where η is a negligible function.

Assume towards contradiction that there exists a pseudorandom unitary ensemble $\mathcal{U}$, any distribution of states Ψ, and a positive polynomial p such that

$$\Pr_{\substack{U\leftarrow\mathcal{U}\\ |\psi\rangle\leftarrow\Psi}}\left[\operatorname*{avg}_{\widehat{U}\leftarrow\mathcal{A}^U} F\left(U|\psi\rangle, \widehat{U}|\psi\rangle\right) - \beta' \geq \epsilon + \eta(N)\right] \geq e^{-\epsilon^2 d/48(1+K)^2} + \frac{1}{p(N)} . \tag{4.51}$$

If the above equation holds, then there is a state $|\psi_0\rangle$ such that this equation holds with $|\psi\rangle := |\psi_0\rangle$:

$$\Pr_{U\leftarrow\mathcal{U}}\left[\operatorname*{avg}_{\widehat{U}\leftarrow\mathcal{A}^U} F\left(U|\psi_0\rangle, \widehat{U}|\psi_0\rangle\right) - \beta' \geq \epsilon + \eta(N)\right] \geq e^{-\epsilon^2 d/48(1+K)^2} + \frac{1}{p(N)} . \tag{4.52}$$

We now construct an efficient algorithm Dist that uses oracle access to the pseudorandom unitary U and also can run the learning algorithm $\mathcal{A}^U$.[51] We show that if our assumption (4.51) is true, then Dist can distinguish between $U \leftarrow \mu$ and $U \leftarrow \mathcal{U}$, thus breaking the pseudorandomness of $\mathcal{U}$.

The distinguisher Dist gets $\mathrm{poly}(N)$ copies of $|\psi_0\rangle$, and is also handed β.[52] Then it proceeds as follows:

1. Dist runs $\mathcal{A}$ and simulates the oracle for $\mathcal{A}$: it receives each query from $\mathcal{A}$, sends the query to its oracle U, and returns the response to $\mathcal{A}$.

2. Dist gives $|\psi_0\rangle$ to $\mathcal{A}$ and receives back $\widehat{U}|\psi_0\rangle$. It also gives $|\psi_0\rangle$ to its oracle U and gets back $U|\psi_0\rangle$.

3. Dist repeats this $\mathrm{poly}(N)$ times and then uses a quantum algorithm (e.g., [90]) to efficiently estimate $\beta'' := F(U|\psi_0\rangle, \widehat{U}|\psi_0\rangle)$ up to some additive error $\hat{\epsilon} > 0$, i.e., $|\beta''_{\text{est}} - \beta''| \leq \hat{\epsilon}$.

4. If $\beta''_{\text{est}} - \beta \geq \tilde{\epsilon}$, where $\tilde{\epsilon} > \hat{\epsilon}$, then Dist outputs 1.

Note that Dist runs in $\mathrm{poly}(N)$-time since it can efficiently query the oracle, and furthermore the quantum algorithm for fidelity estimation is efficient.[53]

[51] Note that the distinguisher can depend on N, as ours does here, since e.g. $|\psi_0\rangle$ depends on N.

[52] The careful reader might worry that handing the distinguisher (multiple copies of) *any* state in $\mathcal{H}$, or β, breaks the computational boundedness of the algorithm. Note, however, that although the algorithms we consider are limited to a polynomial number of queries in $\log\dim\mathcal{H}$, each query can be for any state in $\mathcal{H}$ that the algorithm a) can prepare or b) has been given as "advice". This advice can also include quantities like β. Note that the learning algorithm we consider in, e.g, Theorem 8 gets only *one* copy of the physical state $|\psi\rangle$; the distinguisher that we describe here is a different algorithm and can receive multiple copies as advice.

[53] For example, the quantum algorithm of [90] runs in $\mathrm{poly}(1/\hat{\epsilon})$-time.

We now show that Dist distinguishes between $U \leftarrow \mu$ and $U \leftarrow \mathcal{U}$. For Haar random $U \leftarrow \mu$ as the oracle,

$$\begin{aligned}\Pr_{U\leftarrow\mu}[\mathsf{Dist}=1] &= \Pr\left[\beta''_{\text{est}} - \beta \geq \tilde{\epsilon}\right] = \Pr\left[\beta'' - \beta \geq \tilde{\epsilon} - (\beta''_{\text{est}} - \beta'')\right] \\ &\leq \Pr\left[\beta'' - \beta \geq \tilde{\epsilon} - \hat{\epsilon}\right] \\ &\leq e^{-(\tilde{\epsilon}-\hat{\epsilon})^2 d/48(1+K)^2} \ ,\end{aligned} \tag{4.53}$$

where we used $-\hat{\epsilon} \leq \beta''_{\text{est}} - \beta'' \leq \hat{\epsilon}$ in the second line, and (4.38) (concentration in the Haar measure) for $\epsilon = \tilde{\epsilon} - \hat{\epsilon} > 0$ in the last line.

For pseudorandom $U \leftarrow \mathcal{U}$ as the oracle,

$$\begin{aligned}\Pr_{U\leftarrow\mathcal{U}}[\mathsf{Dist}=1] &= \Pr\left[\beta''_{\text{est}} - \beta \geq \tilde{\epsilon}\right] = \Pr\left[\beta'' - \beta' \geq \tilde{\epsilon} - (\beta''_{\text{est}} - \beta'') - (\beta' - \beta)\right] \\ &\geq \Pr\left[\beta'' - \beta' \geq \tilde{\epsilon} + \hat{\epsilon} + \eta(N)\right] \\ &\geq e^{-(\tilde{\epsilon}+\hat{\epsilon})^2 d/48(1+K)^2} + \frac{1}{p(N)} \ ,\end{aligned} \tag{4.54}$$

where we used $-\eta(N) \leq \beta - \beta' \leq \eta(N)$ in the third line, and our assumption (4.51) in the last line for $\epsilon = \tilde{\epsilon} + \hat{\epsilon}$.

Therefore,

$$\Pr_{U\leftarrow\mathcal{U}}[\mathsf{Dist}=1] - \Pr_{U\leftarrow\mu}[\mathsf{Dist}=1] \geq \frac{1}{p(N)} + e^{-\frac{(\tilde{\epsilon}+\hat{\epsilon})^2 d}{48(1+K)^2}} - e^{-\frac{(\tilde{\epsilon}-\hat{\epsilon})^2 d}{48(1+K)^2}} \ . \tag{4.55}$$

This shows that Dist distinguishes between $U \leftarrow \mu$ and $U \leftarrow \mathcal{U}$. Since Dist runs in poly(N) time, this contradicts the pseudorandomness of $\mathcal{U}$ (Definition 5). Thus based on the pseudorandomness of $\mathcal{U}$, we must have that for fraction $\leq e^{-\epsilon^2 d/48(1+K)^2} + \eta'(N)$ of $U \leftarrow \mathcal{U}$

$$\operatorname*{avg}_{\widehat{U}\leftarrow\mathcal{A}^U} F\left(U\left|\psi\right\rangle, \widehat{U}\left|\psi\right\rangle\right) - \operatorname*{avg}_{\substack{V\leftarrow\mathcal{U} \\ |\varphi\rangle\leftarrow\Psi \\ \widehat{V}\leftarrow\mathcal{A}^V}} F\left(V\left|\varphi\right\rangle, \widehat{V}\left|\varphi\right\rangle\right) \geq \epsilon + \eta(N) \ , \tag{4.56}$$

where $\eta'(N)$ is negligible. In particular, this holds when $|\psi\rangle \leftarrow \Psi$ is Haar random.

Thus, we have proven a measure concentration result for fluctuations around the average fidelity where we average over pseudorandom unitaries (and Haar random states). An analogous proof holds for fluctuations around the average fidelity where we average over pseudorandom states (and any distribution of unitaries $\mathcal{D}$; in particular, it holds for Haar random unitaries $U \leftarrow \mu$). It requires Lemma 8 in [32]. This would result in the following measure concentration result: that for fraction $\leq e^{-\epsilon^2 d/192} + \tilde{\eta}'(N)$ of $|\psi\rangle \leftarrow \Psi$,

$$\operatorname*{avg}_{\widehat{U}\leftarrow\mathcal{A}^U} F\left(U\left|\psi\right\rangle, \widehat{U}\left|\psi\right\rangle\right) - \operatorname*{avg}_{\substack{V\leftarrow\mathcal{U} \\ |\varphi\rangle\leftarrow\Psi \\ \widehat{V}\leftarrow\mathcal{A}^V}} \left[F\left(V\left|\varphi\right\rangle, \widehat{V}\left|\varphi\right\rangle\right)\right] \geq \epsilon + \tilde{\eta}(N) \ , \tag{4.57}$$

where we explicated that the negligible functions need not be the same as in (4.56). Combining (4.56) and (4.57) concludes the proof of Lemma 6. □

We will now use Lemma 6 to prove the first part of Theorem 8. Consider any constant $\epsilon > 0$. Consider any sufficiently large value of $N \in \mathbb{N}$ so that Theorem 7 from [32] holds. Let $U \leftarrow \mathcal{U}$ and $|\psi\rangle \leftarrow \Psi$ denote the pseudorandom unitary and state. Let $\widehat{U}$ denote the operator that $\mathcal{A}^U$ produces. To connect to the learning result (Theorem 7), we average over the output of the algorithm, $\widehat{U}|\psi\rangle \leftarrow \mathcal{A}^U(|\psi\rangle)$. We abbreviate this by $\widehat{U} \leftarrow \mathcal{A}^U$. By Lemma 6, we have that with probability $\geq 1 - e^{-\epsilon^2 d/48(1+K)^2} - e^{-\epsilon^2 d/192} - \mathrm{negl}'(N)$,

$$\operatorname*{avg}_{\widehat{U}\leftarrow\mathcal{A}^U} \left[F\big(U|\psi\rangle, \widehat{U}|\psi\rangle\big)\right] - \operatorname*{avg}_{\substack{V\leftarrow\mathcal{U}\\ |\varphi\rangle\leftarrow\Psi\\ \widehat{V}\leftarrow\mathcal{A}^V}} \left[F\big(V|\varphi\rangle, \widehat{V}|\varphi\rangle\big)\right] < 2\epsilon + \mathrm{negl}(N) \ . \tag{4.58}$$

We call $(U, |\psi\rangle)$ *typical* if the above equation holds. The complexity of learning result (Theorem 7) gives us an upper bound on the average fidelity. Thus we have that for any typical $(U, |\psi\rangle)$,

$$\operatorname*{avg}_{\widehat{U}\leftarrow\mathcal{A}^U} \left[F\left(U|\psi\rangle, \widehat{U}|\psi\rangle\right)\right] \leq 1 - \alpha + 2\epsilon + \mathrm{negl}(N) \ . \tag{4.59}$$

This shows the first part of the theorem (Equation (4.8)).

Haar random unitaries and states. Note that we obtain a stronger result for Haar random unitaries and states: namely, that for fraction $\geq 1 - e^{-\epsilon^2 d/48(1+K)^2} - e^{-\epsilon^2 d/192}$ of $U \leftarrow \mu$ and $|\psi\rangle \leftarrow \mu$,

$$\operatorname*{avg}_{\widehat{U}\leftarrow\mathcal{A}^U} \left[F\left(U|\psi\rangle, \widehat{U}|\psi\rangle\right)\right] \leq 1 - \alpha + 2\epsilon \ . \tag{4.60}$$

Note that for $d = 2^\kappa$ with $\kappa = \mathrm{poly}(N)$, this means that the above bound is violated for only a fraction that is doubly-exponentially suppressed in N.

2. Distinguishing operator. We now show the second part of the theorem. The relationship between fidelity and distinguishing measurements is a known result; for any finite dimensional Hilbert space, this is given in Appendix B in [32]. Thus, for any typical $(U, |\psi\rangle)$, there exists an operator Q (for a given $\widehat{U}$) for which

$$\operatorname*{avg}_{\widehat{U}\leftarrow\mathcal{A}^U} \left|\mathrm{tr}\left(Q\widehat{U}\psi\widehat{U}^\dagger\right) - \mathrm{tr}\left(QU\psi U^\dagger\right)\right| \geq 2\alpha - 4\epsilon - \mathrm{negl}(N) \ . \tag{4.61}$$

This concludes the proof of Theorem 8.

□

4.6.2 Proof of Corollary 4.3.2

In this appendix, we prove Corollary 4.3.2, the existence of a semiclassical distinguishing operator. We repeat Corollary 4.3.2 here for convenience:

[Distinguishing Operator at $N \to \infty$] For any two sequences of states $\{U^{(N)}|\psi^{(N)}\rangle\}_N$ and $\{\widehat{U}^{(N)}|\psi^{(N)}\rangle\}_N$ that have well-defined limits $|\xi^{(\infty)}\rangle$ and $|\widehat{\xi}^{(\infty)}\rangle$, respectively, as $N \to \infty$ and that satisfy the assumptions stated above, both 1. and 2. in Theorem 8 also hold for the

states $\left|\xi^{(\infty)}\right\rangle$ and $\left|\widehat{\xi}^{(\infty)}\right\rangle$. In particular, by taking the constant ϵ to be $< \alpha/2$, we have for any $\left|\widehat{\xi}^{(\infty)}\right\rangle$, the existence of an operator $Q^{(\infty)}$ such that,

$$\operatorname*{avg}_{\widehat{\xi}^{(\infty)}} \left| \operatorname{tr}\left(Q^{(\infty)}\widehat{\xi}^{(\infty)}\right) - \operatorname{tr}\left(Q^{(\infty)}\xi^{(\infty)}\right)\right| \geq O(1) \ .$$

Proof. Note that by the assumptions on the sequences $\{U^{(N)}\left|\psi^{(N)}\right\rangle\}_N$ and $\{\widehat{U}^{(N)}\left|\psi^{(N)}\right\rangle\}_N$, for every sufficiently large N, $U^{(N)}\left|\psi^{(N)}\right\rangle$ and $\widehat{U}^{(N)}\left|\psi^{(N)}\right\rangle$ satisfy 1. and 2. in Theorem 8.

We first show that the result (4.59) remains true at infinite N. By assumption, the two sequences of states $\{U^{(N)}\left|\psi^{(N)}\right\rangle\}_N$ and $\{\widehat{U}^{(N)}\left|\psi^{(N)}\right\rangle\}_N$ have well-defined limits $\left|\widehat{\xi}^{(\infty)}\right\rangle$ and $\left|\xi^{(\infty)}\right\rangle$ as $N \to \infty$. The fidelity itself is well-defined for infinite-dimensional operators of bounded trace norm. Note also that the right-hand side of (4.59) becomes independent of N as $N \to \infty$. Therefore, the result (4.59) should hold for $\left|\widehat{\xi}^{(\infty)}\right\rangle$ and $\left|\xi^{(\infty)}\right\rangle$.

We now turn to the distinguishing operator. Let us first work at finite N. The Fuchs–van de Graaf inequality [91] states that the fidelity of two density operators ρ and σ (in finite dimensions) upper and lower bounds their trace norm distance:

$$1 - \sqrt{F(\rho,\sigma)} \leq \frac{1}{2}||\rho - \sigma||_1 \leq \sqrt{1 - F(\rho,\sigma)} \ . \tag{4.62}$$

When σ is a pure state, the lower bound can be strengthened:

$$1 - F(\rho,\sigma) \leq \frac{1}{2}||\rho - \sigma||_1 \ . \tag{4.63}$$

By this equation and (4.59), for any $(U^{(N)}, \left|\psi^{(N)}\right\rangle)$ which is typical for the distributions $\mathcal{U}^{(N)}$, $\Psi^{(N)}$, we get the following, where we keep the average over $\widehat{U} \leftarrow \mathcal{A}^U$ implicit in what will follow,

$$\left|\left|\widehat{U}^{(N)}\psi^{(N)}(\widehat{U}^{(N)})^\dagger - U^{(N)}\psi^{(N)}(U^{(N)})^\dagger\right|\right|_1 \geq 2\alpha - 4\epsilon - \operatorname{negl}(N) \ . \tag{4.64}$$

Let $\xi^{(N)} = U^{(N)}\psi^{(N)}(U^{(N)})^\dagger$ and $\hat{\xi}^{(N)} = \widehat{U}^{(N)}\psi^{(N)}(\widehat{U}^{(N)})^\dagger$. Below we also write ξ and $\hat{\xi}$ as shorthand for $\xi^{(\infty)}$ and $\widehat{\xi}^{(\infty)}$. Using this, we find that

$$\begin{aligned}
||\hat{\xi} - \xi||_1 &= ||(\hat{\xi} - \hat{\xi}^{(N)}) - (\xi - \xi^{(N)}) + (\xi^{(N)} - \hat{\xi}^{(N)})||_1 \\
&\geq \left| ||\hat{\xi}^{(N)} - \xi^{(N)}||_1 - ||(\hat{\xi} - \hat{\xi}^{(N)}) - (\xi - \xi^{(N)})||_1 \right| \\
&= ||\hat{\xi}^{(N)} - \xi^{(N)}||_1 - ||(\hat{\xi} - \hat{\xi}^{(N)}) - (\xi - \xi^{(N)})||_1 \\
&\geq ||\hat{\xi}^{(N)} - \xi^{(N)}||_1 - ||\hat{\xi} - \hat{\xi}^{(N)}||_1 - ||\xi - \xi^{(N)}||_1 \\
&\geq 2\alpha - 4\epsilon - \operatorname{negl}(N) - ||\hat{\xi} - \hat{\xi}^{(N)}||_1 - ||\xi - \xi^{(N)}||_1.
\end{aligned} \tag{4.65}$$

where we used the reverse triangle inequality $||x - y|| \geq \left|||x|| - ||y||\right|$ in the second line, and the triangle inequality $||x - y|| \leq ||x|| + ||y||$ in the fourth line. In the third line, we use that the last term goes to zero as $N \to \infty$ to remove the absolute value. Taking now the $N \to \infty$

limit, the last two terms vanish by assumption, and $\text{negl}(N) \to 0$, giving $||\hat{\xi} - \xi||_1 \geq 2\alpha - 4\epsilon$. Now, if $\mathcal{A}$ is the algebra of all bounded operators on the Hilbert space $\mathcal{H}^{(\infty)}$, one has that

$$||\rho - \sigma||_1 = \sup_{a \in \mathcal{A}: ||a||_\infty \leq 1} |\operatorname{tr}(a\rho) - \operatorname{tr}(a\sigma)| \tag{4.66}$$

Thus, reinstating averages, there exists some bounded operator Q on $\mathcal{H}^{(\infty)}$, depending on $\widehat{U}$, such that

$$\operatorname*{avg}_{\widehat{\xi}} \left| \operatorname{tr}\left(Q\widehat{\xi}\right) - \operatorname{tr}(Q\xi) \right| \geq 2\alpha - 4\epsilon \ . \tag{4.67}$$

In particular, by taking ϵ to be any constant smaller than $\alpha/2$, we get that there exists an operator whose expectation values for $\widehat{\xi}$ and ξ must differ by a constant. This concludes the proof of Corollary 4.3.2. □

References

[1] R. Penrose, *Gravitational collapse and space-time singularities*, *Phys. Rev. Lett.* **14** (1965) 57–59.

[2] S. W. Hawking, *Singularities in the universe*, *Phys. Rev. Lett.* **17** (Aug, 1966) 444–445.

[3] S. W. Hawking and R. Penrose, *The Singularities of gravitational collapse and cosmology*, *Proc. Roy. Soc. Lond. A* **314** (1970) 529–548.

[4] R. Penrose, *Gravitational collapse and space-time singularities*, *Phys. Rev. Lett.* **14** (Jan, 1965) 57–59.

[5] S. W. Hawking, *Gravitational radiation from colliding black holes*, *Phys. Rev. Lett.* **26** (1971) 1344–1346.

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

[7] R. Gregory and R. Laflamme, *The Instability of charged black strings and p-branes*, *Nucl. Phys. B* **428** (1994) 399–434, [hep-th/9404071].

[8] L. Lehner and F. Pretorius, *Black Strings, Low Viscosity Fluids, and Violation of Cosmic Censorship*, *Phys. Rev. Lett.* **105** (2010) 101102, [arXiv:1006.5960].

[9] R. Gregory, *The gregory-laflamme instability*, arXiv:1107.5821.

[10] G. T. Horowitz, J. E. Santos, and B. Way, *Evidence for an Electrifying Violation of Cosmic Censorship*, *Class. Quant. Grav.* **33** (2016), no. 19 195007, [arXiv:1604.06465].

[11] T. Crisford and J. E. Santos, *Violating the Weak Cosmic Censorship Conjecture in Four-Dimensional Anti–de Sitter Space*, *Phys. Rev. Lett.* **118** (2017), no. 18 181101, [arXiv:1702.05490].

[12] T. Crisford, G. T. Horowitz, and J. E. Santos, *Testing the Weak Gravity - Cosmic Censorship Connection*, *Phys. Rev. D* **97** (2018), no. 6 066005, [arXiv:1709.07880].

[13] G. T. Horowitz and J. E. Santos, *Further evidence for the weak gravity — cosmic censorship connection*, *JHEP* **06** (2019) 122, [arXiv:1901.11096].

[14] Å. Folkestad, *Penrose Inequality as a Constraint on the Low Energy Limit of Quantum Gravity*, *Phys. Rev. Lett.* **130** (2023), no. 12 121501, [arXiv:2209.00013].

[15] F. C. Eperon, B. Ganchev, and J. E. Santos, *Plausible scenario for a generic violation of the weak cosmic censorship conjecture in asymptotically flat four dimensions*, *Phys. Rev. D* **101** (2020), no. 4 041502, [arXiv:1906.11257].

[16] J. Bardeen, *Non-singular general relativistic gravitational collapse*, in *Proceedings of the 5th International Conference on Gravitation and the Theory of Relativity*, p. 87, Sept., 1968.

[17] N. Arkani-Hamed, L. Motl, A. Nicolis, and C. Vafa, *The string landscape, black holes and gravity as the weakest force*, *JHEP* **06** (2007) 060, [hep-th/0601001].

[18] N. Engelhardt and G. T. Horowitz, *Holographic argument for the Penrose inequality in AdS spacetimes*, *Phys. Rev. D* **99** (2019), no. 12 126009, [arXiv:1903.00555].

[19] K. S. Thorne, *Nonspherical gravitational collapse: a short review*, pp. 231–258. W.H. Freeman & Co., 1972.

[20] S. L. Shapiro and S. A. Teukolsky, *Formation of naked singularities: The violation of cosmic censorship*, *Phys. Rev. Lett.* **66** (1991) 994–997.

[21] M. W. Choptuik and F. Pretorius, *Ultra Relativistic Particle Collisions*, *Phys. Rev. Lett.* **104** (2010) 111101, [arXiv:0908.1780].

[22] P. M. Chesler and W. van der Schee, *Early thermalization, hydrodynamics and energy loss in AdS/CFT*, *Int. J. Mod. Phys. E* **24** (2015), no. 10 1530011, [arXiv:1501.04952].

[23] A. Almheiri, D. Marolf, J. Polchinski, and J. Sully, *Black Holes: Complementarity or Firewalls?*, arXiv:1207.3123.

[24] A. Almheiri, D. Marolf, J. Polchinski, D. Stanford, and J. Sully, *An Apologia for Firewalls*, arXiv:1304.6483.

[25] R. Emparan, *Predictivity lost, predictivity regained: a Miltonian cosmic censorship conjecture*, *Int. J. Mod. Phys. D* **29** (2020), no. 14 2043021, [arXiv:2005.07389].

[26] A. C. Wall, *Maximin Surfaces, and the Strong Subadditivity of the Covariant Holographic Entanglement Entropy*, *Class.Quant.Grav.* **31** (2014), no. 22 225007, [arXiv:1211.3494].

[27] N. Engelhardt and A. C. Wall, *No Simple Dual to the Causal Holographic Information?*, *JHEP* **04** (2017) 134, [arXiv:1702.01748].

[28] N. Engelhardt and A. C. Wall, *Coarse Graining Holographic Black Holes*, *JHEP* **05** (2019) 160, [arXiv:1806.01281].

[29] N. Engelhardt, G. Penington, and A. Shahbazi-Moghaddam, *Finding pythons in unexpected places*, *Class. Quant. Grav.* **39** (2022), no. 9 094002, [arXiv:2105.09316].

[30] N. Engelhardt, G. Penington, and A. Shahbazi-Moghaddam, *A World without Pythons would be so Simple*, arXiv:2102.07774.

[31] C. Akers, N. Engelhardt, D. Harlow, G. Penington, and S. Vardhan, *The black hole interior from non-isometric codes and complexity*, arXiv:2207.06536.

[32] L. Yang and N. Engelhardt, *The complexity of learning (pseudo)random dynamics of black holes and other chaotic systems*, arXiv:2302.11013.

[33] Z. Ji, Y.-K. Liu, and F. Song, *Pseudorandom quantum states*, in *Advances in Cryptology – CRYPTO 2018* (H. Shacham and A. Boldyreva, eds.), (Cham), pp. 126–152, Springer International Publishing, 2018.

[34] P. Hayden and J. Preskill, *Black holes as mirrors: quantum information in random subsystems*, *JHEP* **09** (2007) 120, [arXiv:0708.4025].

[35] I. H. Kim and J. Preskill, *Complementarity and the unitarity of the black hole S-matrix*, *JHEP* **02** (2023) 233, [arXiv:2212.00194].

[36] S. H. Shenker and D. Stanford, *Black holes and the butterfly effect*, *JHEP* **03** (2014) 067, [arXiv:1306.0622].

[37] D. A. Roberts, D. Stanford, and L. Susskind, *Localized shocks*, *JHEP* **03** (2015) 051, [arXiv:1409.8180].

[38] J. Maldacena, S. H. Shenker, and D. Stanford, *A bound on chaos*, *JHEP* **08** (2016) 106, [arXiv:1503.01409].

[39] J. S. Cotler, G. Gur-Ari, M. Hanada, J. Polchinski, P. Saad, S. H. Shenker, D. Stanford, A. Streicher, and M. Tezuka, *Black Holes and Random Matrices*, *JHEP* **05** (2017) 118, [arXiv:1611.04650]. [Erratum: JHEP 09, 002 (2018)].

[40] N. Engelhardt and Å. Folkestad, *Holography Abhors Visible Trapped Surfaces*, arXiv:2012.11445.

[41] A. R. Brown, H. Gharibyan, G. Penington, and L. Susskind, *The Python's Lunch: geometric obstructions to decoding Hawking radiation*, *JHEP* **08** (2020) 121, [arXiv:1912.00228].

[42] N. Engelhardt and A. C. Wall, *Quantum Extremal Surfaces: Holographic Entanglement Entropy beyond the Classical Regime*, *JHEP* **01** (2015) 073, [arXiv:1408.3203].

[43] R. Bousso, Z. Fisher, S. Leichenauer, and A. C. Wall, *Quantum focusing conjecture*, *Phys. Rev.* **D93** (2016), no. 6 064044, [arXiv:1506.02669].

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

[45] R. Bousso and G. Penington, *Islands Far Outside the Horizon*, arXiv:2312.03078.

[46] V. A. Belinsky, I. M. Khalatnikov, and E. M. Lifshitz, *Oscillatory approach to a singular point in the relativistic cosmology*, *Adv. Phys.* **19** (1970) 525–573.

[47] V. A. Belinsky and I. M. Khalatnikov, *General Solution of the Gravitational Equations with a Physical Singularity*, *Sov.Phys.JETP* **30** (1970) 1174.

[48] E. Shaghoulian and H. Wang, *Timelike BKL singularities and chaos in AdS/CFT*, *Class. Quant. Grav.* **33** (2016), no. 12 125020, [arXiv:1601.02599].

[49] R. Bousso, Z. Fisher, S. Leichenauer, and A. C. Wall, *Quantum focusing conjecture*, *Phys. Rev. D* **93** (2016), no. 6 064044, [arXiv:1506.02669].

[50] A. C. Wall, *A Proof of the generalized second law for rapidly-evolving Rindler horizons*, *Phys.Rev.* **D82** (2010) 124019, [arXiv:1007.1493].

[51] A. Shahbazi-Moghaddam, *Restricted Quantum Focusing*, arXiv:2212.03881.

[52] N. Engelhardt, G. Penington, and A. Shahbazi-Moghaddam, *Twice Upon a Time: Timelike-Separated Quantum Extremal Surfaces*, arXiv:2308.16226.

[53] C. Akers, N. Engelhardt, G. Penington, and M. Usatyuk, *Quantum Maximin Surfaces*, *JHEP* **08** (2020) 140, [arXiv:1912.02799].

[54] N. Engelhardt and A. C. Wall, *Decoding the Apparent Horizon: Coarse-Grained Holographic Entropy*, *Phys. Rev. Lett.* **121** (2018), no. 21 211301, [arXiv:1706.02038].

[55] A. Hamilton, D. N. Kabat, G. Lifschytz, and D. A. Lowe, *Local bulk operators in AdS/CFT: A Boundary view of horizons and locality*, *Phys.Rev.* **D73** (2006) 086003, [hep-th/0506118].

[56] A. Hamilton, D. N. Kabat, G. Lifschytz, and D. A. Lowe, *Holographic representation of local bulk operators*, *Phys.Rev.* **D74** (2006) 066009, [hep-th/0606141].

[57] A. Hamilton, D. N. Kabat, G. Lifschytz, and D. A. Lowe, *Local bulk operators in AdS/CFT: A Holographic description of the black hole interior*, *Phys. Rev.* **D75** (2007) 106001, [hep-th/0612053]. [Erratum: Phys. Rev.D75,129902(2007)].

[58] N. Engelhardt and H. Liu, *Algebraic ER=EPR and Complexity Transfer*, arXiv:2311.04281.

[59] D. Harlow and P. Hayden, *Quantum Computation vs. Firewalls*, *JHEP* **1306** (2013) 085, [arXiv:1301.4504].

[60] I. Kim, E. Tang, and J. Preskill, *The ghost in the radiation: Robust encodings of the black hole interior*, *JHEP* **06** (2020) 031, [arXiv:2003.05451].

[61] H.-Y. Huang, S. Chen, and J. Preskill, *Learning to Predict Arbitrary Quantum Processes*, *PRX Quantum* **4** (2023), no. 4 040337, [arXiv:2210.14894].

[62] S. Leutheusser and H. Liu, *Emergent times in holographic duality*, arXiv:2112.12156.

[63] T. Faulkner and M. Li, *Asymptotically isometric codes for holography*, arXiv:2211.12439.

[64] D. Marolf, *Microcanonical Path Integrals and the Holography of small Black Hole Interiors*, *JHEP* **09** (2018) 114, [arXiv:1808.00394].

[65] D. A. Roberts and B. Yoshida, *Chaos and complexity by design*, *JHEP* **04** (2017) 121, [arXiv:1610.04903].

[66] R. Gregory and R. Laflamme, *Black strings and p-branes are unstable*, *Phys. Rev. Lett.* **70** (1993) 2837–2840, [hep-th/9301052].

[67] R. P. Geroch, *Local characterization of singularities in general relativity*, *J. Math. Phys.* **9** (1968) 450–465.

[68] Y. B. Zeldovich, *A New Type of Radioactive Decay: Gravitational Annihilation of Baryons*, *Phys. Lett. A* **59** (1976) 254.

[69] L. F. Abbott and M. B. Wise, *Wormholes and Global Symmetries*, *Nucl. Phys. B* **325** (1989) 687–704.

[70] R. Kallosh, A. D. Linde, D. A. Linde, and L. Susskind, *Gravity and global symmetries*, *Phys. Rev. D* **52** (1995) 912–935, [hep-th/9502069].

[71] M. W. Choptuik, *Universality and scaling in gravitational collapse of a massless scalar field*, *Phys. Rev. Lett.* **70** (1993) 9–12.

[72] D. Christodoulou, *Examples of naked singularity formation in the gravitational collapse of a scalar field*, *Annals Math.* **140** (1994) 607–653.

[73] J. Santos, *Connecting the weak gravity conjecture to the weak cosmic censorship*, in *Strings 2018*, (Okinawa, Japan), June, 2018.

[74] R. P. Geroch and G. T. Horowitz, *Global structure of spacetimes*, in *General Relativity: An Einstein Centenary Survey*, pp. 212–293. Cambridge University Press, 1979.

[75] N. Engelhardt, Å. Folkestad, A. Levine, E. Verheijden, and L. Yang, *Cryptographic Censorship*, arXiv:2402.03425.

[76] E. Giorgi, S. Klainerman, and J. Szeftel, *Wave equations estimates and the nonlinear stability of slowly rotating kerr black holes*, 2022.

[77] F. Haake, *Quantum signatures of chaos*. Springer, 1991.

[78] D. N. Page, *Expected entropy of a subsystem*, *Phys. Rev. Lett.* **71** (1993) 1291–1294, [http://arXiv.org/abs/gr-qc/9305007].

[79] J. Choi et al., *Preparing random states and benchmarking with many-body quantum chaos*, *Nature* **613** (2023), no. 7944 468–473, [arXiv:2103.03535].

[80] W. W. Ho and S. Choi, *Exact emergent quantum state designs from quantum chaotic dynamics*, *Phys. Rev. Lett.* **128** (Feb, 2022) 060601.

[81] R. Bousso, V. Chandrasekaran, and A. Shahbazi-Moghaddam, *From black hole entropy to energy-minimizing states in QFT*, *Phys. Rev. D* **101** (2020), no. 4 046001, [arXiv:1906.05299].

[82] N. Engelhardt and S. Fischetti, *Losing the IR: a Holographic Framework for Area Theorems*, *Class. Quant. Grav.* **36** (2019), no. 3 035008, [arXiv:1805.08891].

[83] I. Bena, S. Giusto, E. J. Martinec, R. Russo, M. Shigemori, D. Turton, and N. P. Warner, *Smooth horizonless geometries deep inside the black-hole regime*, *Phys. Rev. Lett.* **117** (2016), no. 20 201601, [arXiv:1607.03908].

[84] J. Cotler and A. Strominger, *The Universe as a Quantum Encoder*, arXiv:2201.11658.

[85] J. Cotler and K. Jensen, *Isometric Evolution in de Sitter Quantum Gravity*, *Phys. Rev. Lett.* **131** (2023), no. 21 211601, [arXiv:2302.06603].

[86] P. Ananth, A. Gulati, F. Kaleoglu, and Y.-T. Lin, *Pseudorandom isometries*, 2023.

[87] S. Lloyd, *Ultimate physical limits to computation*, *Nature* **406** (2000), no. 6799 1047–1054.

[88] G. Katoch, J. Ren, and S. R. Roy, *Quantum complexity and bulk timelike singularities*, *JHEP* **12** (2023) 085, [arXiv:2303.02752].

[89] E. Meckes and M. Meckes, *Spectral measures of powers of random matrices*, *Electronic Communications in Probability* **18** (jan, 2013).

[90] Q. Wang, Z. Zhang, K. Chen, J. Guan, W. Fang, J. Liu, and M. Ying, *Quantum Algorithm for Fidelity Estimation*, *IEEE Trans. Info. Theor.* **69** (2023), no. 1 273–282, [arXiv:2103.09076].

[91] C. A. Fuchs and J. van de Graaf, *Cryptographic distinguishability measures for quantum-mechanical states*, *IEEE Trans. Info. Theor.* **45** (1999), no. 4 1216–1227, [quant-ph/9712042].

Part II

Emergent Spacetimes and their Extremal Submanifolds

Chapter 5

Conformal Rigidity from Focusing

Abstract: The null curvature condition (NCC) is the requirement that the Ricci curvature of a Lorentzian manifold be nonnegative along null directions, which ensures the focusing of null geodesic congruences. In this note, we show that the NCC together with the causal structure significantly constrain the metric. In particular, we prove that any conformal rescaling of a vacuum spacetime introduces either geodesic incompleteness or negative null curvature, provided the conformal factor is non-constant on at least one complete null geodesic. In the context of bulk reconstruction in AdS/CFT, our results combined with the technique of light-cone cuts can be used in vacuum spacetimes to reconstruct the full metric in regions probed by complete null geodesics reaching the boundary. For non-vacuum spacetimes, our results constrain the conformal factor, giving an approximate reconstruction of the metric.

5.1 Introduction

Energy conditions are indispensable tools in the study of gravity. They enable the proof of important results which are true in theories describing nature, but not in Lorentzian geometry in general. Among the most important is the null energy condition (NEC), which underpins the famous area theorem by Hawking [1], the Penrose singularity theorem [2], in addition to results related to holography [3–6], wormholes [7], cosmic censorship [8, 9], topological censorship [10], and causality [11–14].[1] The great utility of the NEC comes from that fact that it ensures the focusing of null geodesics, a physically motivated condition capturing the attractive nature of gravity.

In the Hawking area theorem, an assumption on the causal structure of spacetime together with the NEC is sufficient to derive how areas of certain surfaces behave – that is, properties of the underlying geometry. While the causal structure is invariant under Weyl transformations, areas are not. Thus, when causality is combined with the focusing of null geodesic (as implied by the NEC), we gain control over the geometry. Other remarkable examples of the interplay between causal structure and focusing in constraining geometry were found in [22–25]. In these works, area comparison theorems for extremal surfaces in spacetimes with Anti-de Sitter (AdS) asymptotics were derived from only causal relations and the null curvature condition (NCC), a property equivalent to the NEC when assuming the Einstein equation. Together, these results hint at a pattern where causal structure and energy conditions together constrain geometry to a substantial degree and in physically relevant ways.

In this note we further elaborate on this pattern, and exhibit situations where the NCC together with the conformal structure is enough fix the geometry in various ways, sometimes even uniquely. More specifically, as one of our results we will show that if a conformal class contains a geodesically complete spacetime with vanishing null curvature, then every other geometry in its conformal class either violates the NCC, or suffers from geodesic incompleteness along every geodesic along which the relative conformal factor is non-constant. In other words, in these situations, causal structure, the NCC and geodesic completeness are enough to uniquely fix the metric.

We begin in Section 5.2 by setting up the notation and introducing basic ideas which will be used throughout. Sections 5.3 is devoted to presenting our main results in a self-contained manner, with their proofs relegated to Section 5.4 for ease of reading. Finally, in Section 5.5 we discuss our findings and point out applications of our results to AdS/CFT, particularly for bulk reconstruction.

5.2 Weyl Transformations and Null Curvature

Throughout this note, a spacetime is defined as a pair (M, g) consisting of a C^2 Lorentzian manifold M and a C^2 metric tensor $g : \Gamma(TM) \times \Gamma(TM) \to \mathbb{R}$, where $\Gamma(TM)$ denotes vector fields of the tangent bundle TM of M. The spacetime dimension is denoted by $D \equiv \dim M$ and we assume $D \geq 3$ throughout. The causal structure of a spacetime (M, g) is invariant under *Weyl transformations*, which are rescalings of the metric tensor $g \to \Omega^2 g$ by some

[1] See also e.g. [15–21] for generalizations of the NEC to weaker conditions more relevant to quantum fields.

positive function Ω on M. Hence, the specification of the causal structure of a spacetime is equivalent to specifying the conformal class of its metric, which leaves an overall factor in the metric undetermined.

One may deem a Weyl transformation pathological if its *Weyl factor* Ω is not C^2 or exhibits an undesirable asymptotic behavior. In particular, denoting the conformal boundary of M by ∂M, one could in principle have $\Omega(p) \to 0$ or $\Omega(p) \to \infty$ as p is taken towards ∂M along some sequence of points. In such cases, the extendibility and completeness properties of the spacetime or the topology of its conformal boundary may be altered. It is thus useful to distinguish these more "violent" Weyl transformations from those which do not alter such properties. To do so, it will be helpful to parameterize Weyl transformations using $\Omega = e^{\omega}$, with $\omega : M \to \mathbb{R}$ sometimes referred to as the *Weyl exponent*. In particular, it is easy to see that Ω will not alter completeness properties of the underlying spacetime if ω is bounded (see Lemma 5.4 for a more precise statement).

Consider now a null vector field $k \in \Gamma(TM)$, $k^2 = 0$, and its associated null curvature

$$R_{kk} = \mathrm{Ric}(k, k), \tag{5.1}$$

where Ric is the Ricci curvature of (M, g), also written as R_{ab} in abstract index notation. When $R_{kk} \geq 0$ for every null vector field k we say that the NCC holds for (M, g) and, if $R_{kk} = 0$ for every such k, then we say that (M, g) has zero null curvature. Importantly, the condition of zero null curvature does not trivialize the Ricci tensor of the spacetime – indeed, e.g. any Einstein manifold where $\mathrm{Ric} = \kappa\, g$ for constant κ will satisfy this property.

The NCC is a well-defined geometric condition on Lorentzian manifolds independent of a theory of gravity (be it classical or semiclassical) or any physics at all, even if it is only physically motivated in the classical limit. The most standard forms of field theory matter coupled to Einstein gravity, such as minimally coupled scalars, gauge fields, and p-forms (all with canonical kinetic terms), all source classical spacetimes respecting the NCC. The same is true for the bosonic sector of type IIA, IIB and $D = 11$ supergravity,[2] and in classical bosonic string theory the NCC can be derived to leading order in α' [27]. In the context of AdS/CFT, it is furthermore known that classical matter violating the NCC can lead to various pathologies in the holographic dictionary [5], and so assuming the NCC in the classical limit is well motivated there. The assumption of the NCC is equivalent to the condition that gravity always focuses null geodesic congruences.

Knowledge of the conformal class of a spacetime is clearly not enough to fix its metric, but what if we also demand the NCC? Given a spacetime that obeys the NCC, one conformally equivalent to it need not do so. Under a Weyl transformation with Weyl exponent ω, the Ricci tensor becomes

$$R^{(\omega)}_{ab} = R_{ab} - (D-2)\left(\nabla_a\nabla_b\omega - \nabla_a\omega\nabla_b\omega\right) - \left(\nabla^2\omega + (D-2)(\mathrm{d}\omega)^2\right) g_{ab}. \tag{5.2}$$

[2]This is immediate for $D = 11$ supergravity, whose bosonic sector only contains a free 3-form (besides Chern-Simons terms of no relevance here). For type IIA and IIB supergravity, the bosonic matter stress tensors are essentially those of free bosons and p-forms with canonical kinetic parts, up to overall dilatonic prefactors having no effect on the NEC [26]. Various Kaluza-Klein dimensional reductions can also be checked to preserve the NEC [27] – e.g. for $\mathcal{N} = 2$ supergravity in $D = 4$ one just gets Einstein-Maxwell. We thank Gary Horowitz for helpful discussion on this.

The transformed null curvature then reads

$$R^{(\omega)}_{kk} = R_{kk} - (D-2)k^a k^b \left(\nabla_a \nabla_b \omega - \nabla_a \omega \nabla_b \omega\right). \tag{5.3}$$

It follows that ω defines an NCC-preserving Weyl transformation on (M, g) if and only if, everywhere and for any k,

$$R^{(\omega)}_{kk} \geq 0 \quad \Longleftrightarrow \quad (D-2)k^a k^b \left(\nabla_a \nabla_b \omega - \nabla_a \omega \nabla_b \omega\right) \leq R_{kk}. \tag{5.4}$$

Consider now a null geodesic curve $\gamma : \mathbb{R} \supseteq I \to M$ that is affinely parameterized by $\lambda \in I$. For its tangent vector field $k^a \equiv (\partial_\lambda)^a$, which obeys the geodesic equation $\nabla_k k = 0$, the condition on the right-hand side of (5.3) becomes simply

$$\nabla_k^2 \omega - (\nabla_k \omega)^2 = \frac{\mathrm{d}^2 \omega_\gamma}{\mathrm{d}\lambda^2} - \left(\frac{\mathrm{d}\omega_\gamma}{\mathrm{d}\lambda}\right)^2 \leq \frac{1}{D-2} R_{kk}, \tag{5.5}$$

where $\omega_\gamma \equiv \omega \circ \gamma$ is just the Weyl exponent along the pertinent geodesic. This inequality is the starting point for our work.

5.3 Conformal Rigidity Results

In this section we derive the main inequalities from which our theorems follow. We also state the theorems, with proofs postponed to the next section. In what follows, let $k^a \equiv (\partial_\lambda)^a$ be tangent to an affine null geodesic γ with affine parameter $\lambda \in I \subseteq \mathbb{R}$. When referring to any scalar function on M, we will implicitly work with its composition with the curve γ, as was done above e.g. in (5.5).[3]

Substituting $\omega = -\ln u$ for a strictly positive function u into (5.3), we find that to satisfy the NCC in the Weyl-rescaled spacetime we need

$$R^{(\omega)}_{kk}(\lambda) = R_{kk}(\lambda) + (D-2)\frac{u''(\lambda)}{u(\lambda)} \geq 0. \tag{5.6}$$

Multiplying through by u and integrating once and twice yields the following inequalities:

$$u'(\lambda) \gtrless u'(\lambda_0) - \frac{1}{D-2}\int_{\lambda_0}^{\lambda} \mathrm{d}\lambda'\, R_{kk}(\lambda')u(\lambda'), \qquad \forall \lambda \gtrless \lambda_0, \tag{5.7}$$

$$u(\lambda) \geq u(\lambda_0) + u'(\lambda_0)(\lambda - \lambda_0) - \frac{1}{D-2}\int_{\lambda_0}^{\lambda} \mathrm{d}\lambda'\, (\lambda - \lambda') R_{kk}(\lambda')u(\lambda'), \qquad \forall \lambda \in I, \tag{5.8}$$

for any $\lambda_0 \in I$. As a special case of these inequalities, consider a situation where $I = (-\infty, c)$, and assume that the following limits exist

$$\lim_{\lambda \to -\infty} u(\lambda) \neq 0, \qquad \lim_{\lambda \to -\infty} u'(\lambda). \tag{5.9}$$

[3] In other words, to avoid cluttering the notation, by e.g. ω we will often mean ω_γ – any exception to this should be clear from context.

Then we must have $\lim_{\lambda\to-\infty}(\lambda u'(\lambda)) = 0$, for otherwise we would need $u'(\lambda) \sim \mathcal{O}(1/\lambda)$ at large λ, giving that u diverges logarithmically. Observing that $u = 1/\Omega$ and taking $\lambda_0 = -\infty$, we then get

$$-\frac{\mathrm{d}}{\mathrm{d}\lambda}\left(\frac{1}{\Omega(\lambda)}\right) \leq \frac{1}{D-2}\int_{-\infty}^{\lambda} \mathrm{d}\lambda' \, \frac{R_{kk}(\lambda')}{\Omega(\lambda')}, \tag{5.10}$$

$$\frac{1}{\Omega(\lambda)} \geq \frac{1}{\Omega(-\infty)} - \frac{1}{D-2}\int_{-\infty}^{\lambda} \mathrm{d}\lambda' \, (\lambda-\lambda')\frac{R_{kk}(\lambda')}{\Omega(\lambda')}. \tag{5.11}$$

Thus, in the situation where Ω approaches a non-zero constant $\Omega(-\infty)$ as $\lambda \to -\infty$ so that the limit $\lim_{\lambda\to-\infty}\Omega'(\lambda)$ exists, we encounter two possibilities:

- If the original spacetime has zero null curvature $R_{kk} = 0$ and we want to preserve the NCC, then we see from (5.11) that Ω must be monotonically non-increasing along γ as we move away from the direction in which γ is complete.

- If the original spacetime satisfies the NCC with nonvanishing null curvature $R_{kk} \geq 0$, then the potential increase of $\Omega(\lambda)$ along a complete null geodesic γ is bounded from above by the previous null curvature the geodesic has encountered, as dictated by (5.11). In particular, if Ω satisfies $\Omega(\lambda) \geq \Omega_{\min} > 0$ along some complete null geodesic γ, then

$$-\frac{\mathrm{d}}{\mathrm{d}\lambda}\left(\frac{1}{\Omega(\lambda)}\right) \leq \frac{\mathcal{R}_\gamma}{(D-2)\Omega_{\min}}, \qquad \mathcal{R}_\gamma \equiv \int_{-\infty}^{\infty}\mathrm{d}\lambda' R_{kk}(\lambda'), \tag{5.12}$$

  where we have defined the *averaged null curvature* $\mathcal{R}_\gamma$ along the null geodesic γ.[4] Integrating now (5.12) from λ_0 to λ, we get

$$\frac{1}{\Omega(\lambda)} \gtrless \frac{1}{\Omega(\lambda_0)} - \frac{\mathcal{R}_\gamma}{(D-2)\Omega_{\min}}(\lambda-\lambda_0), \qquad \lambda \gtrless \lambda_0. \tag{5.13}$$

When $R_{kk} = 0$ and we do not assume anything about the asymptotic behavior of Ω, we find the following more general statement, from which most of our results will follow: Let (M, g) be a spacetime with zero null curvature. Let $\Omega \in C^2(M)$ be a positive function, and let γ be any complete null geodesic affinely parameterized by $\lambda \in \mathbb{R}$. Then at least one of the following is true:

1. The spacetime $(M, \Omega^2 g)$ violates the NCC along γ.

2. The Weyl factor Ω is constant along γ.

3. $\lim\limits_{\lambda\to\infty}\Omega(\lambda) = 0$ or $\lim\limits_{\lambda\to-\infty}\Omega(\lambda) = 0$, with Ω decaying at least as fast as $\mathcal{O}(1/|\lambda|)$.

With this lemma in hand, together with another technical lemma relegated to the next section, we find the following series of results. Let (M, g) be a connected and geodesically

[4]This object and the condition that it be nonnegative have appeared previously in e.g. [28, 29]. In general relativity, it is analogous to the averaged null energy, arising in discussions of the averaged NEC [15–17].

complete spacetime with zero null curvature. If $\omega \in C^2(M)$ is bounded, then either ω is constant on M or $(M, e^{2\omega}g)$ violates the NCC.

Let (M, g) be a spacetime with zero null curvature. If $\omega \in C^2(M)$ is bounded and non-constant, then either $(M, e^{2\omega}g)$ is null-geodesically incomplete, or it violates the NCC. If (M, g) is null-geodesically complete, then the previous is true also when ω is only bounded from below. The first theorem states that a conformal rescaling of a geodesically complete vacuum spacetime with bounded ω always causes NCC violation. The second states that in a general vacuum spacetime, after a bounded conformal rescaling, negative null curvature or geodesic incompleteness must always be present. In the next theorem, we show that an NCC-preserving conformal rescaling with bounded ω is only ever possible in incomplete spacetimes:

Let (M, g) be a spacetime with zero null curvature. Let $\omega \in C^2(M)$ be bounded and non-constant. If $(M, e^{2\omega}g)$ preserves the NCC, then (M, g) is null-geodesically incomplete.

Next we turn to some applications for spacetimes with commonly used asymptotics. Let (M, g) be an asymptotically locally AdS (AlAdS) spacetime with zero null curvature. Let $\Omega \in C^2(M)$ be a positive function and $A \subseteq M$ the set of points $p \in M$ lying on at least one complete null geodesic with both endpoints on ∂M. Assume that Ω extends to ∂M with the same value everywhere. Then at least one of the following is true:

1. The spacetime $(M, \Omega^2 g)$ violates the NCC on A.
2. The Weyl factor Ω is constant on A.
3. $(M, \Omega^2 g)$ is not AlAdS.

Let (M, g) be an asymptotically flat spacetime with zero null curvature. Let $\Omega \in C^2(M)$ be a positive function and A the set of points $p \in M$ lying on at least one complete null geodesic with both endpoints lying on $\mathscr{I}^+ \cup \mathscr{I}^-$. Assume that Ω extends to $\mathscr{I}^+ \cup \mathscr{I}^-$ with the same value everywhere. Then at least one of the following is true:

1. The spacetime $(M, \Omega^2 g)$ violates the NCC on A.
2. The Weyl factor Ω is constant on A.
3. $(M, \Omega^2 g)$ is not asymptotically flat.

These two analogous results prove strong versions of conformal rigidity for spacetimes asymptoting to AdS or flat metrics when subject to the NCC. In particular, they show that with prescribed asymptotics, regions probed by boundary-anchored null geodesics do not admit any nontrivial NCC-preserving Weyl transformation where Ω asymptotes to a constant.

Finally, as an example of how causal structure and the NCC can be used to constrain spacetime also in regions without boundary-anchored null geodesics, we now investigate the BTZ black hole [30], where all null geodesics except the closed photon orbits are incomplete [31]. We find the following: Let (M, g) be the BTZ black hole spacetime and r the area radius. Then $(M, e^{2\omega}g)$ violates the NCC unless ω preserves spherical symmetry outside the horizon. Furthermore, if the limits $\lim_{r\to\infty} \omega(r)$ and $\lim_{r\to\infty} \omega'(r)$ exist, then outside the horizon

$$\omega'(r) \geq 0. \tag{5.14}$$

5.4 Proofs

Let (M, g) be a spacetime with zero null curvature. Let $\Omega \in C^2(M)$ be a positive function, and let γ be any complete null geodesic affinely parameterized by $\lambda \in \mathbb{R}$. Then at least one of the following is true:

1. The spacetime $(M, \Omega^2 g)$ violates the NCC along γ.
2. The Weyl factor Ω is constant along γ.
3. $\lim_{\lambda\to\infty} \Omega(\lambda) = 0$ or $\lim_{\lambda\to-\infty} \Omega(\lambda) = 0$, with Ω decaying at least as fast as $\mathcal{O}(1/|\lambda|)$.

Proof. Assume $(M, \Omega^2 g)$ satisfies the NCC and Ω is not everywhere constant along γ. Then there exists an affine parameter λ_0 such that $u'(\lambda_0) \neq 0$. From (5.8), one has

$$u(\lambda) \geq u'(\lambda_0)(\lambda - \lambda_0) + u_0. \tag{5.15}$$

Depending on the sign of $u'(\lambda_0)$, we get that u diverges at least as fast as $|\lambda|$ as either $\lambda \to \infty$ or $\lambda \to -\infty$. Since $\Omega = 1/u$, the result follows.

□

Let γ be an inextendible null geodesic on a spacetime (M, g), and $\Omega \in C^2(M)$ a positive function.

- If $1/\Omega_\gamma$ is bounded above and γ is complete in (M, g), then γ is complete in $(M, \Omega^2 g)$.
- If Ω_γ is bounded above and γ is complete in $(M, \Omega^2 g)$, then γ is complete in (M, g).

Proof. Note first that γ must be inextendible as a geodesic in $(M, \Omega^2 g)$, since $(M, \Omega^2 g)$ and (M, g) are identical at the manifold level, so an endpoint in (M, g) is an endpoint in $(M, \Omega^2 g)$. The affine parameters λ and $\tilde{\lambda}$ of γ respectively in (M, g) and $(M, \Omega^2 g)$ are related by

$$\frac{\mathrm{d}\tilde{\lambda}}{\mathrm{d}\lambda} = c\Omega_\gamma^2, \tag{5.16}$$

where one can fix $c = 1$ and $\lambda(\tilde{\lambda} = 0) = 0$ without loss of generality. Assume $1/\Omega_\gamma$ is bounded above so there exists an $\epsilon > 0$ such that $\Omega_\gamma \geq \epsilon$. If γ is complete in (M, g), then

$$\lim_{\lambda\to\infty} \tilde{\lambda} = \int_0^\lambda \mathrm{d}\lambda\, \Omega_\gamma^2 \geq \epsilon^2 \int_0^\infty \mathrm{d}\lambda = \infty, \tag{5.17}$$

so γ is complete in $(M, \Omega^2 g)$. The proof for Ω_γ bounded is analogous, replacing Ω by $1/\Omega$. □

Let (M, g) be a connected and geodesically complete spacetime with zero null curvature. If $\omega \in C^2(M)$ is bounded, then either ω is constant on M or $(M, e^{2\omega} g)$ violates the NCC.

Proof. Assume $(M, e^{2\omega}g)$ satisfies the NCC. Since (M, g) is complete and ω is bounded by assumption, Lemma 5.3 implies ω is constant along every null geodesic. This will remain true along any broken null geodesic (i.e., any piecewise differentiable curve that is the concatenation of null geodesic segments). This is because all segments are extendible to complete geodesics, with concatenated intersections in pairs. Now, because the manifold is connected, there exists a path γ between any two points on M. One can approximate any such path in an arbitrarily small neighborhood of γ by a broken null geodesic also connecting the two endpoints. Hence ω is the same at any two points of M, proving the desired result. □

Let (M, g) be a spacetime with zero null curvature. If $\omega \in C^2(M)$ is bounded and non-constant, then either $(M, e^{2\omega}g)$ is null-geodesically incomplete, or it violates the NCC. If (M, g) is null-geodesically complete, then the previous is true also when ω is only bounded from below.

Proof. Assume $(M, \Omega^2 g)$ with $\Omega = e^{\omega}$ does not violate the NCC. If Ω is not constant, there must exist a null vector field k and a point p such that $\nabla_k \Omega|_p \neq 0$. Let now γ be an inextendible null geodesic passing through p with tangent k there affinely parameterized by λ. If this geodesic is incomplete in (M, g) we are done: since Ω is bounded, Lemma 5.4 would also imply incompleteness in $(M, \Omega^2 g)$. So assume γ is complete in (M, g).

By Lemma 5.3 we know that $\Omega \sim 1/|\lambda|$ as either $\lambda \to -\infty$ or $\lambda \to \infty$. Assume without loss of generality the latter case. The affine parameter $\tilde{\lambda}$ of γ in $(M, \Omega^2 g)$ is related to λ by (5.16) with the same conventions. Since there must exist constants C and $\epsilon > 0$ such that

$$\Omega(\lambda) < \frac{C}{\lambda}, \qquad \forall \lambda > \frac{1}{\epsilon}, \tag{5.18}$$

one can write

$$\lim_{\lambda \to \infty} \tilde{\lambda} = \int_0^{\infty} \mathrm{d}\lambda \, \Omega(\lambda)^2 = \int_0^{1/\epsilon} \mathrm{d}\lambda \, \Omega(\lambda)^2 + \int_{1/\epsilon}^{\infty} \mathrm{d}\lambda \, \Omega(\lambda)^2. \tag{5.19}$$

But the two terms on the right-hand side are finite, since $\Omega \in C^2(M)$ and

$$\int_{1/\epsilon}^{\infty} \mathrm{d}\lambda \, \Omega(\lambda)^2 < C^2 \int_{1/\epsilon}^{\infty} \frac{\mathrm{d}\lambda}{\lambda^2} = C^2 \epsilon. \tag{5.20}$$

Hence γ is incomplete in $(M, \Omega^2 g)$.

Finally, note that the boundedness of Ω was only used to deduce that γ was complete in (M, g). If (M, g) is geodesically complete, the assumption of Ω being bounded can be removed. □

Let (M, g) be a spacetime with zero null curvature. Let $\omega \in C^2(M)$ be bounded and non-constant. If $(M, e^{2\omega}g)$ preserves the NCC, then (M, g) is null-geodesically incomplete.

Proof. Boundedness of ω implies boundedness of Ω and $1/\Omega$. By boundedness of Ω and Theorem 5.3, $(M, \Omega^2 g)$ is null-geodesically incomplete. By boundedness of $1/\Omega$ and Lemma 5.4, this means (M, g) is geodesically incomplete. □

Let (M, g) be an asymptotically locally AdS (AlAdS) spacetime with zero null curvature. Let $\Omega \in C^2(M)$ be a positive function and $A \subseteq M$ the set of points $p \in M$ lying on at least one complete null geodesic with both endpoints on ∂M. Assume that Ω extends to ∂M with the same value everywhere. Then at least one of the following is true:

1. The spacetime $(M, \Omega^2 g)$ violates the NCC on A.
2. The Weyl factor Ω is constant on A.
3. $(M, \Omega^2 g)$ is not AlAdS.

Proof. Assume $(M, \Omega^2 g)$ obeys the NCC. By definition of AlAdS [32] there exists a function z on M such that $z^2 g$ extends smoothly to a metric on the conformal boundary ∂M, with $z|_{\partial M} = 0$, $\mathrm{d}z|_{\partial M} \neq 0$ and $z|_{\mathrm{int}\, M} > 0$. By Lemma 5.3, on each boundary-anchored null geodesic γ, if Ω is not constant on γ, then $\Omega \sim c/|\lambda|$ as either $\lambda \to \infty$ or $\lambda \to -\infty$ for some $c > 0$. Assume without loss of generality that it vanishes as $\lambda \to \infty$. In Poincaré coordinates of pure AdS, one can show that an affine null geodesic near the boundary has coordinate $z(\lambda) = \frac{\alpha}{|\lambda|}$ for some α not depending on λ. The same scaling will hold asymptotically at large λ in any AlAdS spacetime, as can be seen by using a Fefferman-Graham expansion near ∂M [33]. Hence we have the asymptotic scaling $\Omega \sim z$ as $\lambda \to \infty$ along γ. In order for $(M, \Omega^2 g)$ to be AlAdS would need a function $\hat{z}$ such that $\hat{z}^2 \hat{g} = \hat{z}^2 \Omega^2 g$ extends smoothly to a metric on ∂M. This requires the asymptotic scaling $\hat{z}\Omega \sim z$ with some nonzero coefficient of proportionality or, equivalently, that $\hat{z}(\lambda)$ asymptotes to a nonzero constant as $\lambda \to \infty$. However, this contradicts the requirement $\hat{z}|_{\partial M} = 0$ of a defining function, so $(M, \Omega^2 g)$ cannot be asymptotically AdS. It follows that if $(M, \Omega^2 g)$ is AlAdS, then Ω is constant along every boundary-anchored null geodesic γ. Since each $p \in A$ has a null geodesic ending on the boundary and Ω has a fixed value there, Ω is constant on A. □

Let (M, g) be an asymptotically flat spacetime with zero null curvature. Let $\Omega \in C^2(M)$ be a positive function and A the set of points $p \in M$ lying on at least one complete null geodesic with both endpoints lying on $\mathscr{I}^+ \cup \mathscr{I}^-$. Assume that Ω extends to $\mathscr{I}^+ \cup \mathscr{I}^-$ with the same value everywhere. Then at least one of the following is true:

1. The spacetime $(M, \Omega^2 g)$ violates the NCC on A.
2. The Weyl factor Ω is constant on A.
3. $(M, \Omega^2 g)$ is not asymptotically flat.

Proof. By definition of asymptotically flat [34], there exists a function z on M such that $z^2 g$ extends smoothly to a metric $\partial M = \mathscr{I}^+ \cup \mathscr{I}^-$, with $z|_{\partial M} = 0$, $\mathrm{d}z|_{\partial M} \neq 0$ and $z|_{\mathrm{int}\, M} > 0$. Now, any asymptotically flat space has an asymptotic coordinate system near $\mathscr{I}^\pm$ given by [34]

$$\mathrm{d}s^2 = \frac{1}{z^2}(\mathrm{d}u\, \mathrm{d}z + g_{\mathbb{S}^{D-2}}) + \mathcal{O}(z^{-1}), \tag{5.21}$$

where $g_{\mathbb{S}^{D-2}}$ is the metric of the unit $(D-2)$–sphere. To leading order in z and λ, the solution of the null geodesic equation for the z–coordinate is $z(\lambda) = c/|\lambda|$ where $c > 0$ for a boundary-anchored geodesic. Notice that z is defined only up to a rescaling $z \to f(z)z$

by a strictly positive function $f \in C^2(M \cup \partial M)$. If $(M, \Omega^2 g)$ is asymptotically flat, then there is a function $\hat{z}$ satisfying all properties above in the spacetime $(M, \Omega^2 g)$. Since z is uniquely defined up to any such f, so is $\hat{z}$, i.e., $\hat{z} = f(z)z/\Omega$. Assume now $(M, \Omega^2 g)$ obeys the NCC, and that there exists a boundary-anchored null geodesic γ on which Ω is somewhere non-constant. Then $\Omega \sim \mathcal{O}(1/|\lambda|)$ either as $\lambda \to \infty$ or $\lambda \to -\infty$. Assume without loss of generality the former. At large λ we then have $\Omega \sim \mathcal{O}(1/|\lambda|) \sim \mathcal{O}(z)$ with a nonzero coefficient. This implies $\hat{z}|_{\partial M} \neq 0$, yielding a contradiction. Thus $(M, \Omega^2 g)$ is asymptotically flat only if Ω is constant along every boundary-anchored null geodesic. Since each $p \in A$ has a null geodesic ending on the boundary and Ω has a fixed value there, Ω is constant on A. □

Let (M, g) be the BTZ black hole spacetime and r the area radius. Then $(M, e^{2\omega} g)$ violates the NCC unless ω preserves spherical symmetry outside the horizon. Furthermore, if the limits $\lim_{r\to\infty} \omega(r)$ and $\lim_{r\to\infty} \omega'(r)$ exist, then outside the horizon

$$\omega'(r) \geq 0. \tag{5.14}$$

Proof. Firstly, note the curious property that any spherically-symmetric BTZ black hole has circular photon orbits at every possible radius outside the horizon [31]. In other words, there exist complete null geodesics at every fixed radius in both exteriors along which Ω must be constant by Lemma 5.3. This means an NCC-preserving Weyl transformation on BTZ can only possibly have radial dependence outside the horizon. Finally, if $\lim_{r\to\infty} \omega(r)$ and $\lim_{r\to\infty} \omega'(r)$ exist, then (5.11) implies ω is monotonically non-decreasing along a radial null geodesic moving outwards, meaning that $\omega'(r) \geq 0$ outside the horizon. □

5.5 Discussion

Relation to previous work: In this brief note we have shown how the combination of causal structure and the null curvature condition significantly constrain the conformal factor of the metric. To our knowledge, these results are new and have not appeared elsewhere in the literature. There are nonetheless some recent results in [35] of a similar flavor. In their work, it is shown that conformal rescalings of a class of generalized Robertson-Walker spacetimes must introduce geodesic incompleteness when requiring the strong energy condition. This fits well with the pattern explored in our paper, and shows that useful results can be obtained also for non-vacuum spacetimes.

We also note that the interaction between conformal transformations, geodesic completeness, and global causality conditions (without the consideration of energy conditions) were studied in [36, 37].

Metric reconstruction in AdS/CFT with light-cone cuts: In [38–40], it was shown that knowledge of $(D+2)$-point correlators of local CFT operators on the conformal boundary ∂M of a D-dimensional AlAdS spacetime is enough to reconstruct the conformal metric in a large portion of the causal wedge $C_W = J^+(\partial M) \cap J^-(\partial M)$. This method of bulk reconstruction requires no knowledge of the bulk equations of motion, allows for obtaining the internal dimensions of the bulk spacetime which trivialize asymptotically [40], and currently constitutes the only constructive approach to partially obtaining general bulk metrics from

CFT data.[5] The downside is that light-cone cuts recover only the causal structure of the spacetime, i.e., the conformal class of its metric. But if the bulk has vanishing null curvature, our results provide the missing conformal factor, since from Theorem 5.3 only a single choice of conformal factor will lead to a NCC-respecting AlAdS geometry. Note that our method only reconstructs the conformal factor on a subset of C_W which is probed by boundary-anchored null geodesics. This constitutes a subset of the region where light-cone cuts reconstruct the conformal metric, but as long as $D \geq 4$ we expect both regions to be large subsets of C_W. For $D = 3$ this is not necessarily true due to the fact that gravitational interactions do not fall off with distance. In the case of the BTZ black hole this causes all non-spherical null geodesics to be incomplete, leading to the failure of light-cone cuts. Our methods fail at determining completely the conformal factor anywhere in C_W in the BTZ geometry for the same reason.

Future directions: Instead of analyzing complete null geodesics, one could study the effect of Weyl transformations on either complete timelike geodesics or complete spatial slices. In these cases the more useful energy condition would likely be the weak energy condition.[6] Carrying out the analysis on a spatial slice would require PDE analysis, but could on the other hand have the potential to constrain the conformal factor also in black hole interiors.

In cases where the bulk has null curvature, we would get constraints on the conformal factor through (5.8) (and the less constraining but more explicit (5.13)). It would be interesting to study more quantitatively how much (5.8) constrains geometry, and thus how good the approximate bulk reconstruction is when null curvature is present.

The geodesic NCC in (5.5) can be recast into the bound $S(f_\gamma) \leq \frac{2}{D-2} R_{kk}$, where S is the Schwarzian derivative and $f'_\gamma(\lambda) = e^{2\omega(\lambda)}$. It would be interesting to investigate if the well-studied properties of the Schwarzian derivative would allow us to derive rigidity results when $R_{kk} \neq 0$. It could also be useful to try to study the NCC directly at the level of (5.4), i.e., more globally rather than restricting to single null geodesics as we did. A possible step forward in this direction is to re-interpret (5.4) as the NEC for matter fields in a Brans-Dicke theory with ω playing the role of the dilatonic scalar [42]. This way, the NCC could be studied as the physicality condition that it be possible to add matter fields to Brans-Dicke such that solutions to the equations of motion obey the NEC.

Finally, in Theorem 5.3, we explored the implications of our results on BTZ, deriving a condition on the conformal factor for spacetimes in its conformal class to obey the NCC. This low-dimensional example illustrated the usefulness of our methods even in the absence of boundary-anchored null geodesics. Higher-dimensional spacetimes provide an arena where Theorems 5.3 and 5.3 would play an important role in large regions of spacetime, but where we also expect there to be regions in C_W which are not probed by boundary-anchored null geodesics, particularly near black holes. Hence it would be interesting to explore the implications of our methods in these regions of higher-dimensional black hole spacetimes [43, 44].

[5] The reconstruction of bulk operators is a different problem which generally assumes knowledge of a background bulk geometry in the first place – see [41] for more details. Of course, operator reconstruction includes perturbative metric reconstruction as a special case, but this still requires a reference background on top of which to consider graviton perturbations.

[6] We could also consider the dominant or strong energy conditions, but the latter is not physically well motivated and the former is stronger than the weak one.

Acknowledgments

It is a pleasure to thank Netta Engelhardt for collaboration in the early stages of this project, Gary Horowitz for helpful comments on various parts of the paper, and Gregory Galloway and Elena Giorgi for guidance on the existing mathematical literature. ÅF is supported in part by NSF grant PHY-2011905 and the MIT department of physics, and in part by an Aker Scholarship. SHC is supported by NSF grant PHY-1801805 and the University of California, Santa Barbara.

References

[1] S. W. Hawking, *Gravitational radiation from colliding black holes*, *Phys. Rev. Lett.* **26** (1971) 1344–1346.

[2] R. Penrose, *Gravitational collapse and space-time singularities*, *Phys. Rev. Lett.* **14** (1965) 57–59.

[3] S. Gao and R. M. Wald, *Theorems on gravitational time delay and related issues*, *Class. Quant. Grav.* **17** (2000) 4999–5008, [gr-qc/0007021].

[4] A. C. Wall, *Maximin Surfaces, and the Strong Subadditivity of the Covariant Holographic Entanglement Entropy*, *Class. Quant. Grav.* **31** (2014), no. 22 225007, [arXiv:1211.3494].

[5] M. Headrick, V. E. Hubeny, A. Lawrence, and M. Rangamani, *Causality & holographic entanglement entropy*, *JHEP* **12** (2014) 162, [arXiv:1408.6300].

[6] R. Bousso and N. Engelhardt, *New Area Law in General Relativity*, *Phys. Rev. Lett.* **115** (2015), no. 8 081301, [arXiv:1504.07627].

[7] M. S. Morris, K. S. Thorne, and U. Yurtsever, *Wormholes, Time Machines, and the Weak Energy Condition*, *Phys. Rev. Lett.* **61** (1988) 1446–1449.

[8] B. Chen, F.-L. Lin, B. Ning, and Y. Chen, *Constraints on Low-Energy Effective Theories from Weak Cosmic Censorship*, *Phys. Rev. Lett.* **126** (2021), no. 3 031102, [arXiv:2006.08663]. [Erratum: Phys.Rev.Lett. 126, 119903 (2021)].

[9] N. Engelhardt and Å. Folkestad, *Holography Abhors Visible Trapped Surfaces*, arXiv:2012.11445.

[10] J. L. Friedman, K. Schleich, and D. M. Witt, *Topological censorship*, *Phys. Rev. Lett.* **71** (1993) 1486–1489, [gr-qc/9305017]. [Erratum: Phys.Rev.Lett. 75, 1872 (1995)].

[11] F. J. Tipler, *Causality violation in asymptotically flat space-times*, *Phys. Rev. Lett.* **37** (1976) 879–882.

[12] S. W. Hawking, *The Chronology protection conjecture*, *Phys. Rev. D* **46** (1992) 603–611.

[13] K. D. Olum, *Superluminal travel requires negative energies*, *Phys. Rev. Lett.* **81** (1998) 3567–3570, [gr-qc/9805003].

[14] M. Visser, B. Bassett, and S. Liberati, *Superluminal censorship*, *Nucl. Phys. B Proc. Suppl.* **88** (2000) 267–270, [gr-qc/9810026].

[15] L. H. Ford and T. A. Roman, *Averaged energy conditions and quantum inequalities*, *Phys. Rev. D* **51** (1995) 4277–4286, [gr-qc/9410043].

[16] N. Graham and K. D. Olum, *Achronal averaged null energy condition*, *Phys. Rev. D* **76** (2007) 064001, [arXiv:0705.3193].

[17] T. Hartman, S. Kundu, and A. Tajdini, *Averaged Null Energy Condition from Causality*, *JHEP* **07** (2017) 066, [arXiv:1610.05308].

[18] R. Bousso, Z. Fisher, J. Koeller, S. Leichenauer, and A. C. Wall, *Proof of the Quantum Null Energy Condition*, *Phys. Rev. D* **93** (2016), no. 2 024017, [arXiv:1509.02542].

[19] S. Balakrishnan, T. Faulkner, Z. U. Khandker, and H. Wang, *A General Proof of the Quantum Null Energy Condition*, *JHEP* **09** (2019) 020, [arXiv:1706.09432].

[20] B. Freivogel and D. Krommydas, *The Smeared Null Energy Condition*, *JHEP* **12** (2018) 067, [arXiv:1807.03808].

[21] S. Leichenauer and A. Levine, *Upper and Lower Bounds on the Integrated Null Energy in Gravity*, *JHEP* **01** (2019) 133, [arXiv:1808.09970].

[22] A. May, G. Penington, and J. Sorce, *Holographic scattering requires a connected entanglement wedge*, arXiv:1912.05649.

[23] A. May, *Holographic quantum tasks with input and output regions*, arXiv:2101.08855.

[24] A. May and D. Wakeham, *Quantum tasks require islands on the brane*, arXiv:2102.01810.

[25] A. May, *Bulk private curves require large conditional mutual information*, arXiv:2105.08094.

[26] M. J. D. Hamilton, *The field and Killing spinor equations of M-theory and type IIA/IIB supergravity in coordinate-free notation*, arXiv:1607.00327.

[27] M. Parikh and J. P. van der Schaar, *Derivation of the Null Energy Condition*, *Phys. Rev. D* **91** (2015), no. 8 084002, [arXiv:1406.5163].

[28] A. Borde, *Geodesic focusing, energy conditions and singularities*, *Class. Quant. Grav.* **4** (1987) 343–356.

[29] N. Engelhardt and S. Fischetti, *The Gravity Dual of Boundary Causality*, *Class. Quant. Grav.* **33** (2016), no. 17 175004, [arXiv:1604.03944].

[30] M. Banados, C. Teitelboim, and J. Zanelli, *The Black hole in three-dimensional space-time*, *Phys. Rev. Lett.* **69** (1992) 1849–1851, [hep-th/9204099].

[31] N. Cruz, C. Martinez, and L. Pena, *Geodesic structure of the (2+1)-dimensional black hole*, *Class. Quant. Grav.* **11** (1994) 2731–2740, [gr-qc/9401025].

[32] D. Marolf, W. Kelly, and S. Fischetti, *Conserved Charges in Asymptotically (Locally) AdS Spacetimes*, pp. 381–407. Springer Berlin Heidelberg, 11, 2012. arXiv:1211.6347.

[33] C. Fefferman and C. R. Graham, *Conformal invariants*, in *Élie Cartan et les mathématiques d'aujourd'hui*, no. S131 in Astérisque, pp. 95–116. Société mathématique de France, 1985.

[34] R. M. Wald, *General Relativity*. Chicago Univ. Pr., Chicago, USA, 1984.

[35] G. J. Galloway and E. Ling, *Remarks on the existence of CMC Cauchy surfaces*, arXiv:2104.02136.

[36] C. J. S. Clarke, *On the Geodesic completeness of causal space-times*, *Proc. Cambridge Phil. Soc.* **69** (1970) 319–323.

[37] J. K. Beem, *Conformal changes and geodesic completeness*, *Commun. Math. Phys.* **49** (June, 1976) 179–186.

[38] N. Engelhardt and G. T. Horowitz, *Towards a Reconstruction of General Bulk Metrics*, *Class. Quant. Grav.* **34** (2017), no. 1 015004, [arXiv:1605.01070].

[39] N. Engelhardt and G. T. Horowitz, *Recovering the spacetime metric from a holographic dual*, *Adv. Theor. Math. Phys.* **21** (2017) 1635–1653, [arXiv:1612.00391].

[40] S. Hernández-Cuenca and G. T. Horowitz, *Bulk reconstruction of metrics with a compact space asymptotically*, *JHEP* **08** (2020) 108, [arXiv:2003.08409].

[41] A. Hamilton, D. N. Kabat, G. Lifschytz, and D. A. Lowe, *Local bulk operators in AdS/CFT: A Boundary view of horizons and locality*, *Phys. Rev. D* **73** (2006) 086003, [hep-th/0506118].

[42] C. Brans and R. H. Dicke, *Mach's principle and a relativistic theory of gravitation*, *Phys. Rev.* **124** (1961) 925–935.

[43] R. Emparan and H. S. Reall, *Black Holes in Higher Dimensions*, *Living Rev. Rel.* **11** (2008) 6, [arXiv:0801.3471].

[44] G. T. Horowitz, ed., *Black Holes in Higher Dimensions*. Cambridge Univ. Pr., 2012.

Chapter 6

Subregion Independence in Gravity

Abstract: In gravity, spacelike separated regions can be dependent on each other due to the constraint equations. In this paper, we give a natural definition of subsystem independence and gravitational dressing of perturbations in classical gravity. We find that extremal surfaces, non-perturbative lumps of matter, and generic trapped surfaces are structures that enable dressing and subregion independence. This leads to a simple intuitive picture for why extremal surfaces tend to separate independent subsystems. The underlying reason is that localized perturbations on one side of an extremal surface contribute negatively to the mass on the other side, making the gravitational constraints behave as if there exist both negative and positive charges. Our results support the consistency of islands in massless gravity, shed light on the Python's lunch, and provide hints on the nature of the split property in perturbatively quantized general relativity. We also prove a theorem bounding the area of certain surfaces in spherically symmetric asymptotically de Sitter spacetimes from above and below in terms of the horizon areas of de Sitter and Nariai. This theorem implies that it is impossible to deform a single static patch without also deforming the opposite patch, provided we assume spherical symmetry and an energy condition.

6.1 Introduction

When are spacelike separated regions in a quantum theory of gravity independent? This is a central question that holography forces us to confront. On the one hand, the constraint equations of GR tells us that some gravitational subregions depend on each other, since generically one cannot specify initial data on two regions independently. On the other hand, the theory of entanglement wedges and subregion-subregion duality [1–13] in AdS/CFT demonstrates that independent regions do exist – at least perturbatively in the $G_N \to 0$ limit. Thus, some spatially separated regions are mutually dependent, while others are not.

For perturbative canonical quantization of GR around flat space, any compact region in the bulk in some sense depends on a region surrounding spatial infinity, since every compactly supported operator must be dressed to infinity when coupled to gravity [14], due to diffeomorphism invariance. No excitation can be turned on in the center of the spacetime without disturbing the gravitational field at infinity. An even stronger notion of dependence on infinity has been shown for perturbations of pure AdS: to first order in perturbation theory, two Wheeler-deWitt (WdW) functionals that agree on observables in a time-band near the boundary must be equal [15]. It might be tempting to think that perturbatively quantized GR itself is holographic, in the sense that perturbative data near the conformal boundary can be used to access perturbative data in the deep bulk, without relying on non-perturbative corrections in G_N^{-1}.[1] But what about other backgrounds? While gravity contains no truly local operators, once the background has sufficiently rich structure, there exist quasilocal operators [22–28] that are dressed features other than conformal infinity. For such backgrounds, one can construct operators that commute with operators near the boundary to all orders in perturbation theory [26, 29, 30]. This shows that we must be cautious about extrapolating lessons about subregion dependence learned from working perturbatively around featureless spacetimes like Minkowski or AdS. At the classical level, these satisfy rigidity theorems [31–34] that strongly constrain the nature of perturbations around these spacetimes.

In this paper, we will show that the question of subregion independence is surprisingly interesting and illuminating already at the classical level. We will demonstrate that the nature of the background strongly influences how and when regions become independent. This will lead to a simple physical picture where classical extremal surfaces play an important role in ensuring subregion independence.

To approach the precise question we will ask in this paper, consider two spatial boundary subregions L, R on the conformal boundary of an asymptotically AdS (AAdS) spacetime. Assume that these are spacelike separated by a small but finite gap – see Fig. 6.1. By the so-called split property of local QFT [35–39], we then have a notion of independence for the CFT states on L, R. Namely, for any two independent choices of states ω_L and ω_R on these regions, there exists a pure state $|\psi\rangle$ on the full Hilbert space that agrees with ω_L, ω_R on L, R:[2]

$$\langle\psi|\,\mathcal{O}_L\mathcal{O}_R\,|\psi\rangle = \langle\mathcal{O}_L\rangle_{\omega_L}\,\langle\mathcal{O}_R\rangle_{\omega_R}\,, \tag{6.1}$$

[1] See [16–19] for the development of the idea that a version of holography is implied by the fact that the Hamiltonian is boundary term, and also for discussion of the importance of non-perturbative corrections. See also [20, 21] for recent work on these ideas.

[2] Strictly speaking, the split property guarantees a state on $R \cup L$ with this property. But we expect that this state can be purified on the full Hilbert space.

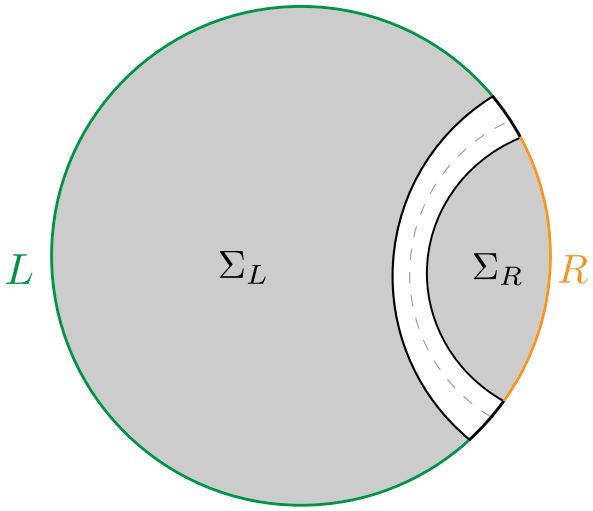


Figure 6.1: Illustration of an AdS Cauchy slice Σ containing Cauchy slices Σ_L, Σ_R for the entanglement wedges of the boundary regions L, R. Taking the classical limit of AdS/CFT, we expect that we can make perturbations of the classical initial data in Σ_L and Σ_R independently.

where $\mathcal{O}_L, \mathcal{O}_R$ are operators in the algebra of (the domain of dependence of) L, R, respectively. Assuming we are in a holographic CFT, ω_L, ω_R encodes data about the entanglement wedges of L, R. Now, let Σ_L, Σ_R be Cauchy slices of the entanglement wedges of L, R, contained in a full Cauchy slice Σ shown in Fig. 6.1. In the strictly classical limit of AdS/CFT, we expect that Σ_L, Σ_R inherit some notion of independence from the QFT split property. A natural expectation would be the following: for any two small perturbations of initial data in Σ_L, Σ_R, there exists some complete set of GR initial data on all of Σ that agrees with our perturbed data on Σ_L, Σ_R. Furthermore, this complete initial data is perturbatively close to the initial data we started with. In more pedestrian language: we can wiggle the initial data in the two subregions independently, without making drastic non-perturbatively large changes to the full spacetime.[3] However, there is no one-liner argument in GR showing that this is true, since the constraint equations are elliptic – they have no lightcones. Generically, modifications of matter (or pure gravity degrees of freedom) in some compact region A source gravitational fields spacelike to A, so how do we know that there are no perturbations of initial data in Σ_L that requires us to change fields in Σ_R? We might be tempted to mumble words about dressing observables to R and L, but this is not quite satisfactory. First, we should be able to ask this question at the level of solutions of the classical constraint equations of GR. Second, and more importantly, this leaves no room for the extremal surface playing any kind of special role – this explanation would lead us to conclude that any two disjoint regions $\Sigma_1, \Sigma_2 \subset \Sigma$ connected to the asymptotic boundary and separated by a gap are independent. It is not clear that this is true.

To illustrate the problem more clearly, let us for a brief moment consider a different example where we model gravity as electromagnetism coupled to matter with only positive charge density, living on a fixed background geometry. In this case, Gauss' law is a stand-in

[3] In AdS/CFT language, we do not have to leave the classical limit of our code subspace where the entanglement wedges were defined.

for the Hamiltonian constraint of GR, the electric field E^i for the gravitational field, and the charge density ρ for the energy density. Consider now this theory on a fixed AdS-Schwarzschild background, and let Σ_L, Σ_R be the left and right sides of the bifurcation surface on a canonical $t = 0$ slice. We let Σ_L, Σ_R terminate slightly to the left and right of the bifurcation surface, so that there is a small open region C between these slices that contains the bifurcation surface.[4] See Fig. 6.2. A perturbation of the trivial $E^i = \rho = 0$ initial data on Σ_L is the following: add a spherically symmetric shell of charged matter and choose that the electric field is unchanged near left infinity. This is compatible with Gauss' law on Σ_L. However, now Gauss' law requires the new electric field lines sourced by the shell to travel towards the right. Because our theory contains no negative charges, there is nothing we can do in C to absorb the field lines before they reach Σ_R. They will travel all the way to right infinity. See Fig. 6.2. The initial data on Σ_L and Σ_R cannot be chosen independently.

Now let us turn to gravity proper. If the Hamiltonian constraint behaved like Gauss' law coupled to matter with only positive charge densities, this would not bode well for the independence of initial data perturbations in Σ_R and Σ_L in gravity. However, except in the special case of perturbations around pure AdS or Minkowski, Gauss' law coupled to purely positive charge densities is a poor analogy for the Hamiltonian constraint. Crucially, the total gravitational mass is not just an integral of the local energy density weighted by something positive. As we will explain, extremal surfaces play a very practical role in ensuring that complementary entanglement wedges are independent (after you cut out a small splitting region). Specifically, we are going to show that if you live in Σ_R, positive energy density added to the other side of the bifurcation surface, from your perspective, acts like a negative energy density! Furthermore, this effect is present for general extremal surfaces. This means that positive energy densities can be used to terminate gravitational fields sourced by positive energy densities on the opposite sides of an extremal surface.[5] For compact extremal surfaces, this effect cannot be understood by studying the Hamiltonian constraints perturbatively around AdS or flat space, or by relying on analogies to Gauss' law. It relies on the non-linear nature of gravity, which causes structures of the background to significantly affect when subregions allow independent perturbations of initial data. We argue that this effect relieves a tension that was argued to exist between islands and massless gravity in [40] (see also [29, 30] for arguments that islands and massless gravity are consistent). It also ensures the independence of Σ_L and Σ_R in the example considered above.

Next, we will prove or argue, depending on the setup, that trapped regions of spacetime always have a kind of indeterminate energetic behavior. From the perspective of the mass in any asymptotic region, there are always some modes in a trapped region that, when turned on, increase the mass, and others that reduce it.[6] Together, these enable perturbations of spatial

[4]This case has a qualitative difference from the previous example. Σ_L and Σ_R are slightly smaller than Cauchy slices of the entanglement wedges of the left and right boundaries, R, L. This is because ω_L, ω_R now cannot be chosen completely independently. If their entanglement entropies disagree, there is no pure state $|\psi\rangle$ on the joint Hilbert space that reduces to ω_R, ω_L. So we leave a small number of degrees of freedom flexible, to avoid this obstruction.

[5]Strictly speaking, we are able to show this in all dimensions only with spherical symmetry, while in four spacetime dimensions we are able to remove this assumption.

[6]This has of course long been understood for stationary black holes, where the Killing vector that defines the asymptotic energy flips signature in the bulk, but we will make it clear more generally that trapped regions always have this behavior.

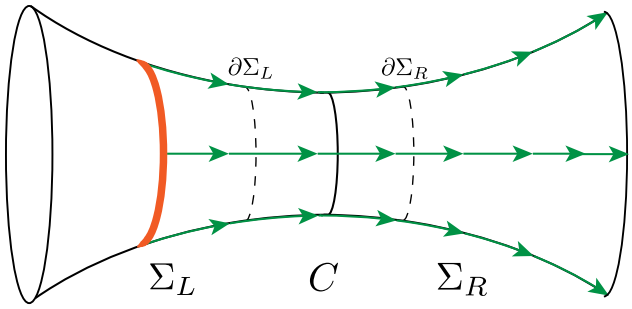


Figure 6.2: Initial data for electromagnetism coupled to matter on a fixed AdS-Schwarzschild background, on a canonical $t = 0$ slice. The regions Σ_L, Σ_R are not independent when electromagnetism is coupled to charge of a fixed sign. If we pick initial data on Σ_L corresponding to adding a charged shell of matter, and with the $E^i = 0$ at left infinity, the Gauss law imposes that $E^i \neq 0$ at right infinity.

compact support. This gives subregions of a trapped region a larger degree of independence from nearby regions, and we will use it to argue that the islands found in the evaporating black holes of [41, 42] easily can host a large number of localized excitations that never affect the black hole exterior. Along the way we will also show the following fact: an object's contribution to the ADM mass approaches zero as the object localizes on an extremal surface (see Secs. 6.4.2, 6.4.3, for precise statements). So around an extremal surface, we can add arbitrarily high energy objects (as measured locally) with no or arbitrarily small change in the ADM mass.[7]

In the following, we first give a precise definition of subregion independence in GR in Sec. 6.2 and argue why it is a useful definition. We explore our definition in a series of illuminating examples in Sec 6.3. We find that regions separated by extremal surfaces "shields" of matter tend to classically independent. Sometimes trapped surfaces can play a similar role. Of course, we do not expect that this independence generally persists (or can be defined) in full non-perturbative quantum gravity, but it does appear plausible that this independence could be inherited by a perturbative quantization of GR around a fixed background.

Then, in Sec. 6.4, we conduct a more general analysis. For spacetimes with spherical symmetry, we prove that the behaviors we found in the examples are general: gravity truly behaves as if it has negative and positive energies when extremal or trapped surfaces are present. In four-dimensional spacetime, we are also able to remove the spherical symmetry assumption, giving strong arguments at the physics level of rigor. We do expect that the physical effects we identify to survive in other dimensions, although we do not know what are the right tools to show this in spacetimes without symmetries.

Then, guided by findings when studying independence of subregions of de Sitter in Sec. 6.3, in Sec. 6.5 we prove a new theorem upper and lower bounded the area of a family of surfaces that include certain extremal and trapped surfaces in asymptotically dS spacetimes.

[7]This plays nicely with the ideas of [43], which propose that Wilson lines threading a wormhole really, in the bulk UV, are split into factorizing operators by heavy charges residing near the bifurcation surface. This ensures factorization. The fact that general classical extremal surfaces can host heavy objects without large cost in energy supports the viability of this idea in a large class of backgrounds.

A corollary of the theorem states that under spherical symmetry and given an energy condition, adding matter in one static patch necessitates adding matter in the opposite patch. Furthermore, adding matter always reduces the area of a certain extremal/trapped surface. This result holds at the full non-linear level.

Finally, in Sec. 6.6 we discuss implications of our findings for semiclassical gravity, quantum extremal surfaces, islands in massless gravity, localized observables, and the Python's lunch.

6.2 Subregion Independence

6.2.1 The constraint equations

To understand subsystem independence, we need to deal with the gravitational constraint equations. Let Σ be a manifold, possibly with boundary, and let

$$S \equiv (h_{ab}, K_{ab}, \Phi) \tag{6.2}$$

be a regular initial dataset for the Einstein equations. Here h_{ab} is a Riemannian metric on Σ, K_{ab} a symmetric tensor, and Φ a label that collectively denotes initial data for all of the matter fields. For example, when the matter is a scalar field, Φ consists of the value of the scalar and its time-derivative on Σ. For any choice of S, assuming the matter is reasonably well behaved, there is a unique (up to diffeomorphism) globally hyperbolic spacetime (M, g) corresponding to the maximal evolution of (Σ, S) [44–47]. The pair (Σ, S) represents a moment of time in (M, g), with Σ being an embedded spatial slice, and with h_{ab}, K_{ab} being the induced metric and extrinsic curvature of Σ, respectively. Note that we never consider initial data such that Σ has a kink or ends in a singularity. If we extract an initial data slice (Σ, S) from an existing spacetime (N, g), unless otherwise specified, we always assume it is globally (AdS-)hyperbolic, with Σ being a Cauchy slice.[8]

Due to diffeomorphism invariance, initial data is not freely specifiable. Instead, it has to satisfy a set of constraint equations. In the case of Einstein gravity minimally coupled to matter, the constraint equations read (in units of $8\pi G_N = 1$)

$$\mathcal{R} - K_{ab}K^{ab} + K^2 - 2\Lambda = 2\mathcal{E}, \tag{6.3}$$

$$\mathcal{D}^b K_{ba} - \mathcal{D}_a K = \mathcal{J}_a, \tag{6.4}$$

where $\mathcal{R}$ is the Ricci scalar of h_{ab}, $\mathcal{D}_a$ the h_{ab}-compatible connection on Σ, $K = K^a{}_a$, and Λ the cosmological constant. $\mathcal{E}$ and $\mathcal{J}_a$ are the matter energy and momentum densities, respectively. From the spacetime perspective, if n^a is the future unit normal to Σ, P_{ab} the projector onto the tangent space of Σ, and T_{ab} the matter stress tensor, then $\mathcal{E} = T_{ab}n^a n^b$ and $\mathcal{J}_a = T_{cb}n^b P^c{}_a$. Equations (6.3) and (6.4) are known as the Hamiltonian and momentum constraints, respectively. Unless otherwise stated, we will assume the weak energy condition (WEC), which says that the local energy density is positive:

$$\mathcal{E} \geq 0. \tag{6.5}$$

[8] For AAdS spacetimes, which are technically not globally hyperbolic, we assume the AdS notion of global hyperbolicity [5].

As will become clear soon, we are in an unusual situation where we are assuming the WEC to make our lives harder, not easier. The positivity of local energy densities is a significant obstruction to achieving independence of subregions in gravity, so if two regions are independent despite the WEC, then we expect them to be independent also when the WEC is violated.[9]

As alluded to in the introduction, the central property of the constraint equations for us is that they are not hyperbolic equations. There are no lightcones. A change in the matter fields with support on some region $A \subset \Sigma$ might demand a change in the gravitational or gauge fields in a different spacelike separated region $B \subset \Sigma$. This type of behavior is of course already present in the Maxwell equations. For example, at $t = 0$ in Minkowski spacetime, it is impossible to add an electron without also adding an electric field which either reaches infinity, or alternatively, also adding a positron that the field can terminate on. From the QED perspective, this is because there are no gauge-invariant operators that create a single electron at a point x. Instead one needs to dress the operator to restore gauge invariance, for example by adding a Wilson line that reaches out to infinity:

$$e^{i\int_x^\infty A}\psi(x), \tag{6.6}$$

where A, ψ are the gauge and Dirac fields, respectively. This adds precisely the required electric field to satisfy the constraints. However, there is no need study the quantum theory to see the need for dressing – the effect is already there in the classical Gauss' law. This suggests that a careful understanding of the constraints, even at the classical level, can give hints about what gauge invariant operators, and thus what algebras of observables, can exist. This is an additional motivation for this study. We will see the Einstein constraint equations yield behavior significantly richer than Gauss' law.

6.2.2 Subregion independence and dressing

To define subsystem independence, we need to look at perturbations to the constraints. We will use $\delta S \equiv (\delta h_{ab}, \delta K_{ab}, \delta\Phi)$ to denote a regular perturbative solution of the constraint equations around some initial dataset S (Σ is often left implicit). The small perturbation parameter is the amplitude of the perturbation. Unless otherwise specified, we assume δS is a formal series solution to all orders in perturbation theory in this amplitude, and not just a linearized solution. We always choose the lowest order terms in the perturbation so that the leading order correction to the stress tensor and the metric is at the same order. Note that some perturbations do not correspond to a physical change of the spacetime, but instead to a diffeomorphism, either changing coordinates within Σ, or moving us to a different slice nearby Σ in the same spacetime. We allow these.

Next, consider a subregion $A \subset \Sigma$. We will use the notation $\delta S|_A$ to refer to a perturbative constraint solution restricted to A, and we will refer to a perturbative constraint solution δS specified on all of Σ as an *extension* of $\delta S|_A$ if it restricts to $\delta S|_A$ on A. If an extension exists, it will generally not be unique. If we specify a perturbation of the constraint on two disjoint subregions A, B, we take $\delta S|_A \cup \delta S|_B$ to mean the obvious concatenated perturbation on $A \cup B$. Finally, since we have in mind working in some particular theory, when we talk about

[9] Modulo other complications that might arise if WEC breaking comes from considering quantum effects.

perturbations, we only ever talk about perturbations that respect the boundary conditions imposed on the boundaries of Σ by our theory – either at finite boundaries or at spatial infinity.

We are now ready to define when A and B are independent:

Definition 9 (Independence)**.** *Let (Σ, S) be an initial dataset, and let $A, B \subset \Sigma$ be two disjoint subregions. We say that A and B are independent if for any two perturbations $\delta S|_A$ and $\delta S|_B$, there exists a perturbative extension δS of $\delta S|_A \cup \delta S|_B$ to all of Σ.*

Definition 9 simply says that it is possible to wiggle the initial data in A, B completely independently, provided these wiggles are perturbatively small.[10] If A and B are not independent, we say that they are dependent. If A and B are independent, it is natural to say that classical gravity satisfies the split property across $C \equiv \Sigma - A - B$. This means that for any (perturbative) choice of the classical state on A and B, there exist a classical state on all of Σ that agrees with the chosen states on A and B. Note that it is important that we work within a fixed theory, which comes with a set of allowed matter fields and boundary conditions. We might think that we can use gluing constructions supported by delta functions shocks to make more or less anything independent, but this is not so. Gluings will generically produce distributional stress tensors that do not satisfy any energy conditions. Even if they do, there is no guarantee that regularizations of these shocks can be achieved by matter in our theory. So we might as well stick with genuinely regular initial data.

Next, note that since pure diffeomorphisms are allowed, perturbing the boundaries of the regions A, B count as perturbing the initial data in A and B. To see this, consider the data S and assume that the regions A, B are fixed in some gauge. Assume now that a diffeomorphism $\psi : \Sigma \to \Sigma$ maps the boundaries of A, B to new locations. An initial dataset in which the boundaries have moved, but with the spacetime and the slice Σ left fixed, is $(\Sigma, \psi[A], \psi[B], S)$. By diffeomorphism invariance, this is equivalent to $(\Sigma, A, B, \psi_*[S])$, where ψ_* is the pull-back.[11] This logic can easily be generalized to a diffeomorphism perturbing A, B out of the slice. Thus, the regions A and B are not completely fixed – they are fuzzy at the perturbative scale. Alternatively said, the perturbative edges modes [48] of A and B can be activated when testing for independence. Before we apply a physical perturbation in A that changes the spacetime, we are allowed to deform the boundaries of A by a perturbative amount first – or the other way around.

Definition 9 can be applied to classical non-gravitational theories as well. For theories without constraints, such as a scalar field ϕ, all regions A, B separated by an open set C are independent. We can always match the initial data $\phi, \dot{\phi}$ in A, B smoothly across C, since they are completely unconstrained there. However, the *splitting region* C was important. Even in free scalar theory, two regions with intersecting closures are dependent. Next, even in the presence of a constraint like Gauss' law of electromagnetism, on Minkowski space all

[10] Rather than perturbative deformations, we could alternatively work with finite but arbitrarily small deformations with respect a C^k or Sobolev norm defined by the background metric, as is common in mathematical studies of the constraints. See for example [46, 47] for a discussion of such norms. We expect that the physical effects we discuss in this paper are robust to taking this approach instead.

[11] Note that if ψ is a "large diffeomorphism" (not large in the perturbative sense, but in the sense that it has non-trivial action on the boundary of spacetime), then we have both changed the full physical state, in addition to moving A and B.

spacelike separated subregions are completely independent, provided we couple to charged matter with both signs, as explained in the introduction. If we add a positive charge density in A, we just need to make sure to compensate with a negative charge density in C to screen any new multipole moments incurred in B.[12]

In gravity things are more interesting. Because of the positivity of local energy densities, it is generally harder to screen perturbations, and the question of independence is non-trivial. As an example, consider Σ to be a canonical $t = 0$ slice of Minkowski or AdS. Let $A \subset \Sigma$ be a compact region and $B \subset \Sigma$ an infinite annulus containing spatial infinity. By the positive mass theorem (PMT) [32–34, 51, 52], any perturbation in A which is not a pure diffeomorphism must increase the spacetime's mass. Since the ADM Hamiltonian is a boundary term localized to B, this implies that the gravitational data in B must be perturbed as well. Thus A and B are dependent. Pure AdS and Minkowski support no perturbations of compact support, except for pure gauge.[13] From the perspective of perturbative quantum gravity around flat space and AdS, this can be seen as the classical avatar of the statement that operators must be dressed to infinity [14], for example with a gravitational Wilson line (see for example [53–56]). Effectively, there is no negative energy density to terminate the Wilson line on.

Going away from pure AdS or Minkowski, since matter has positive energy density, it naively seems like we are always forced to dress a perturbation to asymptotic infinity, in turn causing every compact subregion to depend on an annulus around spatial infinity. However, this conclusion is much too quick. When certain geometric features are present, such as extremal and trapped surfaces or lumps of matter, localized dressings become possible.

To streamline the discussion going forward, it is useful to give a precise notion of classical dressing of a perturbation (as opposed to an observable):

Definition 10 (Dressing)**.** *Let (S, Σ) be an initial dataset. Let A be a subregion and $\delta S|_A$ a perturbation of the constraints on A. If δS is an extension of $\delta S|_A$, we say that δS is a dressing of $\delta S|_A$ to the region supp(δS).*

We should emphasize that given some $\delta S|_A$, the choice of gravitational dressing is highly non-unique. The independence of A and B implies that for every perturbation of A, there exists some choice of dressing that does not intersect B. Furthermore, after dressing any perturbation on A to C, we still have the ability to dress any perturbation in B to C as well. The statement of independence is not that a certain perturbation in A could not be dressed to B, if so desired.

Next, let us address two questions that naturally arise from our definition. What about regions in spacetime that are not conveniently described as lying on the same slice? Why only perturbative deformations?

First, in globally (AdS-)hyperbolic spacetimes,[14] it is straightforward to extend our notion

[12] In vacuum electromagnetism on Minkowski, we also have independence of subregions [49]. However, on non-trivial manifolds, the question of independence is more interesting. For example, as pointed out in [50], vacuum electromagnetism on a spatial torus will not have subregion independence. For example, on T^2, this is because the value of the electric charge on two homologous non-contractible S^1's must match in different regions.

[13] It is crucial to not just work to linear order in perturbations, since compact perturbations of Minkowski exist at linear order [49].

[14] We can also apply this definition within a causal diamond of an AdS spacetime without reflecting boundary conditions.

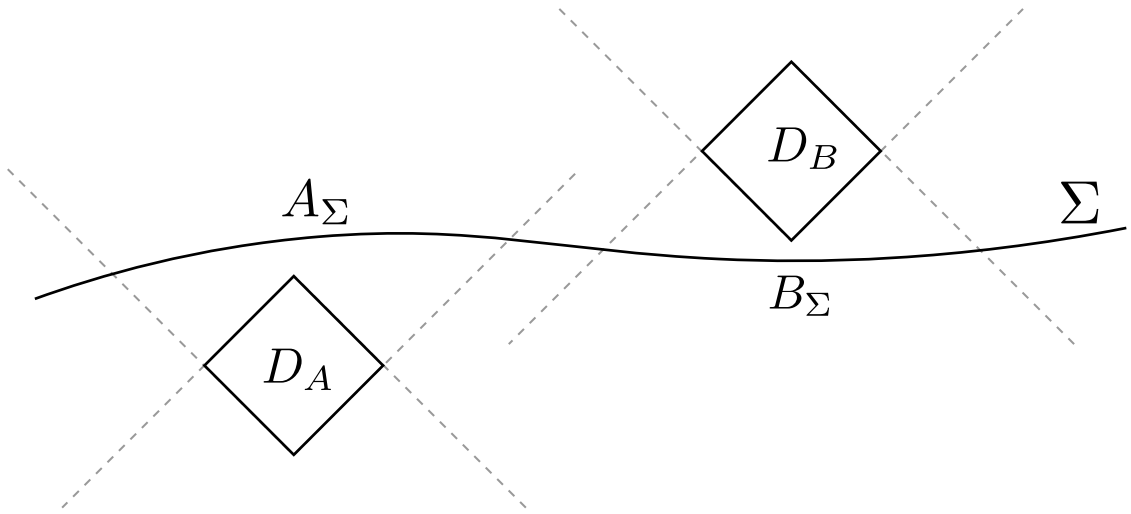


Figure 6.3: D_A and D_B are independent if there exist a slice Σ such that A_Σ and B_Σ are independent.

of independence to two causal diamonds D_A and D_B. First, if D_A and D_B overlap or are timelike separated, they should clearly not be considered independent, so assume that they are spacelike separated. Let Σ be any (AdS-)Cauchy slice Σ, and define

$$\begin{aligned} A_\Sigma &= \Sigma \cap (J^+[D_A] \cup J^-[D_A]), \\ B_\Sigma &= \Sigma \cap (J^+[D_B] \cup J^-[D_B]), \end{aligned} \tag{6.7}$$

where $J^+[X]$ ($J^-[X]$) is the usual causal future (past) of a given set X. We say that D_A and D_B are independent if there exists a Cauchy slice Σ such that B_Σ and A_Σ are independent, in the sense already defined. See Fig. 6.3. This is natural, since by causality and global hyperbolicity, the gravitational initial data on A_Σ, B_Σ contains all information about D_A, D_B. Thus, any perturbation in D_A and D_B can, through the evolution equations, be translated into a corresponding perturbation on A_Σ and B_Σ.

Second, the reason that we only consider perturbative deformations is that if we allow large deformations of the initial data, there is no sense in which the regions are fixed. If we could modify the data in A, B arbitrarily much, we could deform it to any data that is compatible with the topologies of A, B, or use a diffeomorphism to move the boundaries of A, B arbitrarily much. But then we would in the end only be asking the following question: does there exist interpolating initial data between every possible choice of initial data on two regions with the given topologies of A and B. But why do we even care to keep topologies fixed at this point? Thus, going beyond the perturbative level, we quickly lose all connection to our original question. This problem is reflective of the deeper underlying fact that, on the global phase space of GR, there is probably no useful diffeomorphism invariant notion of a subregion. However, when we restrict to perturbative deformations, we let the regions and the states become fuzzy, but only at the scale of the perturbative parameter. In language used in AdS/CFT contexts, asking about independence of two spacetime regions seems to be a natural question only in a code subspace built of perturbations around a single background

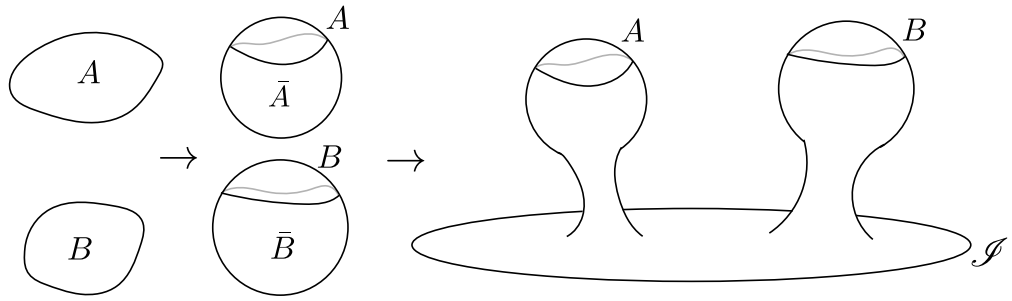


Figure 6.4: Using the gluing construction of [57] to embed two compact datasets in a complete dataset with a pre-specified conformal infinity $\mathscr{I}$.

(and, at the quantum level, in the small$-G_N$ limit).

It is worth making a few more comments on the non-perturbative case. However, this is a detour, and the reader can freely skip the next section without any loss of coherence in the rest of the paper.

Gluing constructions and independence

Consider making small deformations of initial data in $A, B \subset \Sigma$, but not requiring that the change outside $A \cup B$ is small – i.e. we allow jumping to a point in the GR phase space that is not near our original point. Can we then find an initial dataset interpolating between A and B, and which still has the requisite number of conformal boundaries? While this question has little relevance to understanding perturbative quantization around a background, it is nevertheless interesting to see what happens in this case. In vacuum gravity, it can be partially answered, thanks to the gluing result of [57]. This result says that in vacuum gravity, if Σ is a smooth initial dataset and $\Omega_1, \Omega_2 \subset \Sigma$ are two open sets whose domains of dependence have no Killing vectors, then we can cut out geodesic balls in both Ω_1, Ω_2 and glue in a handle $[0,1] \times S^{d-1}$ to connect Ω_1, Ω_2 with a wormhole. Furthermore, this can be achieved so that the final initial data is smooth and unchanged outside $\Omega_1 \cup \Omega_2$. Using this, if $A, B \subset \Sigma$ are compact and sufficiently generic, then we can embed them in a common initial dataset through the following procedure: adjoin A and B to some compact initial datasets $\bar{A}$ and $\bar{B}$ such that $\bar{A} \cup A$ and $\bar{B} \cup B$ are complete compact initial datasets, with $\bar{A}, \bar{B}$ both having at least one neighbourhood whose domain of dependence has no Killing vectors. Then take another sufficiently generic initial dataset with the required number of conformal boundaries, and use the gluing result to attach two wormholes to it – one connecting to $\bar{A}$ and one to $\bar{B}$. See Fig. 6.4. One can see that a similar procedures works if A and/or B merely have compact boundaries, so that A and B might contain complete connected components of conformal infinity. However, if ∂A and ∂B themselves are anchored to conformal boundaries, while we again can find a common initial dataset containing A and B using gluings, the complete set of initial data might have too many conformal boundaries, since the naive type of procedure described above will not connect any of the subregions of the conformal boundary that might be present in A and B.

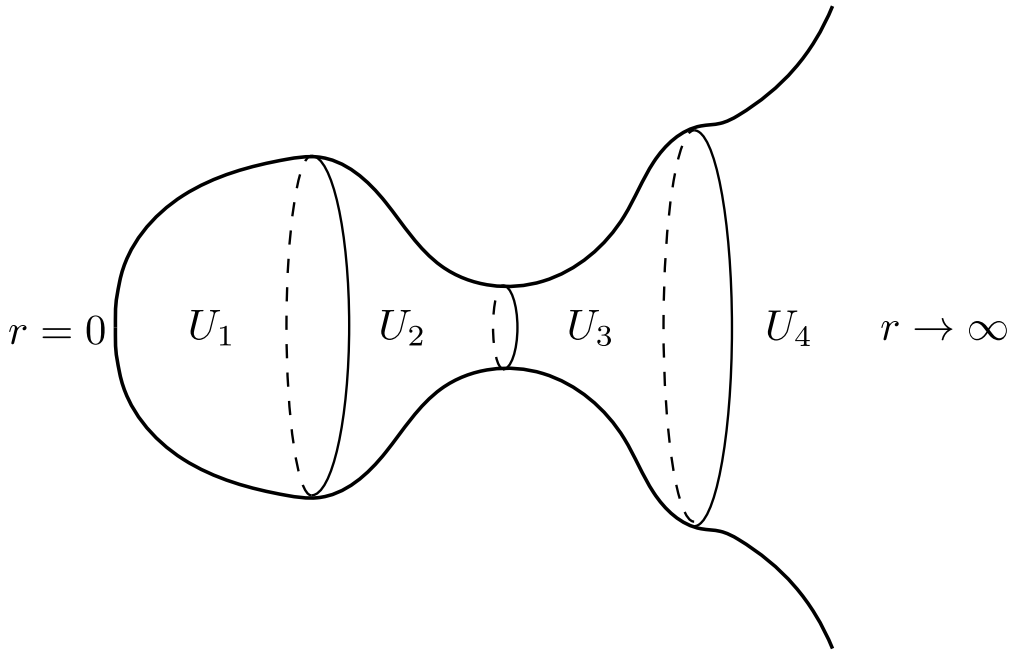


Figure 6.5: An example of a $d = 2$ spacelike slice Σ covered by four coordinate patches of the type (6.9), separated by every type of stationary surface. At these surfaces, $B = \infty$.

6.3 How Background Structures Enable Independence

In this and the following section, we will show how extremal surfaces, matter, and generic trapped surfaces can be used to dress perturbations, avoiding dressing to asymptotic regions. They are background structures that enable subregion independence. We will first study a set of illustrative examples in this section, where we restrict to particular backgrounds with spherical symmetry. Then, in Sec. 6.4 we will give a more general discussion, relaxing many of the simplifications made in our examples.

6.3.1 Setup

Our example will be Einstein gravity in $d+1$ spacetime dimensions coupled to a massless scalar, with action

$$I = \int \mathrm{d}^{d+1}x\sqrt{-g}\left[R + \frac{d(d-1)}{L^2} - \nabla_a\phi\nabla^a\phi\right]. \tag{6.8}$$

For positive cosmological constant, L is imaginary, and we set $L = iL_{\mathrm{dS}}$.

Consider now a spherically symmetric spacelike initial data slice Σ. Because of spherical symmetry, we can cover Σ with coordinate patches U_i, such that on each patch, we have local coordinates

$$h_{\mu\nu}\mathrm{d}y^\mu\mathrm{d}y^\nu = B(r)\mathrm{d}r^2 + r^2\mathrm{d}\Omega^2, \tag{6.9}$$

where $\mathrm{d}\Omega^2$ is the round metric S^{d-1}. These coordinates break down at some given radius $\hat{r}$ if and only if $B(\hat{r}) = \infty$. $B(\hat{r}) = 0$ can be shown to correspond to curvature singularity

for h_{ab}, which we do not consider. By computing the mean curvature of a sphere of radius r within Σ, we can see that $B = \infty$ precisely corresponds to the sphere having vanishing mean curvature [58], meaning that the sphere is locally minimal, maximal, or saddle-like within Σ. See Fig. 6.5.

Next, assume furthermore the mean extrinsic curvature in spacetime vanishes: $K_a{}^a = 0$. For nonpositive cosmological constant, this usually means that Σ is a maximal volume slice, while for a positive cosmological constant, it can also mean that Σ has minimal volume. The extrinsic curvature is conveniently parametrized by a single function $\mathcal{K}(r)$ as[15]

$$K_{\mu\nu}\mathrm{d}y^\mu \mathrm{d}y^\mu = \mathcal{K}(r)\left[B(r)\mathrm{d}r^2 - \frac{r^2}{B(r)(d-1)}\mathrm{d}\Omega^2\right]. \tag{6.10}$$

$\mathcal{K}(r)$ is simply the rr-component in an orthonormal basis, and so must be everywhere finite, otherwise Σ would have a singular embedding in spacetime. Furthermore, on a slice Σ whose embedding in spacetime does not have any kinks, $\mathcal{K}(r)$ must be continuous across coordinate patches.

It turns out that our analysis will greatly simplify if we introduce the function $\omega(r)$ as

$$B(r) = \frac{1}{1 + \frac{r^2}{L^2} - \frac{\omega(r)}{r^{d-2}}}. \tag{6.11}$$

In terms of this variable, a stationary surface is characterized by

$$0 = 1 + \frac{\hat{r}^2}{L^2} - \frac{\omega(\hat{r})}{\hat{r}^{d-2}}. \tag{6.12}$$

When we patch together two coordinate systems at some $r = \hat{r}$, ω must match in the two patches, since $\omega(\hat{r})$ is uniquely fixed by $\hat{r}$. Next, we have that in a patch containing conformal infinity, where we can take $r \to \infty$, that [58][16]

$$M = \frac{(d-1)\mathrm{Vol}[S^{d-1}]}{16\pi G_N}\omega(\infty), \tag{6.13}$$

where M is the ADM mass, or the AdS analogue. We refer to ω as the Hawking mass.[17]

Computing the constraints in our setup, we find the equations

$$(d-1)\frac{\omega'(r)}{r^{d-1}} = \left(1 + \frac{r^2}{L^2} - \frac{\omega}{r^{d-2}}\right)\phi'(r)^2 + \dot{\phi}(r)^2 + \frac{d}{d-1}\mathcal{K}(r)^2, \tag{6.14}$$

$$\frac{\mathrm{d}}{\mathrm{d}r}\left[r^d\mathcal{K}(r)\right] = r^d\phi'(r)\dot{\phi}(r), \tag{6.15}$$

where the first equation is the Hamiltonian constraint, and the second the momentum constraint. Since each coordinate patch has its own set of functions $\{\omega, \mathcal{K}, \phi, \dot{\phi}\}$, we will

[15] There is only one function's worth of degrees of freedom in K_{ab}, thanks to spherical symmetry and $K^a{}_a = 0$.

[16] This relies on $K^a{}_a = 0$, or that $K_a{}^a \to 0$ sufficiently fast at infinity.

[17] There are two versions of the Hawking mass: the Riemannian (also known as the Geroch-Hawking mass) [33, 59–61] and the Lorentzian versions [62–65], with the former not being directly sensitive to K_{ab}. ω is the Riemannian version, suitably generalized to $d \neq 3$ and $\Lambda \neq 0$, in the spacial case of spherical symmetry.

sometimes write ω_{U_i} to indicate the ω-function on the i-th patch, and similarily for other quantities. To obtain the solution on all of Σ, we solve these equations on each patch, matching $\omega, \mathcal{K}$ across each junction. In a spacetime with two asymptotic regions, there are two integration constants, corresponding to ω and $\mathcal{K}$ at a single value of r. These are the purely gravitational degrees of freedom in spherical symmetry. Without matter, where every solution is (AdS-)Schwarzschild by the Birkhoff-Jebsen theorem [66–69], these numbers determine M and $t_L + t_R$, where t_L, t_R are the times on the left and right conformal boundaries at which Σ is anchored. With only one conformal boundary, or in a spacetime without any conformal boundary, $r = 0$ is present on at least one coordinate patch. The spacelike nature of Σ ($B > 0$) then requires that $\omega(r = 0) = 0$, while smoothness of the embedding of Σ in spacetime requires $\mathcal{K}(r = 0) = 0$.

Now come the important observations. Since we are on a spacelike slice, we must have $B(r) > 0$, meaning that the prefactor of $\phi'(r)^2$ term in (6.14) is positive, and thus also the full right hand side. We conclude that

$$\omega'(r) \geq 0. \tag{6.16}$$

ω is monotonically non-decreasing as we move towards increasing r. This conclusion does not rely on our particular theory. It is true as long as we have the WEC (6.5). It is the monotonicity of ω that makes excitations in gravity more difficult to screen. However, even with one conformal boundary, the direction of increasing r does not always point towards the conformal boundary. This is at the heart of all our subsequent results. It means that adding matter in a patch of spacetime were r is decreasing towards the boundary can decrease the ADM mass, despite the fact that the matter has positive local energy density.

We are now going to use Eqs. (6.14) and (6.15) to study subregion independence and dressing of gravitational perturbations. This requires us to study the linearized versions of (6.14) and (6.15) around some background geometry. We will restrict to spherical perturbations preserving $K^a_a = 0$ in our explicit analysis, and give arguments that our conclusions are unchanged under general perturbations. Now, the scalar matter corresponds to completely free data; thus we can organize our perturbative expansion as follows:

$$\begin{aligned}
\phi &= \phi_0(r) + \kappa \delta\phi(r), \\
\dot{\phi} &= \dot{\phi}_0(r) + \kappa \delta\dot{\phi}(r), \\
\omega &= \omega_0(r) + \sum_{i=1}^{\infty} \kappa^i \delta_i \omega(r), \\
\mathcal{K} &= \mathcal{K}_0(r) + \sum_{i=1}^{\infty} \kappa^i \delta_i \mathcal{K}(r),
\end{aligned} \tag{6.17}$$

where $\delta\phi, \delta\dot{\phi}$ is free data, and with $\delta_i\omega, \delta_i\mathcal{K}$ determined by the constraints. κ is the small perturbative parameter controlling our expansion.[18] We could also adjust $\delta\phi, \delta\dot{\phi}$ more finely at higher orders, but we will not need this.

[18] Morally, we could think of κ as $\sqrt{G_N}$, although in the purely classical theory, there is no preferred universal scale. If we were to sum the series, κ should be adjusted according to the scales of the background solution.

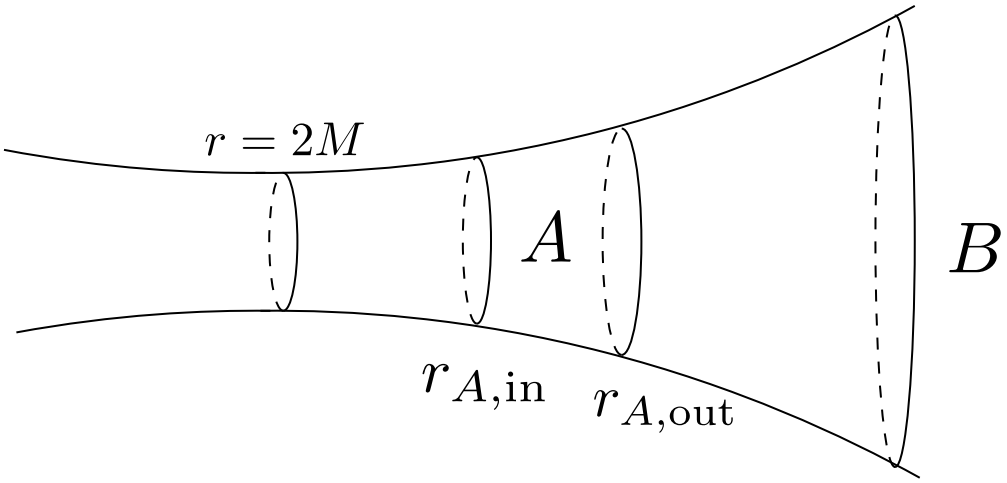


Figure 6.6: A bounded annular region A on a $t = 0$ slice in Schwarzschild, with one spatial dimension suppressed. A and B are dependent regions.

Note that the fully non-perturbative solutions of (6.14), (6.15) can be written down exactly as nested integrals over the sources $\phi(r), \dot{\phi}(r)$ (see for example [58]), but we will not need these solutions.

6.3.2 Subregion independence in asymptotically flat and AdS spacetimes

We have already shown how a compact region in the center of AdS or Minkowski depends on an annulus around infinity. From the perspective of spherically symmetric perturbations, this follows from integrating (6.14) from $r = 0$ to $r = \infty$ and using $\omega(r = 0) = 0$, giving that

$$\omega(\infty) = \omega(\infty) - \omega(0) = \int_0^\infty \mathrm{d}r(\text{positive}) > 0. \tag{6.18}$$

Thus, let us now consider more interesting geometries. For notational convenience we take $d = 3$ and a vanishing cosmological constant – the analysis is virtually identical for AdS and/or $d \neq 3$. We will treat the $\Lambda > 0$ case separately.

Schwarzschild

Let us consider the simplest non-trivial background: Schwarzschild with mass M. Let Σ be a canonical $t = 0$ slice, which has $K_{ab} = 0$. Let B be an infinite annulus containing right infinity, and let A be a finite annulus in the right region, restricted to $r \in [r_{A,\text{in}}, r_{A,\text{out}}]$, with $r_{A,\text{in}} > 2M$. See Fig. 6.6. The perturbed leading order constraints read[19]

$$\begin{aligned} 2\delta_2\omega'(r) &= r^2 f(r)\delta\phi(r)^2 + r^2\delta\dot{\phi}^2, \\ \frac{\mathrm{d}}{\mathrm{d}r}\left(r^3\delta_2\mathcal{K}(r)\right) &= r^3\delta\phi'(r)\delta\dot{\phi}(r), \end{aligned} \tag{6.19}$$

[19]Since the leading order perturbation of δT_{ab} is $\mathcal{O}(\kappa^2)$, we do not consider $\delta_1\omega \neq 0$.

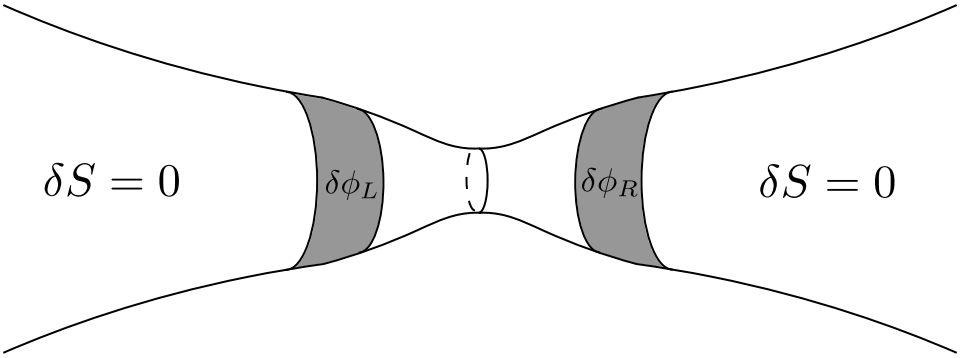


Figure 6.7: A perturbation $\delta\phi_R$ on the right side dressed to a perturbation on the left side, leaving the initial data both asymptotic regions unchanged. This causes the area of the minimal surface to decrease.

with $f(r) = 1 - 2M/r$. The free data in A are the functions $\delta\phi(r), \delta\dot{\phi}(r)$ for $r \in [r_{A,\text{in}}, r_{A,\text{out}}]$, and two numbers, corresponding to $\delta_2\omega, \delta_2\mathcal{K}$ a single value of the radius in A. These two numbers are the only spherically symmetric gravitational degrees of freedom. To test for independence, we are free to consider a perturbation with $\delta_2\omega(r_{A,\text{in}}) = \delta_2\mathcal{K}(r_{A,\text{in}}) = 0$, and add some matter in A. With the choice $\delta_2\omega(r_{A,\text{in}}) = \delta_2\mathcal{K}(r_{A,\text{in}}) = 0$, we are choosing $\delta S|_A$ such that the new gravitational field sourced by $\delta\phi$ is forced to travel to the right. Then, integrating (6.19) from $r_{A,\text{in}}$ to $r = \infty$ gives that $\delta_2\omega(\infty) > 0$ on the right, and so A and B are dependent. The background has no matter that we can remove outside A to screen the perturbation. One could wonder if a non-symmetric perturbation outside A could be used to screen the perturbation, but in this case, we can appeal to a rigorous theorem, the so-called Riemannian Penrose inequality [70–72],[20] to show that no perturbation whatsoever can save us.[21] This theorem says that if we have an asymptotically flat initial dataset with $K_a{}^a = 0$, with one asymptotic end with mass M, and with an inner boundary σ that is an outermost minimal surface, then

$$M \geq \sqrt{\frac{\text{Area}[\sigma]}{16}}, \tag{6.20}$$

with equality if and only if we are identically Schwarzschild. Our chosen perturbation in A was compatible with keeping the geometry at the minimal surface fixed, and so if a non-symmetric dressing that did not reach B existed, it would violate the rigidity statement of the Riemannian Penrose inequality. So graviton degrees of freedom cannot make A and B independent.

[20] In the AdS case, this inequality is still conjectural, and only partial proofs exist [58, 73–76].

[21] We do assume that $K^a_a = 0$ is preserved under these perturbations, but is not a severe restriction, since small perturbations should not break the existence of a slice with maximal volume. So even if we perturbed away from $K^a_a = 0$, there should be a nearby slice in spacetime where $K^a_a = 0$, so we can always take our perturbation to be a combination of the perturbation that changes spacetime, together with a diffeomorphism that moves us to this slice.

We thus see that any bounded region A lying strictly in the right exterior is dependent on right infinity. Is A dependent on left infinity? To test this, again pick a perturbation where we add some matter in A. However, we now use our gravitational degree of freedom to set $\delta_2\omega(r_{A,\text{out}}) = \delta_2\mathcal{K}(r_{A,\text{out}}) = 0$ instead, forcing the new gravitational field lines sourced by $\delta\phi$ to travel to the left. This seems dangerous: will we now be forced to dress to left infinity? The answer is no! We can screen the perturbation from left infinity by **adding** more matter, provided we add it to the left of the bifurcation surface. Integrating the constraints, and matching $\delta_2\omega, \delta_2\mathcal{K}$ across the coordinate systems, we find that the functions on the left side are[22]

$$2\delta_2\omega_L(r) = \int_{2M}^{r} \mathrm{d}\rho\rho^2 \left[f(\rho)(\delta\phi_L')^2 + (\delta\dot{\phi}_L)^2\right] + \int_{r_{A,\text{out}}}^{2M} \mathrm{d}\rho\rho^2 \left[f(\rho)(\delta\phi_R')^2 + (\delta\dot{\phi}_R)^2\right], \quad (6.21)$$

$$\delta_2\mathcal{K}_L(r) = \int_{2M}^{r} \mathrm{d}\rho\rho^3\delta\phi_L'(r)\delta\dot{\phi}_L + \int_{r_{A,\text{out}}}^{2M} \mathrm{d}\rho\rho^3\delta\phi_R'(r)\delta\dot{\phi}_R. \quad (6.22)$$

The crucial thing to note in these equations is that the last term in (6.21) is negative, caused by that fact that r increases in opposite directions on each side of the bifurcation surface. Locally positive energy densities on one-side of an extremal surface contribute negatively to the Hawking (and thus ADM) mass on the other side. Thus, we see that we can always pick $\delta\phi_L, \delta\dot{\phi}_L$ such that $\delta_2\omega_L(r) = 0, \delta_2\mathcal{K}_L(r) = 0$ for all $r > 2M + \epsilon$ for any $\epsilon > 0$ (assuming ϵ does not scale with κ to a positive power). So a positive energy density shell can be dressed to another positive energy density shell, provided they are separated by an extremal surface. See Fig. 6.7. As advertised earlier, now gravity behaves more like electromagnetism. Positive energy densities seen from the other side of an extremal surface behaves as a negative energy density.

What about perturbations in A that are not spherically symmetric, but where we keep the solution in A fixed near the rightmost boundary, so that we are forcing the gravitational field to change towards the left. Can these perturbations be screened from left infinity? It is physically quite clear that they can, although we will give an argument rather a proof. First, using a technique known as inverse mean curvature flow, which we explain in Sec. 6.4.3, it can be shown for $d = 3$ that non-spherical additions of matter on the right will again contribute negatively to the mass on the left. So again we can simply add matter on the left to bring this back up, keeping left the mass unchanged. Of course, now we might also have to cancel the momentum and angular momentum, but this is much easier since these are quantities without a preferred sign. Let us assume we added some amount of angular momentum in A that is backreacting leftwards. We might worry that the required matter or gravitons we need to add to cancel this angular momentum forces us to overshoot the mass, causing a left mass increase. But this can always be avoided. Rather than cancelling the angular momentum by adding matter on the left, we can add it on the right side, between A and the minimal surface, so that we bring the angular momentum closer to zero, while at the same time decreasing the mass. Thus, as long as A is slightly separated from the bifurcation surface, so that we

[22]The reader might be worried that $\delta_2\omega \neq 0$ at $r = 2M$, since the rr-metric perturbation there reads, in our gauge, $\delta_2 h_{rr} = r\delta\omega(r - 2M)^{-2}$, which diverges at $r = 2M$. However, this is simply a gauge artifact that can be fixed with the addition of a term $\nabla_{(a}\xi_{b)}$. It is simply reflecting the fact the location of the minimal surface, and thus coordinate breakdown location of our gauge, has moved by a perturbative amount.

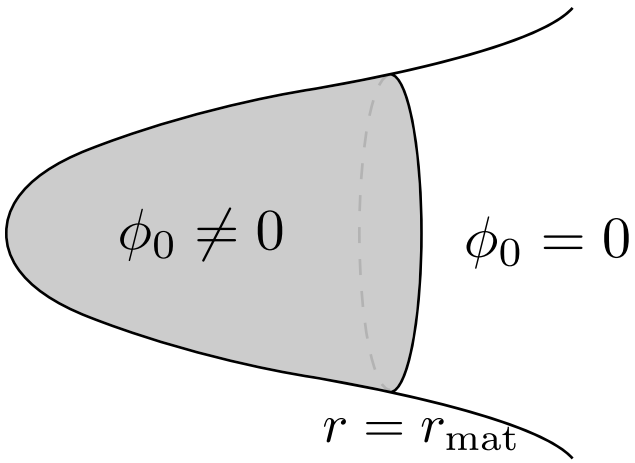


Figure 6.8: A one-sided spacetime with a lump of matter around $r = 0$. We argue that a ball A with radius r_A is independent from an annulus B around infinity if and only if $r_A < r_{\rm mat}$

have the flexibility of perturbing on both sides of it, A ought to be independent of B.

What we see from this discussion is that gravity appears to have a classical version of the split property across the Schwarzschild bifurcation surface. Let A and B be the regions $r \geq 2M + \epsilon$ on the left and right, respectively, with $\epsilon > 0$ not scaling with κ. By the discussion above, we see that A, B are likely independent of each other. However, it is crucial that ϵ is non-zero, and in fact non-perturbatively large, so that we can always fit enough matter without leaving perturbation theory. Thanks to the fact that the small band between A, B contains an extremal surface, we have access to perturbations with all combinations of signs of charges within this band. We will see later that this property is a feature of general extremal surfaces.

A spacetime with a lump of matter

Consider now a one-sided spherically symmetric geometry at a moment of time-symmetry, $\mathcal{K}_0 = \dot{\phi}_0 = 0$. Assume that it has no stationary surfaces, so that one coordinate patch covers everything. Next, we take there to be a lump of matter $\phi_0(r)$ with compact support on $r \in [0, r_{\rm mat}]$. See Fig. 6.8. We demand that $\phi_0'(r)$ is non-zero at least in $r \in [r_{\rm mat} - \epsilon, r_{\rm mat})$ for some $\epsilon > 0$.

Let A be a ball with radius r_A, and take B to be an annulus around infinity. Consider turning on any scalar field profiles $\delta\phi(r), \delta\dot{\phi}(r)$ in A. The perturbative constraints are now non-trivial at first order in κ, and read

$$\begin{aligned} \delta_1\omega'(r) &= r^2\left[1 - \frac{\omega_0(r)}{r}\right]\phi_0'(r)\delta\phi'(r) - \frac{1}{2}r\delta_1\omega(r)\phi_0'(r)^2, \\ \frac{\mathrm{d}}{\mathrm{d}r}\left(r^3\delta_1\mathcal{K}\right) &= r^3\phi_0'(r)\delta\dot{\phi}(r). \end{aligned} \tag{6.23}$$

Can we then extend the perturbation to $r > r_A$ such that no backreaction leaks to infinity?

If $r_A < r_{\rm mat}$, then the answer is yes. The first order constraint solutions are

$$\delta_1\omega(r) = e^{-\frac{1}{2}\int_0^r \mathrm{d}zz\phi_0'(z)^2}\int_0^r \mathrm{d}\rho\rho^2 e^{\frac{1}{2}\int_0^\rho \mathrm{d}zz\phi_0'(z)^2}\left[1-\frac{\omega_0(\rho)}{\rho}\right]\phi_0'(\rho)\delta\phi'(\rho) \tag{6.24}$$

$$\delta_1\mathcal{K}(r) = \frac{1}{r^3}\int_0^r \mathrm{d}\rho\rho^3\phi_0'(\rho)\delta\dot{\phi}(\rho) \tag{6.25}$$

We now see that neither integral has fixed sign, so as long as $r_A < r_{\rm mat}$[23] we can always pick $\delta\phi, \delta\dot{\phi}$ on $(r_A, r_{\rm mat}]$ so that there is some $\hat{r} \in (r_A, r_{\rm mat}]$ such that

$$\delta_1\omega(r \geq \hat{r}) = 0, \quad \delta_1\mathcal{K}(r \geq \hat{r}) = 0. \tag{6.26}$$

Physically, what has happened is that we have removed some matter from the background shell in order to avoid having the backreaction of our newly added matter leak out to infinity. Since we are considering infinitesimal perturbations, the background reservoir is effectively infinitely large, so we can keep doing this to any order in perturbation theory. All spherically symmetric perturbations in A can be dressed to the background lump of matter. Similarly, if we alter the mass in B, we can always add or remove some of the matter in the shell lying in C to compensate, keeping A fixed. Assuming we can dress non-symmetric perturbations to the lump as well, we thus find that A, B now are independent. It is physically reasonable that it should also be possible to dress these perturbations to B. If our non-spherical perturbation in, say, A, adds some momentum or angular momentum, we can just add more moving or rotating matter between A and B to cancel these charges, since they do not have a preferred sign. The only real worry is that this matter will increase the energy at the same time, but since we are working perturbatively, we have an infinite background reservoir of matter, so we can always remove a bit of matter from the lump to keep the energy fixed. So the background lump of matter restores independence of the ball A from infinity, and perturbations in A need not be dressed to infinity.

If we instead took $r_A > r_{\rm mat}$, independence breaks down. This is because the spacetime looks like Schwarzschild for $r > r_{\rm mat}$, and if we pick a perturbation on A that changes the region $(r_{\rm mat}, r_A]$, but keeps everything fixed for $r \in [0, r_{\rm mat}]$, then we have specified a perturbation where no matter from the background is allowed to be removed, and we have no choice but to let the backreaction leak to infinity. This time there is no minimal surface to dress to. Non-symmetric perturbations in C cannot save us, since these would constitute all-order compactly supported perturbations of pure Schwarzschild in one exterior, and these perturbations could be used contradict the rigidity-part of the Riemannian Penrose inequality.

This example illustrated dressing of perturbations to background matter. It was important that this background matter was not just perturbatively small, since then the argument would break down – there would exist perturbations that could not be dressed to matter, since the added matter then exceed what we might have the ability to remove. At the level of the full perturbative series, the radius of convergence of κ thus has to be less than infinity.

Schwarzschild in pure gravity

We already discussed subsystem independence in Schwarzschild, but we crucially relied on matter. What happens in vacuum gravity? Consider again a canonical $t = 0$ slice of

[23] And $r_A - r_{\rm mat} = \mathcal{O}(\kappa^0)$.

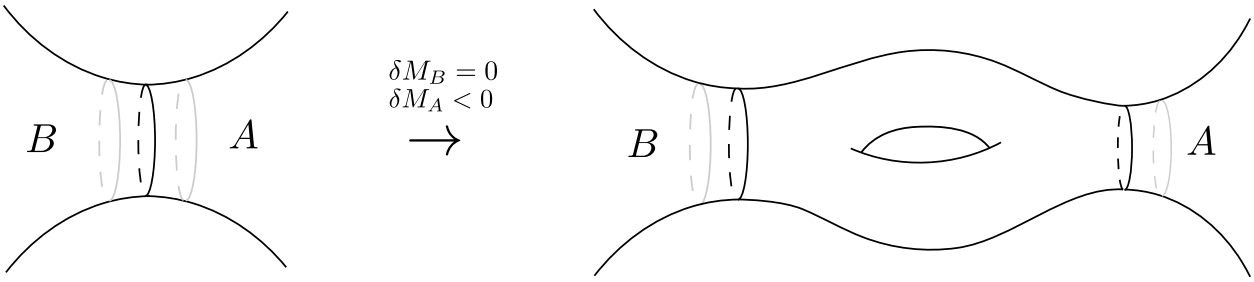


Figure 6.9: Using pure gravity degrees of freedom to change the mass of the BTZ region on the A-side without changing it on the B-side. The right geometry looks identical to BTZ in A and B.

(AdS-)Schwarzschild, and let A and B be the regions $r \geq 2M + \epsilon$ on the right and left, respectively. Consider now picking a perturbation $\delta S|_A$ that corresponds to simply deforming A to Schwarzschild with a lower mass. Pure gravity has no local spherically symmetric degrees of freedom, so we can clearly not hide this perturbation from B within spherical symmetry. However, it appears likely that the perturbation can be hidden from B if we break spherical symmetry, so that we unlock the graviton degrees of freedom. Let us see how this happens in a similar situation: the BTZ black hole. Unlike higher dimensions, we do not have local degrees of freedom. However, vacuum gravity still possesses topological degrees of freedom. Once we leverage these, it is well known (see for example [77, 78]) that we can reduce the BTZ mass in A without changing B by filling in a higher-genus surface between A and B – see Fig. 6.9. Of course, in this case, the change is not perturbative, since we changed the topology, so pure 3d gravity at best has a non-perturbative notion of independence. Nevertheless, the example is brought up to make it less surprising that pure gravity degrees of freedom might do the job for $d \geq 3$. In fact, there are mathematical results in GR supporting this possibility. As long as we are allowed to change the geometry at the bifurcation surface, we are not constrained by the Riemannian Penrose inequality, and Corvino and Schoen [79, 80] have shown that being identically Schwarzschild in a neighbourhood of infinity is not a rigid feature.[24] Furthermore, by Theorem 1.1 in [81], Schwarzschild is known to support first order pure gravity perturbations of compact support. Thus, it is a possibility that any perturbation on A can be shielded from B by utilizing pure gravity degrees of freedom in C.

6.3.3 Subregion independence in de Sitter

Let us now treat a case with positive cosmological constant. Let the background be a minimal slice of dS_4, which has $K_{ab} = 0$, and which can be covered by two coordinate patches U_L, U_R, each having a range $r \in [0, L_{\rm dS}]$. The metric on each patch is

$$ds^2 = \frac{dr^2}{1 - \frac{r^2}{L_{\rm dS}^2}} + r^2 d\Omega^2, \tag{6.27}$$

[24] Specifically, they proved that for any asymptotically flat vacuum solution (h_{ab}, K_{ab}) of the constraints on $\mathbb{R}^3$, and for any choice of compact subset $\Omega \subset \mathbb{R}^3$, there exists new vacuum initial data (h'_{ab}, K'_{ab}) that looks identical to Kerr near infinity, and which agrees with (h_{ab}, K_{ab}) on Ω.

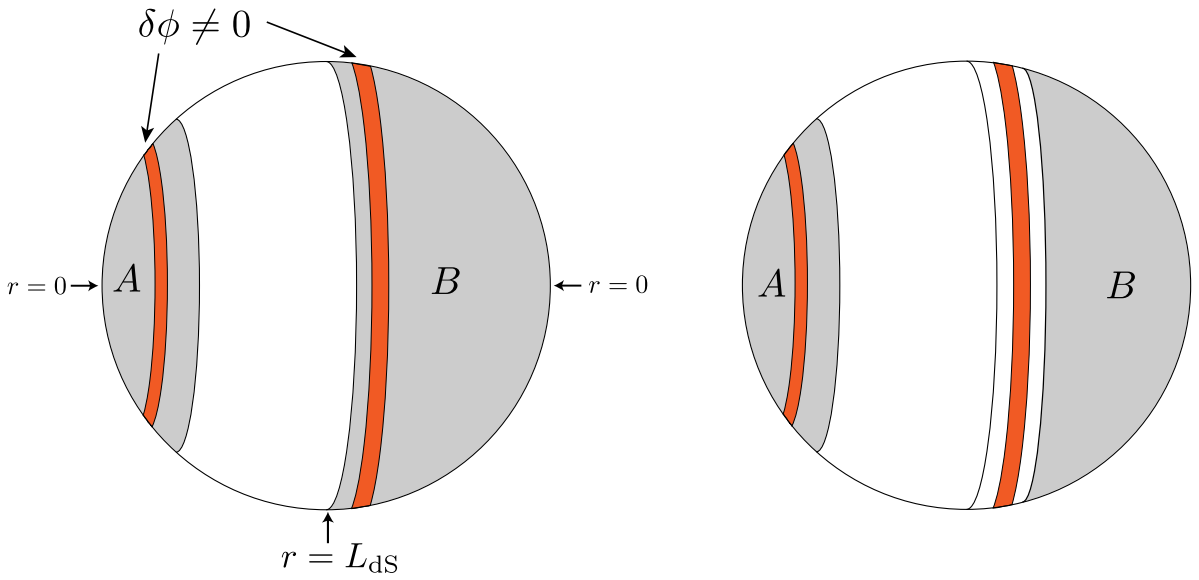


Figure 6.10: A perturbation of a minimal slice of de Sitter, which is just a round sphere of radius $L_{\rm dS}$. The two shells are dressed to each other, so the solution looks like pure de Sitter near the poles of A and B. At the level of spherical symmetry, adding a shell to the one static patch requires the addition of one to the other. Thus it is possible that A and B are dependent in the left scenario. On the right, we can screen perturbations in A from B by adding matter around the maximal surface.

i.e. $\omega_0(r) = 0$. Together these two patches make up a round sphere of radius $L_{\rm dS}$. The spherically symmetric constraints are non-trivial at $\mathcal{O}(\kappa^2)$, and given by (6.19), except now $f(r) = 1 - r^2/L_{\rm dS}^2$. At $r = 0$, we as usual need that $\delta_2\omega(0) = 0$. Integrating (6.19) and using that $\delta_2\omega$ is continuous at $r = L_{\rm dS}$, all perturbations are constrained to satisfy

$$\int_0^{L_{\rm dS}} {\rm d}\rho\rho^2 \left[f(\rho)(\delta\phi_L')^2 + (\delta\dot{\phi}_L)^2\right] = \int_0^{L_{\rm dS}} {\rm d}\rho\rho^2 \left[f(\rho)(\delta\phi_R')^2 + (\delta\dot{\phi}_R)^2\right] \tag{6.28}$$

This is just a special case of the first law of thermodynamics for positive cosmological constant [82] applied to spherical perturbations of pure dS. Since U_L and U_R are Cauchy slices for the left and right static patches, we see that within the domain of spherical symmetry, it is impossible to perturb one static patch without also perturbing the other. See Fig. 6.10. It is easy to show the same result for all $d \geq 2$. Thus, it seems plausible in ${\rm dS}_{d+1}$ for $d \geq 2$, any subregion of a given static patch depends on the full opposite static patch. However, we have no proof that we cannot break spherical symmetry on the right and leverage pure vacuum gravity degrees of freedom to shield this perturbation from the left. As we discuss in Sec. 6.5, there very likely exist at least some non-symmetric non-trivial initial data that is strictly localized to one static patch. Either way, A is probably independent of any B that is a strict subset of the opposite static patch, provided B is such that C contains a neighbourhood of the cosmological horizon. In this case, we can satisfy (6.28) by adding matter in a small neighbourhood around the cosmological horizon, as illustrated in Fig. 6.10.

In recent work, a von Neumann algebra of diffeomorphism invariant operators was constructed for a static patch of de Sitter that contains an observer [83]. This algebra

consisted of operators dressed to the observer. It was found that in order to have a global Hilbert space also describing the opposite static patch, it was necessary to include an observer in the second patch as well. It is interesting to speculate whether this could be connected to (6.28). An observer ought to backreact on the geometry, and the Hamiltonian constraint forbids backreaction in a single static patch (in spherical symmetry). If the observer were to travel on a worldline, this perturbation preserves spherical symmetry, and so some backreaction must thus be added in the other patch. Adding a second observer solves this. In our setup, it would however be appropriate to treat an observer as a feature of the background, rather than a perturbation, so that dressed observables act as perturbations with respect to the observer. To treat this case, and to get a better sense of when we can deform a single static patch, we prove a stronger theorem in Sec. 6.5, showing that for full spherically symmetric non-linear backreaction and $d \geq 2$, it is not possible to change a single static patch. We do this by showing that the area of the cosmological horizon must be reduced by any of these deformations. However, before we do this, we first study gravitational independence at the more general level.

6.4 Dressing Across Extremal and Trapped Surfaces

Above we studied examples of spherically symmetric backgrounds, and we argued that at moments of time-symmetry, we can always dress perturbations across minimal or maximal surfaces, at least when the theory has matter. The assumption of time-symmetry and spherical symmetry is however restrictive. Furthermore, it would be more useful to characterize the surfaces that we are able to dress across in terms of their spacetime properties, rather than their properties on a particular slice. We are now going to argue that we can dress across extremal surfaces and generic trapped surfaces. We start with general spherically symmetric spacetimes and perturbations in Secs. 6.4.1,6.4.2. Then in Sec. 6.4.3, we remove symmetry assumptions in the case of four spacetime dimensions.

6.4.1 Spherical symmetry

In spherical symmetry, the general constraint equations on a ${K_a}^a = 0$ slice Σ read

$$(d-1)\frac{\omega'(r)}{r^{d-1}} = 2\mathcal{E} + \frac{d}{d-1}\mathcal{K}(r)^2, \tag{6.29}$$

$$\frac{\mathrm{d}}{\mathrm{d}r}\left(r^d\mathcal{K}\right) = r^d\mathcal{J}_r. \tag{6.30}$$

By virtue of $\omega'(r) \geq 0$, as we cross any minimal or maximal surface σ, we always have a flip in the direction in which ω grows as it encounters matter. So if we add some matter on one side of σ, we can add some matter on the other side to compensate, keeping ω unchanged outside a neighbourhood of σ. The assumption of time-symmetry in our examples was irrelevant to this central point. However, at first sight, minimal and maximal surfaces in a $K^a_a = 0$ slice seem like a very restricted set of surfaces. It turns out this is not true. A simple computation shows that stationary surfaces on a ${K_a}^a = 0$ slice are always either strictly (anti)trapped or extremal (see for example Eq. (A.55) of [58]). In a moment we are going to show a partial

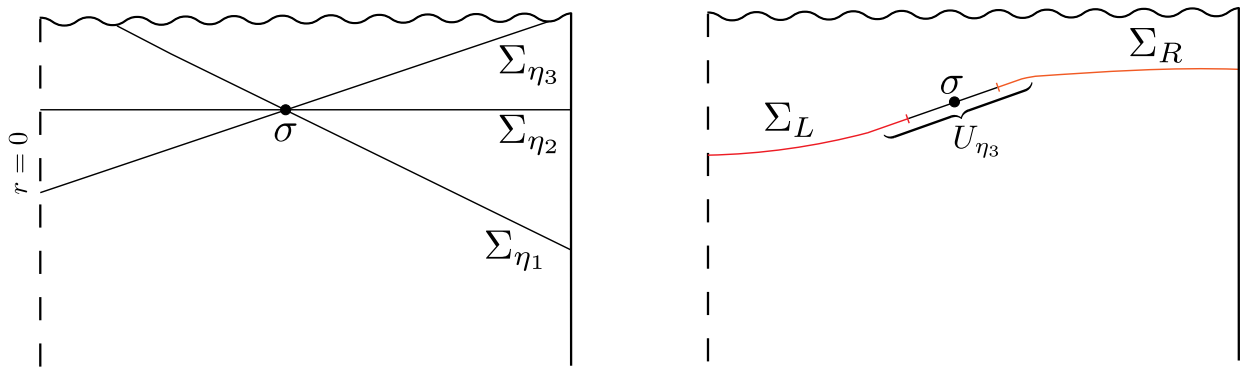


Figure 6.11: (Left) Three $K^a{}_a = 0$ slices fired off σ, with $\eta_3 > \eta_2 > \eta_1$. Only Σ_{η_2} is smooth and complete. (Right) Deforming Σ_{η_3} to a smooth complete slice, preserving a neighbourhood U_{η_3} around Σ that has $K^a{}_a = 0$.

converse in spherical symmetry: maximal or minimal surfaces on $K_a{}^a = 0$ slices include all (spherical) extremal surfaces and also all generic trapped surfaces, provided we do not require Σ to satisfy $K_a{}^a = 0$ globally.[25] This will be enough for us.

Consider a general spherically symmetric spacetime (N, g_{ab}) and let σ be any sphere. Then there always exists a one-parameter family Σ_η of spherically symmetric $K_a{}^a = 0$ slices locally defined in a neighbourhood around σ. To see this, note that the equation that determines Σ_η is a second order ODE. This ODE is obtained by extremizing the induced volume of Σ with respect to a single embedding coordinate, say $t(r)$. One integration constant is fixed by demanding that Σ_η passes through σ. The second integration constant is just the boost angle η at which Σ_η is "fired off" σ. This is completely analogous to how we can fire a one-parameter family of radial spacelike geodesics off some given point. Once η is fixed we can integrate the ODE determining the location of Σ_η to try to extend it to a full slice, but generically one does not get a full smooth Cauchy slice. See the left panel of Fig. 6.11. In a one-sided spacetime, for example, this could happen because we hit $r = 0$ with the wrong boost angle, so that Σ_η has a kink, corresponding to $\mathcal{K} \to \infty$. Or we could fall into a singularity. However, crucially, we will not need a full $K_a{}^a = 0$ slice. We can just terminate the integration at some finite value and then arbitrarily continue the slice in a smooth way, now giving up the requirement that the mean curvature vanishes. This way we produce a one-parameter family of Cauchy slices Σ_η, each containing σ and each having a neighbourhood $U_\eta \subset \Sigma_\eta$ that has $K^a{}_a = 0$. See the right panel of Fig. 6.11 for one such slice.

Consider now σ being a surface such that there exists an η so that σ is minimal or maximal on Σ_η. Let Σ_L and Σ_R be the parts of Σ_η lying to the left and right sides of σ, respectively, except for a small neighbourhood around σ that is included in neither – see Fig. 6.11. Thus, Σ_L and Σ_R are separated by an open neighbourhood containing σ that has $K^a{}_a = 0$. We now see that any spherical perturbation in Σ_L can be screened from Σ_R, and vice versa, by exactly the mechanism we described earlier.[26]

[25] Allowing saddle-type surfaces, all trapped surfaces are included.

[26] Strictly speaking, we should also consider perturbations that satisfy $\delta K_a{}^a \neq 0$ at $\partial\Sigma_R, \partial\Sigma_L$, so we should technically have consider the spherically symmetric Einstein equations for general $K^a_a = 0$. This leads to no complications. This should be clear from the following Lorentzian analysis.

Note that $\omega(r)$ is defined everywhere on the slice by (6.11), but ω generally has no monotonicity properties away from U_η. This does not matter to the argument, however. Thus, what we now to show is that for any extremal or generic trapped surface, there is some η on which σ is maximal or minimal on Σ_η.

Let us now first assume that σ is extremal. That means that the area of σ is stationary under all variations, so for every choice of η, σ is either maximal, minimal, or a saddle in Σ_η. Being a saddle is clearly a fine-tuned case, and it can always be avoided. To see this, note that when the boost angle η approaches $+\infty$ or $-\infty$, U_η gets closer and closer to one of the two the null congruences fired off σ. By the focusing theorem, the area of these congruences are shrinking both to the future and past, so once $|\eta|$ is large enough, σ will be a locally maximal surface on U_η. In the fine tuned case where the congruences fired from σ are stationary, we are just locally in a standard spherical black hole, in which case we can just pick one of the other slices, where we know that σ is minimal or maximal.

Next, assume that σ is a trapped or antitrapped surface, meaning that the two future null expansions have the same sign, with both strictly nonzero. Fixing an arbitrary zero-point of the boost angle η, let n^a be the future timelike unit normal to $\Sigma_{\eta=0}$, and let r^a be a unit normal to σ that is tangent to $\Sigma_{\eta=0}$. Two independent future directed null normals to σ can be taken to be

$$k^a_\pm = \frac{1}{\sqrt{2}}(n^a \pm r^a), \tag{6.31}$$

where k^a_+ conventionally defines "outwards". As shown in the appendix of [58], with this normalization of the null normals, the mean curvature $H_0[\sigma]$ of σ within $\Sigma_{\eta=0}$ reads

$$H_0[\sigma] = D_a r^a = \frac{1}{\sqrt{2}}(\theta_+ - \theta_-), \tag{6.32}$$

where the null expansions are given by $\theta_\pm = (g^{ab} + 2k_+^{(a}k_-^{b)})\nabla_a k_b$. Consider now boosting our slice to Σ_η. Since n^a, r^a form an orthonormal basis of the normal bundle of σ, they transform in the canonical way under a Lorentz boost. It is easily seen that the null normals canonically normalized with respect to the boosted n^a, r^a are

$$k^a_{\pm,\eta} = e^{\pm\eta}k^a_\pm, \tag{6.33}$$

so the mean curvature of σ within Σ_η is

$$H_\eta[\sigma] = \frac{1}{\sqrt{2}}(e^\eta\theta_+ - e^{-\eta}\theta_-), \tag{6.34}$$

where θ_+, θ_- still are the expansions at $\eta = 0$. Because θ_+, θ_- are nonzero and have the same sign, we can always take $\eta = \frac{1}{2}\log(\theta_-/\theta_+)$, giving $H_\eta[\sigma] = 0$. Hence, locally there is always a $K^a{}_a = 0$ slice on which σ is stationary. Assuming we do not have the fine tuned situation where this surface is a saddle, we see that σ posses a slice one which the left and right sides are independent – at least with respect to spherical perturbations.

So in total, if D_A and D_B are two spacelike separated causal diamonds whose edges are spacelike to a common extremal or generic trapped surface, then we expect that these diamonds are independent, since they are contained in the domains of dependence of the regions Σ_R and Σ_L that we expect to be independent – see Fig. 6.12.[27]

[27] Assuming that the edges are not parametrically close to null separated, so that the splitting region is

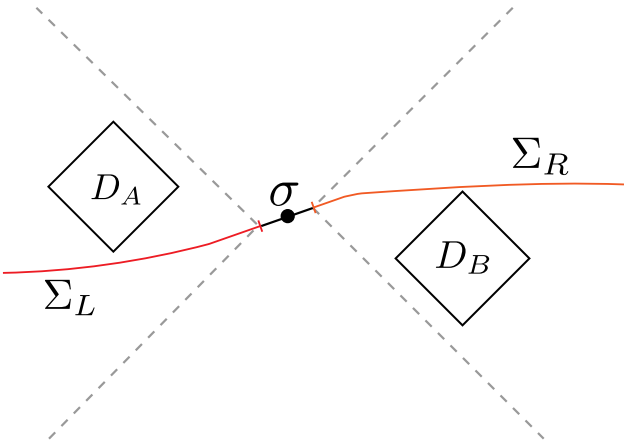


Figure 6.12: Two causal diamonds separated by a trapped or extremal surface σ.

6.4.2 A Lorentzian analysis

It is illustrative to reach the same conclusion using a spacetime approach. Let us now do this, at the same type generalizing to also allow planar or hyperbolic symmetry, in addition to spherical. This will also let us show that arbitrarily large energy densities near an extremal surface make arbitrarily small contributions to the ADM mass.

Let (N, g_{ab}) be a $(d+1)$-dimensional spacetime with one of these three types of symmetries. This means that N can be foliated by a two-parameter family of codimension-2 surfaces σ that have the intrinsic metric of a sphere, a plane, or the hyperbolic plane (or quotients thereof). We can pick a double null gauge

$$\mathrm{d}s^2 = -2e^{-f(x^+,x^-)}\mathrm{d}x^+\mathrm{d}x^- + r(x^+,r^-)^2\mathrm{d}\Sigma_k^2, \tag{6.35}$$

where $\mathrm{d}\Sigma_k^2$ is the metric of the unit sphere, plane, and hyperbolic plane for $k = -1, 0, 1$, respectively (or quotients thereof). For the leaves of our symmetric foliation, given by constant x^+, x^-, we now define the following function[28]

$$\mu(x^+,x^-) = kr^{d-2} + \frac{r^d}{L^2} - \frac{2\theta_+\theta_-}{k_+\cdot k_-(d-1)^2}, \tag{6.36}$$

where k_+^a, k_-^a are any two independent future-directed null normals to the surface, and θ_+, θ_- the corresponding null expansions. The quantity on the RHS is covariantly defined, since the area radius r can be viewed as a coordinate independent scalar on spacetime.[29] μ is a spacetime analog of ω, and it has some special properties. First, in AAdS or AF spacetimes,

parametrically small.

[28] μ is a generalization of the Lorentzian Hawking mass [62–65] to $d \neq 3$, in the special case of spatial symmetries. Without symmetries, the proper definition of the Hawking mass is unknown for $d \neq 3$.

[29] This is not strictly true in planar symmetry. In this case, there is an ambiguity in an overall scaling of r, reflective of the fact that there is no canonical conformal frame for the boundary of AdS, when the boundary is conformal to Minkowski. In this case, there is also a scaling ambiguity in the mass. But this ambiguity can be fixed in some particular spacetime.

it can be shown to reduce to the mass at spatial infinity:[30]

$$M = \frac{(d-1)\text{Vol}[\Sigma_k]}{16\pi G_N}\mu|_{r=\infty}. \tag{6.37}$$

Second, using the Einstein equations, it was shown in [76] to satisfy

$$\partial_\pm \mu = \frac{2e^f r^d}{(d-1)^2}(T_{+-}\theta_\pm - \theta_\mp T_{\pm\pm}). \tag{6.38}$$

Let now X^a be any spacelike vector pointing outwards ($X_a k_+^a \geq 0$), and assume the dominant energy condition (DEC), which says that

$$T_{ab}U^a V^b \geq 0 \quad \forall \text{ timelike } U^a, V^a, \tag{6.39}$$

and which implies the WEC. The DEC implies that $T_{\pm\pm} \geq 0, T_{+-} \geq 0$. Thus, in a "normal" region, where $\theta_+ \geq 0, \theta_- \leq 0$, μ is monotonically non-decreasing in any outwards spacelike direction:

$$X^a \nabla_a \mu \geq 0 \qquad \text{when} \quad \theta_+ \geq 0, \quad \theta_- \leq 0. \tag{6.40}$$

Next, in an "anti-normal" region, we have monotonicity in the inwards direction instead:

$$X^a \nabla_a \mu \leq 0 \qquad \text{when} \quad \theta_+ \leq 0, \quad \theta_- \geq 0. \tag{6.41}$$

This is the spacetime analog of monotonicity of ω, and it constitutes an obstruction to subregion independence (if we break the DEC, this obstruction only becomes weaker). However, we see that if σ is an extremal surface that separates a normal and an anti-normal region, then the insertion of matter causes μ to increase in opposite directions on opposite sides of σ, and so we regain independence between the opposite sides. If we instead find ourselves in an (anti)trapped region of spacetime, μ is no longer monotonic, and by the appropriate choice of turning on either T_{++} or T_{--}, we can push μ up or down as we are moving in any fixed spacelike direction.

We also see from (6.37) and (6.38) that the contribution to the asymptotic mass is given by a integral of $T_{++}\theta_-, T_{--}\theta_+, T_{+-}\theta_-, T_{+-}\theta_+$ weighted by positive factors that are bounded in a neighbourhood of an extremal surface. Thus, arbitrarily large energy densities make arbitrarily small contributions to the ADM mass, provided they are localized to an extremal surface. Similarly, a marginally trapped surface (say, $\theta_+ = 0$) can support modes with very large T_{--} at low cost to the ADM mass, provided we do not make T_{+-} large as well.

6.4.3 No symmetries

We now want to understand what happens if our spacetime does not have any symmetries. Furthermore, even if our spacetime has symmetries, we might want to consider perturbations or surfaces that break the symmetries of the spacetime. A prototypical example of the latter occurs in AdS/CFT, where we might consider two regions A and B separated by a region containing a boundary anchored extremal surface, as illustrated in Fig. 6.13.

[30] Provided matter fields fall off fast enough. The important thing is that μ is a spacetime function with certain monotonicities that sometimes act as an obstruction to independence.

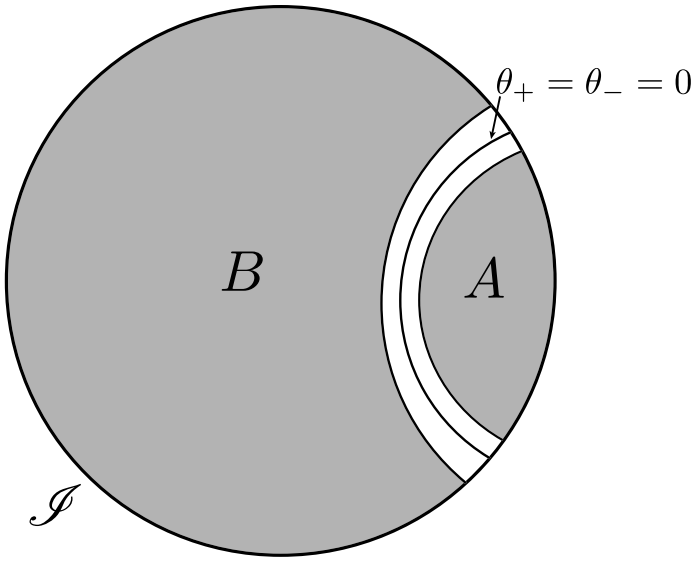


Figure 6.13: A timeslice Σ of an asymptotically AdS spacetime containing two subregions A and B that are separated by a set containing an HRT surface anchored to the conformal boundary $\mathscr{I}$.

To deal with the case of no symmetry, we need an appropriate generalization of ω. To the author's knowledge, the appropriate generalization is known only in four spacetime dimensions, and so we will only treat this case. However, based on expectations from AdS/CFT and the fact that the spherically symmetric considerations work for all $d \geq 2$, we expect a similar story holds in other dimensions, but we do not know what is the right tool to use.

Let us thus assume four spacetime dimensions. If (Σ, S) is some initial dataset with a spacelike two-dimensional surface σ, we define

$$\omega[\sigma] = \frac{1}{16\pi}\sqrt{\frac{\text{Area}[\sigma]}{4\pi}}\int_\sigma \left[2\mathfrak{R} - H[\sigma]^2 + \frac{4}{L^2}\right], \tag{6.42}$$

where the integral is taken in the induced volume form on σ, and where $\mathfrak{R}$ is the Ricci scalar for the induced metric on σ. Consider now a one-parameter family of surfaces σ_τ, where increasing τ corresponds to flowing the surfaces along the vector field

$$v^a = \frac{1}{H[\sigma_\tau]} r^a_\tau, \tag{6.43}$$

with r^a_τ being a unit normal to Σ in σ. Thus, our one-parameter family of surfaces correspond to a flow where the velocity of the flow at $p \in \sigma_\tau$ is set by the inverse of the mean curvature of σ_τ at p. This flow is known as inverse mean curvature (IMC) flow,[31] and it can be shown

[31] In four spacetime dimensions, a Lorentzian version of IMC flow also exists [65, 84, 85], but is less understood. It generalizes the monotonicities (6.40) and (6.41) to cases with no symmetries.

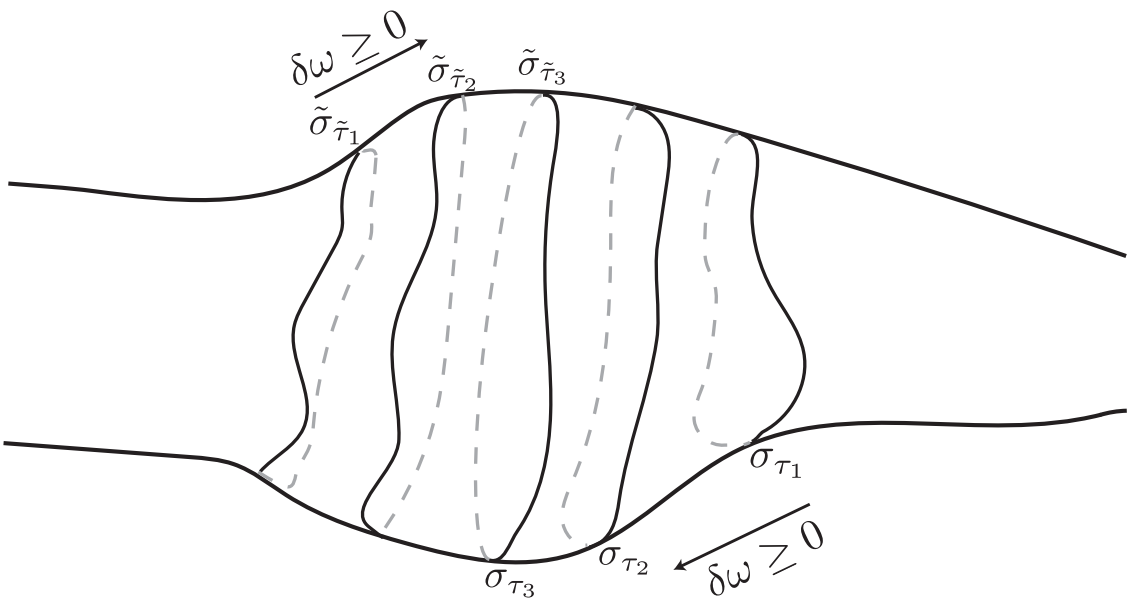


Figure 6.14: Two IMC flows σ_τ and $\tilde{\sigma}_{\tilde{\tau}}$ accumulating at a maximal surface $\tilde{\sigma}_{\tilde{\tau}_3} = \sigma_{\tau_3}$.

that on a slice with $K^a{}_a = 0$, assuming the WEC, we have [59]

$$\frac{\mathrm{d}}{\mathrm{d}\tau}\mathrm{Area}[\sigma_\tau] \geq 0, \tag{6.44}$$

$$\frac{\mathrm{d}}{\mathrm{d}\tau}\omega[\sigma_\tau] \geq 0. \tag{6.45}$$

This is the generalization of $\omega'(r) \geq 0$. Minimal and maximal area surfaces are special locations where the flow terminates. See [86] for a review of these facts for compact surfaces, and for the proof that (6.44) and (6.45) remains true also for boundary anchored surfaces in asymptotically AdS_4 spacetimes (ω is finite even when σ_τ is boundary anchored [86]).

Consider a surface that is a minimal or maximal surface on a $K^a{}_a = 0$ slice. Then, as illustrated in Fig. 6.14, IMC flow goes in opposite directions, so again we have a mechanism to bypass the monotonicity of $\omega[\sigma_\tau]$. Positive energy density can now likely screen positive energy density. Of course, to show this rigorously is challenging,[32] but it would not be surprising if it is true. It would fit perfectly with what we know about AdS/CFT, and it is plausible that the monotonicity of ω is the sole obstruction to subregion independence in gravity, since charges other than the mass tend to not have a preferred sign. In fact, in various gluing results, there is typically only a finite-dimensional space of obstructions to gluing, corresponding to a set of charges that must match or have a certain relationship [79, 87]. In four spacetime dimensions, this is a set of 10 charges, and the mass is the only one with a preferred sign.[33] Thus, we conjecture the following:

[32]Note that IMC flow can have singularities. In this case, a weak version of the flow can be defined. In this case, when a surface becomes singular, it makes a jump to a surface with larger area and Hawking mass. However, since we are trying to screen perturbative excitations by adding matter near the minimal/maximal surface at which we start the flow, should be able to screen the perturbation before we reach a jump, since the jump-time is non-perturbative, so it will not get radically affected by our perturbations.

[33]The other charges corresponds to linear momentum, angular momentum, and center of mass.

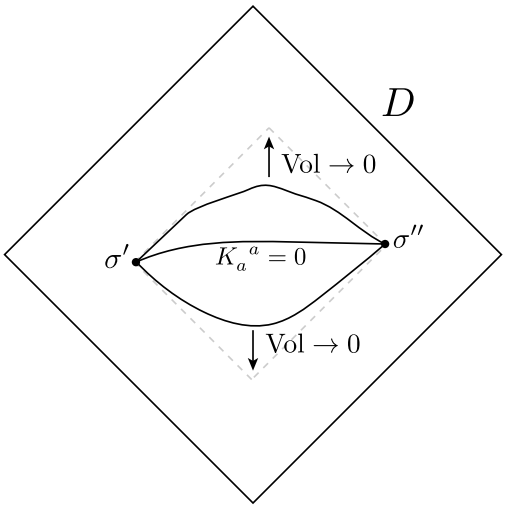


Figure 6.15: A small domain of dependence D. For any two spacelike separated surfaces σ', σ'', we expect there to be a maximal volume slice Σ with $\partial\Sigma = \sigma' \cup \sigma''$.

Conjecture. *Let (Σ, S) be a smooth d-dimensional initial dataset in Einstein gravity minimally coupled to matter, with $K_a{}^a = 0$, and $d \geq 3$. Let A, B be two closed disjoint subregions such that $\Sigma - A - B$ contains a locally minimal or maximal surface that is homologous to ∂A and ∂B. For any sufficiently small deformations $\delta S|_A$ and $\delta S|_B$, there exists a small extension δS of $\delta S|_A \cup \delta S|_B$.*

We are deliberately vague about what we mean with small here. We might either consider formal perturbative solutions, or perhaps better, we could try to require that with an appropriate choice of C^k or Sobolev norm defined by the background metric, a small extension always exists once the norm of $\delta S|_A \cup \delta S|_B$ is sufficiently small. Note that the above might also be true for $d = 2$ if matter is included, but for $d = 2$ vacuum gravity, we saw that we were forced to change the topology of C to match perturbations made in A, B, so there was no sense in which the perturbation was small.

Next, we again really want to talk about surfaces that are trapped or extremal, rather that maximal or minimal on a $K_a{}^a = 0$ slice. By the same logic as in the spherically symmetric discussion, we thus want that every extremal and generic trapped surface σ is minimal or maximal on some slice Σ that satisfies $K_a{}^a = 0$ locally in a neighbourhood of σ. While we will not attempt a proof, this appears very likely to be true. Consider a region of spacetime D corresponding to a domain of dependence containing σ. Let us shrink D enough so that its closure is compact, so that the null generators of its boundary do not terminate on singularities or future infinity. Then for any two spacelike separated surfaces σ', σ'' in D, there should exist a maximal volume hypersurface Σ bounded by σ', σ''. See Fig. 6.15. This makes sense, since if Σ is deformed towards the future or past null congruences fired from σ', σ'', its volume goes to zero. Provided we pick D small enough so that it does not contain a portion of a future/past infinity where volume of space could be forever expanding, like future infinity of de Sitter, we should not be able to make the volume of Σ arbitrary large. Thus, there ought to be a maximal volume slice bounded by σ', σ''. This surface is necessarily a $K_a{}^a = 0$ slice, and so in D we ought to have a $K_a{}^a = 0$ slice for every choice of two spacelike

separated surfaces σ', σ''. Allowing these surfaces to vary, this likely provides more than enough freedom for finding such a slice that contains σ, again suggesting that the setup described by Fig. 6.12 is true without spherical symmetry as well.

6.5 de Sitter Rigidity and Area Bounds

In our discussion of pure AdS and Minkowski, we saw that these spacetimes were very rigid. No deformation in the interior of the spacetime can be made without it altering the geometry at infinity. Can some form of rigidity statement be made for dS? A natural question to ask is the following: is it possible to deform the initial data in one static patch without altering the other? At the level of spherical symmetry, we saw that the answer was no at first order in perturbation theory. We will now prove a non-linear version of this statement. However, as we discuss below, once we break spherical symmetry, rigidity appears to no longer hold, unlike the case of Minkowski and AdS. This likely gives rise to large class of spacetimes that look identical to dS in a single static patch.

We prove the following[34]

Theorem 11. *Let (Σ, S) be a regular spherically symmetric initial dataset for the Einstein equations with positive cosmological constant, $K_a{}^a = 0$, and satisfying the WEC. If Σ has the topology of a hemisphere (i.e. a ball) of dimension $d \geq 2$, then every sphere σ in Σ has an area radius r satisfying*

$$r \leq L_{\text{dS}}. \tag{6.46}$$

Equality is achieved if and only if $K_{ab} = 0$, with h_{ab} the metric of a round sphere. Next, if σ_* is a locally maximal sphere of radius r_*, then

$$r_* \geq \sqrt{\frac{d-2}{d}} L_{\text{dS}} = r_{\text{Nariai}}. \tag{6.47}$$

Before the proof, let us make a few comments. First, at leading order in perturbation theory around pure dS, the upper bound (6.46) is just a special case of the first law [88]. Next, for $d = 3$, the upper bound (6.46), which is equivalent to $A(\sigma) \leq 4\pi L_{\text{dS}}^2$, was proven without spherical symmetry in [89], given certain other assumptions (see also [90]). Second, extremal surfaces are always stationary on any slice, so for an extremal surface X in a spherical asymptotically dS spacetime with a simply connected Cauchy slice, we expect the area bounds (6.46), (6.47) to apply to X, provided X is a cosmological horizon type surface – i.e. it is maximal rather than minimal/a saddle on Σ. Third, we are not aware if a bound like (6.47), which lower bounds the area of cosmological horizon-type surfaces in terms of the event horizon of the Nariai black hole, has been shown before (beyond the dS-Schwarzschild family). Finally, the result shows that a spacetime with the following properties cannot exist when the WEC holds: (1) σ is a sphere that splits a Cauchy slice in two, (2) the wedge D_R of points right-spacelike to σ looks identically like a static patch of de Sitter, and (3) the wedge of D_L points left-spacelike to σ has a Cauchy slice Σ with $K^a{}_a = 0$. See Fig. 6.16. Now to the proof:

[34] The author is grateful to Matt Headrick for proposing that Nariai might set a lower area bound, and for extensive discussions that lead directly to this result.

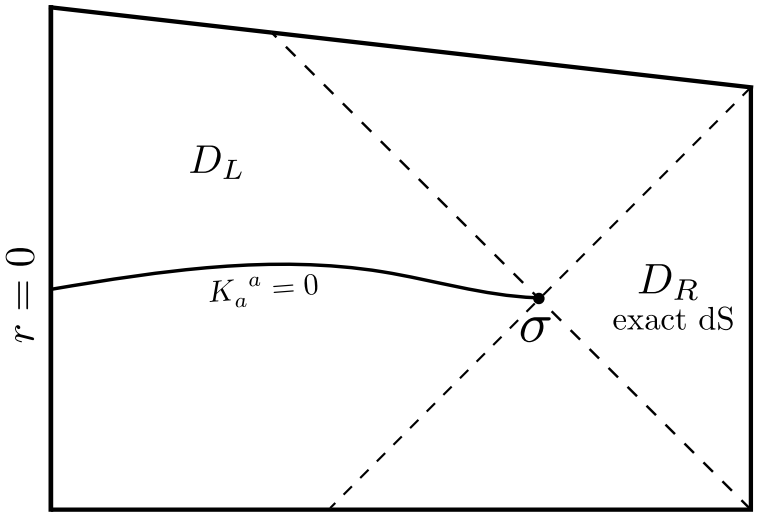


Figure 6.16: A hypothetical spherically symmetric spacetime that is ruled out by Theorem 11 when the WEC holds.

Proof. By the fact that Σ is spacelike, we have

$$1 - \frac{r^2}{L_{\rm dS}^2} - \frac{\omega(r)}{r^{d-2}} \geq 0, \tag{6.48}$$

as discussed in Sec. 6.3.1. Thus, if we can show that $\omega(r) \geq 0$ on every patch of Σ, then we get that $r \leq L_{\rm dS}$. Let us now show this. The Hamiltonian constraint together with $K_a{}^a = 0$ implies that $\omega'(r) \geq 0$. Furthermore, by assumption, Σ contains the point $r = 0$, where we must have $\omega(0) = 0$. Hence, integrating (6.29) outwards, we find that $\omega \geq 0$ on the patch containing $r = 0$. If one patch covers Σ, we are done. Thus, assume that Σ is covered by multiple patches. Then the $r = 0$ patch is separated from the next patch by either a maximal surface or a saddle – see Fig. 6.17. If it is a saddle, it does not alter the monotonicity properties of ω as we move from $r = 0$ towards $\partial\Sigma$, so ω remains positive on the next patch. Thus, assume we have a maximal surface instead. Consider first integrating ω from $r = 0$ to the first maximal surface, with radius $r_{\rm max}$. We already showed that $\omega \geq 0$ there, so we find that $r_{\rm max} \leq L_{\rm dS}$. Transitioning to the next patch, r and ω must now be decreasing as we move towards $\partial\Sigma$. Either we hit $\partial\Sigma$ and we are done since then $r|_{\partial\Sigma} < r_{\rm max} \leq L_{\rm dS}$, or we hit a minimal surface with radius $r = r_{\rm min} < r_{\rm max} \leq L_{\rm dS}$ (again, we can hit a saddle, but nothing interesting happens at these). By stationarity we get

$$\frac{\omega(r_{\rm min})}{r_{\rm min}^{d-1}} = 1 - \frac{r_{\rm min}^2}{L_{\rm dS}^2} > 1 - \frac{r_{\rm max}^2}{L_{\rm dS}^2} \geq 0. \tag{6.49}$$

Repeating this exact argument as we move past any number of minimal, maximal, or saddle surfaces, we find that $\omega \geq 0$ everywhere, showing that $r \leq L_{\rm dS}$. Next, we only find a sphere with $r = L_{\rm dS}$ if Σ a subset of exact de Sitter. This is seen by the fact that chain of potential equalities above is broken if we encounter matter anywhere. This proves rigidity.

To prove the lower bound, note that approaching a stationary surface implies that $1/B(r)$ is approaching 0 from a positive value. The surface being maximal with radius $r = r_*$ implies

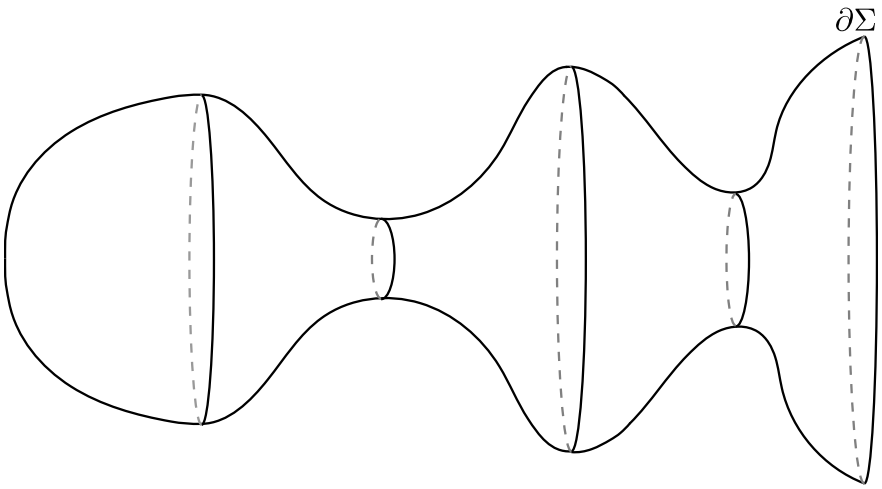


Figure 6.17: A spherically symmetric spatial manifold with a smooth $r = 0$ and a locally maximal boundary. When $K_a{}^a = 0$, the WEC implies that all spheres satisfy $r \leq L_{\rm dS}$

that we are approaching r_* from a smaller value of r. Thus, $\frac{\mathrm{d}}{\mathrm{d}r}B^{-1}|_{r=r_*} \leq 0$, giving that

$$\begin{aligned} 0 &\geq r_* \frac{\mathrm{d}}{\mathrm{d}r}\left[1 - \frac{r^2}{L_{\rm dS}^2} - \frac{\omega(r)}{r^{d-2}}\right]\Big|_{r=r_*} \\ &= -\frac{2r_*^2}{L_{\rm dS}^2} - \frac{\omega'(r_*)}{r_*^{d-3}} + (d-2)\frac{\omega(r_*)}{r_*^{d-2}} \\ &= -\frac{dr_*^2}{L_{\rm dS}^2} + d - 2 - \frac{\omega'(r_*)}{r_*^{d-3}}. \end{aligned} \tag{6.50}$$

Using that $\omega' \geq 0$, this then gives the lower bound on r_*. □

Let us finally discuss the case without spherical symmetry, which turns out to be interesting. It would be tempting to conjecture the following: if Σ is a $K_a{}^a = 0$ slice where $\partial\Sigma$ is an extremal surface, then $A[\partial\Sigma] \leq A[\partial\Sigma_{\rm dS}]$, with equality only in the case of pure dS (assuming the WEC). While the area inequality might hold, the rigidity part, i.e. equality only in pure dS, is overwhelmingly likely to be false for $d \geq 3$. The rigidity part is closely related to a former conjecture by Min-Oo [91], who conjectured the following: assume h_{ab} is a metric on a hemisphere Σ with (1) a Ricci scalar lower bound $\mathcal{R} \geq d(d-1)L_{\rm dS}^{-2}$,[35] and (2) with the intrinsic and extrinsic geometry of $\partial\Sigma$ matching that of the boundary of a round hemisphere of radius $L_{\rm dS}$. Then Σ must be the round hemisphere.[36] While according to [92] this conjecture was proven in [93] for $d = 2$, after being open for 16 years, Min-Oo's conjecture was disproven in [92] for all $d \geq 3$. They showed that there are Riemannian metrics distinct from the canonical round sphere metric that nevertheless looks identical to it in a neighbourhood around $\partial\Sigma$, for any $d \geq 3$. Several properties of these solutions are known. They can be constructed to have arbitrarily large volume [94],[37] they always have a minimal surface [96],

[35]This is the lower bound one would get from the constraints, assuming the WEC and $K_a{}^a = 0$.

[36]With hemisphere, we mean that $\partial\Sigma$ is totally geodesic in Σ – it is really a half sphere, not just a cap.

[37]None of the solutions with volume greater than that of the static patch lie perturbatively close to the

and they can have a wide range of topologies [94, 97]. Furthermore, for h_{ab} sufficiently close to pure dS, the metric must disagree with the round metric somewhere in the band [98]

$$\sqrt{\frac{d-1}{d+3}} L_{\rm dS} < r < L_{\rm dS}, \tag{6.51}$$

so these deformations cannot localize in a tiny cap. If any of these metrics can be realized as a solution of the constraints, either with some choice of $K_{ab} \neq 0$ and/or with some choice of matter, then the rigidity part of the upper bound in Theorem 11 cannot be true without spherical symmetry (when $d > 2$). Thus, in four spacetime dimensions and higher, when we break symmetries there might exist a large flexibility in changing the initial data in only one static patch. The behavior of these solutions are reminiscent of another example we have already discussed: black holes in GR with a negative cosmological constant and three spacetime dimensions. In this case there is an infinite number of one-sided black holes that look identical to the BTZ black hole in the exterior(s). However, they all break spherical symmetry in the interior, and none are close to a spherically symmetric metric. Like black holes, the de Sitter metrics discussed above also have minimal surfaces, so it seems likely that they will be black holes as well. It would be interesting if these could be used to give a semiclassical counting interpretation of the area of the cosmological horizon using the Euclidean path integral approach of [99, 100].

6.6 Discussion

In this work, we have given a simple natural definition of subregion independence in classical gravity. We argued that extremal surfaces, generic trapped surfaces, and background distributions of matter are structures that enable subregion independence, i.e. independent initial data perturbations. For extremal surfaces X, we saw that excitations added on the opposite side of X contribute with opposite sign to any given asymptotic mass. This enables positive energy densities to terminate the gravitational fields sourced by positive energy densities on the other side of X, providing a simple physical picture for why an extremal surface is a good location to separate independent subregions.

We now discuss some further implications of our perspective.

6.6.1 The semiclassical case and quantum extremal surfaces

Consider studying semiclassical gravity, where we couple the expectation value of the stress tensor of quantum matter to the classical Einstein equation. In this case, the constraint equations are unchanged, so our analysis is still relevant, albeit incomplete. The matter source is now more exotic, since classical energy conditions are violated. However, in our case, the classical energy conditions were an obstruction to independence, so we expect that this particular effect pushes subregions to a greater tendency for independence. Nevertheless, while quantum fields typically do not have local energy conditions, they often have global

static patch, since the results of [95] imply that WEC-respecting metrics C^2-close to the static patch must have smaller volume.

ones, so we do not expect that the problem is trivial – the Hawking mass in a normal region still ought to have certain monotonicity properties over sufficiently large distances.

Next, while our analysis does not directly say anything about general QESes [7], our results are informative in many particular cases. We might for example find that QESes themselves are classically (anti)trapped, so that we still avoid monotonicity of the Hawking mass, boding well for the perturbative independence of the complement wedges of the QES. Next, if a QES is a Planckian distance away from a classical extremal or (anti)trapped surface, we can tell a similar story, since we always want to draw some splitting region around the QES, which then contain a classically trapped or extremal surface. However, we leave a more careful study of QESs to the future.

6.6.2 Islands in massless gravity

The new QESs discovered after the Page time [41, 42], giving rise to the island phenomenon, are at the heart of recent breakthroughs on the information paradox [41, 42, 99, 101]. Islands are regions of spacetime that furnish entanglement wedges for the Hawking radiation after the Page time. They have the special property that they are compact, not reaching any asymptotic boundaries. Since the island is supposed to be encoded in the radiation, acting with operators in the island should correspond to acting with operators on the radiation.

In [40] it was argued that this behavior is inconsistent with having massless gravity. An important part of their argument was the following claim: any excitation localized to the island must have compact support, and by the Heisenberg uncertainty principle it should have finite energy, thus altering the ADM energy. However, as should be clear by now, non-linearities of the Einstein equations make this is false at the classical level. While turning on excitations adds positive energy density locally, this does not imply that the ADM energy is changed. It is true for perturbations around Minkowski space or AdS, but these examples turn out to not be good analogies for the general case. As discussed in Sec. 6.4.1, in trapped regions of spacetime, ingoing and outgoing modes can be added so as to affect the the ADM mass with either sign. In the case of the evaporating black holes studied in [41, 42], the island is in a classically trapped region of spacetime, so there is no obstruction to implementing large classes of localized perturbations of the constraints with no influence on the geometry outside the island.

Another part of the argument in [40] was the closely related claim that diffeomorphism invariant operators that probe local physics in the island must be dressed to the boundary, and so they must have non-zero commutators with operators in the complement entanglement wedge, leading to a contradiction with subregion-subregion duality. However, at the level of perturbative quantization in $\sqrt{G_N}$ around some background, there does exist localized operators that are not dressed to the boundary [22–28]. Operators like this were constructed in the CFT by [29, 30], and they argued this relieves the tension between islands and massless gravity. We agree, and will add a few additional remarks.

It was suggested in [29, 30] that one might need to dress operators to the Hawking radiation. While perhaps possible, this can be avoided for the evaporating black holes of [41, 42]. To begin with, let Φ_0 be some scalar field that is part of the background we are

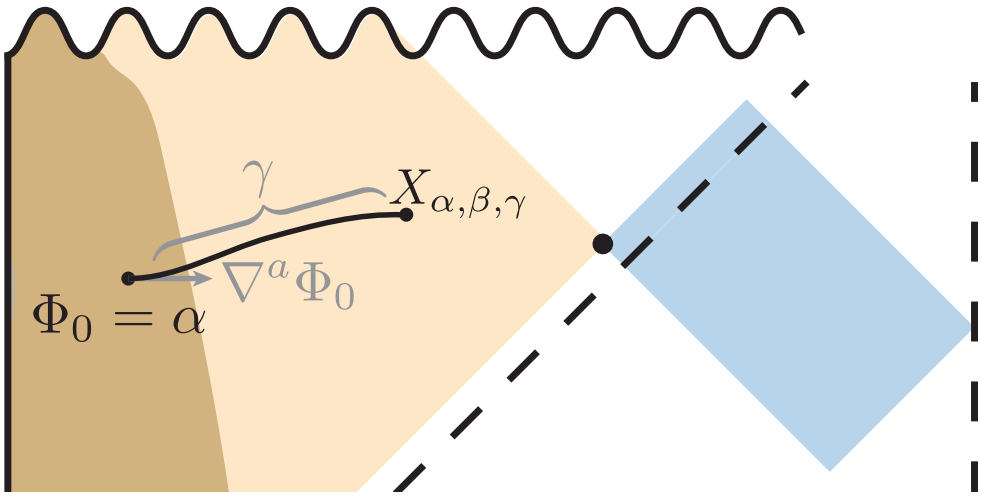


Figure 6.18: A diffeomorphism invariant operator localized to the island, dressed to matter that formed the black hole, and which probes the "barren" part of the island without becoming highly delocalized as $G_N \to 0$.

considering, and let ϕ be a dynamical quantum field.[38] Then

$$O(\alpha) = \int \mathrm{d}^{d+1}x\sqrt{-g}\delta\left(\Phi_0(x) - \alpha\right)\phi(x) \tag{6.52}$$

is a one-paramater family of diffeomorphism invariant observables. In the quantum case, we should smear this over a window of α-values to get a proper operator, but as long as the gradient of Φ_0 is not extremely small (i.e. scaling with G_N to a positive power), this gives an approximately localized operator on appropriate backgrounds. The existence of single-integral observables like (6.52) was acknowledged in [40]. However, they argued that single-integral observables cannot probe large parts of the island. This is based on the fact that once we reach the Page time, the black hole interior volume (on some nice slice) has grown approximately linearly for a time of order $t \sim \mathcal{O}(G_N^{-1})$, so unless the black hole has been constantly fed matter, there are large regions in the black hole where there is no matter, or where the matter is extremely dilute. So if we try to get the delta function in $O(\alpha)$ to "click" somewhere in the dilute region, once we smear α a tiny bit, the operator strongly delocalizes. There are however other better operators that do not have this problem. First, note that the matter that collapsed to form the black hole never dilutes, and it is present in the island.[39] So for some windows of α, we could get operators localized inside the matter distribution that formed the original black hole. But we can do better. Let us now consider forming the following three-parameter family of diffeomorphism invariant observables[40]

$$O(\alpha,\beta,\gamma) = \int \mathrm{d}^{d+1}x\sqrt{-g}\delta\left(\Phi_0(x) - \alpha\right)\delta\left(\nabla_a\Phi_0(x)\nabla^a\Phi_0(x) - \beta\right)\phi(X_{\alpha,\beta,\gamma}(x)) \tag{6.53}$$

[38] Φ_0 could also be quantum field with a non-zero VEV.

[39] Alternatively, if we instead evaporate a past-eternal black hole in AdS, the same goes through for the shell of matter that falls into the black hole due to the sudden coupling of the CFT to the reservoir where we dump the Hawking radiation.

[40] Here g_{ab}, ∇_a are the full metric and connections, so these should be expanded in a perturbative series. Similar for $X_{\alpha,\beta,\gamma}$.

where the $X_{\alpha,\beta,\gamma}(x)$ is the point obtained when firing a geodesic from x along the direction of $t^a = \nabla^a \Phi_0$ and following it for a proper distance/time γ. See Fig. 6.18. Since we get to fix the sign of β, we also get to fix the signature of t^a, so we know whether we are using a spacelike or timelike geodesic – i.e. the observable is well defined. To get an observable with the potential to be promoted to a proper operator when quantizing, we could smear over a small window of α, β, γ. If we pick α-window appropriately, then the x-integral can localize within the concentrated matter that formed the black hole. Next, by increasing γ up to values of order $\mathcal{O}(1/G_N)$, we can reach into the "barren" matterless region of the island. We could also reach this region by dressing to matter that falls into the black hole around the Page time, if such matter exist.

All in all, perturbation theory in massless gravity appears to be consistent with compact entanglement wedges. Of course, once we worry about exponential corrections in G_N, things are much more subtle.[41] But once we consider exponential corrections, we should worry about what right we have to talk about concepts like spacetime regions or entanglement wedges, and it is hard to draw strong conclusions either way. But we see no clear sign of inconsistencies between islands and massless gravity at the perturbative level.

6.6.3 Gravitational splitting and local algebras

In this paper, we have argued that classical gravity has a sort of perturbative split property, provided the splitting region is sufficiently generic – i.e. it contains an extremal surface, a non-perturbative amount of matter, or a generic (anti)trapped surface. It would be very interesting to study the perturbative WdW equation to see whether this remains true in the quantum case. Specifically, working around a sufficiently generic background, can we choose the WdW wavefunctional on generic separated spacelike subregions separated by a finite gap independently? [15] showed that this is not true around pure AdS, but there are many hints that this story changes on other backgrounds. If we indeed can choose the WdW wavefunctional independently on different generic regions (up to the local constraints on these regions), this suggest that non-trivial perturbative algebras of localized observables exist for generic compact regions. This would be good news, given the recent interesting developments on algebras and their entropies in gravity [83, 103–115].

Note that we have not directly discussed the relation between independence and localized observables, which is strictly speaking more directly tied to algebras that the states themselves. We now turn to this.

6.6.4 From independence to localized observables

Our notion of independence deals with the structure of the phase space $\mathcal{P}$ of GR in a sufficiently small neighbourhood $U \subset \mathcal{P}$ around a point $\Psi \in \mathcal{P}$, which corresponds to our background. It would be interesting to understand whether the independence of $A, B \subset \Sigma$ means that there exist observables $f_{AC}, f_{BC} : U \mapsto \mathbb{R}$ that are sensitive only to the dynamical fields in the domains of dependence of $A \cup C$ and $B \cup C$, respectively. For example, if B is a

[41] See [102] for an illuminating discussion of the challenges of defining non-perturbatively diffeomorphism invariant observables.

collar around spatial infinity, f_{AC} would correspond to a (perturbatively) diffeomorphism invariant observable that is not dressed to the boundary, and it would Poisson-commute with observables localized to B [26]. It would be natural if independence implied that such functions exist. If they do, they are candidates for localized operators upon quantization. Let us outline a non-rigorous argument that localized observables exist. We do not claim that this completely settles the issue – there might be devils in the details.

Assume that there exist some gauge fixing procedure, so that the coordinate values of $\partial A, \partial B$ are fixed, and so that every point $\delta\Psi \in U$ has a unique value of $\delta S|_A$, $\delta S|_B$ and $\delta S|_C$. These three collections of classical field values on A, B, C are generally dependent on each other. However, if A and B are independent, then $\delta S|_A$ and $\delta S|_B$ can be specified independently. So they provide valid coordinates on U, and we can write a coordinate representation

$$\delta\Psi = \begin{pmatrix} \delta S|_A \\ \delta S|_B \\ \widehat{\delta S}|_C \end{pmatrix}, \tag{6.54}$$

where $\widehat{\delta S}_C$ is a collection of coordinates parametrizing the remaining freedom in C independent of the degrees of freedom in A and B. Clearly, by the constraints, some of the degrees of freedom in C are dependent on the degrees of freedom in $A \cup B$, so $\widehat{\delta S}|_C$ must be a strict subset of $\delta S|_C$.

Now, since $\delta S|_A, \delta S|_B$ can be used as coordinates, the phase space coordinate functions in the A– and B–slots of (6.54) then seem like candidates for f_{AC}, f_{BC}. It might be tempting to say that these functions are sensitive to the dynamical fields in just A or just B, since these functions always return information about the dynamical fields in A or B. However, this is not true. A choice of $\delta S|_C$, influences the possible values of $\delta S|_A$, and thus the possible output of the coordinate functions of the A-slot. Disregarding diffeomorphism invariance for a second, an analogy in a theory with a dynamical field ϕ and two points $x_A \in A, x_C \in C$ would be a function like $f_{AC} = \phi(x_A)\theta(\phi(x_C))$, where θ is the heaviside step function.

Note that if A and B were dependent, this construction would not work. The possible values of $\delta S|_A$ would depend on the state on B, so any function of $\delta\Psi$ that returned $\delta S|_A$ could not be sensitive only to the fields in A, since the possible value of the fields there depends on the fields in B.

6.6.5 The Python's lunch

Consider an asymptotically AdS spacetime, and let R be a spatial subregion of the conformal boundary – or perhaps a complete connected component. Next, assume that R has three extremal surfaces homologous to it: X_{min}, X_{out} and X_{bulge}, as shown in Fig. 6.19 in the case where R is a complete boundary component. X_{min} is the HRT surface, which defines the entanglement wedge. X_{out} is an extremal surface forming a candidate HRT surface, but it has area greater than X_{min}. X_{bulge} lies in between the two, has greater area than either, and it is not minimal on any spatial slice. These three surfaces together form a structure known as a "Python's lunch" – see Fig. 6.19. It was conjectured in [116] that the computational

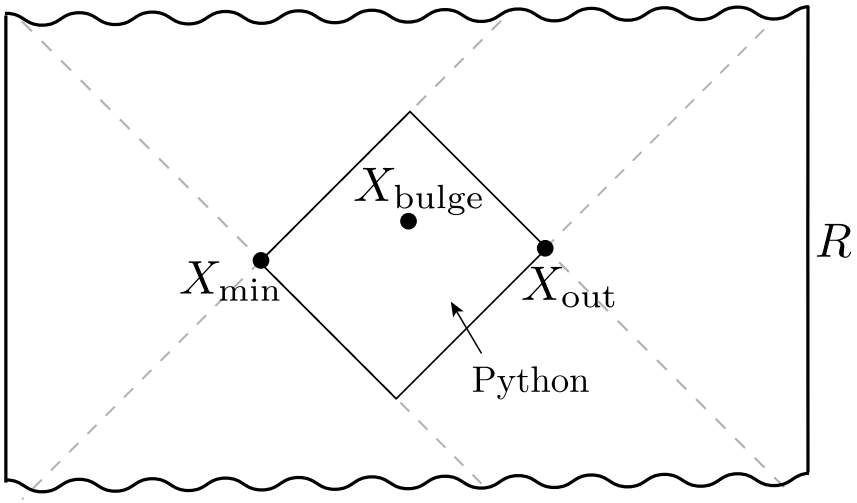


Figure 6.19: A spacetime with a Python's lunch for a complete boundary component R. The wedge to the left and right of X_{min} are the entanglement wedges of the left and the right CFT, respectively.

complexity required to reconstruct operators in the Python's lunch from the CFT scales as

$$\propto \exp\left(\frac{\text{Area}[X_{\text{bulge}}] - \text{Area}[X_{\text{out}}]}{8G_N}\right). \tag{6.55}$$

In the semiclassical case, we replace HRT with QES and Area/$4G_N$ with the generalized entropy. Thus, the reconstruction of operators in the lunch from operators near the boundary is non-perturbative in G_N, and should not be achievable when just accessing the perturbative bulk description (including gravitons). Further evidence for this was found in [117], where they showed that classical boundary sources, together with repeated forwards and background time evolution (i.e. time-folds), cannot be used to expose the lunch region and make its data causally accessible from the boundary.

The findings in this paper play nicely with the Python's conjecture. The bulge provides a surface that can be dressed across, giving it a functional role. It enables perturbations with spatial compact support around the bulge, never disturbing the complement wedges to the lunch, including the exterior of X_{out}, which is the part of the entanglement wedge that is believed to be simple to reconstruct. Furthermore, at least in spherical symmetry, bulge surfaces are supported by matter, which provides additional background structures to anchor diffeomorphism invariant with respect observables to. So it is quite reasonable that the perturbatively quantized theory supports a non-trivial localized algebra of observables inside the Python that commutes with observables in the simple wedge, and thus observables near the boundary, to all orders in perturbation theory in G_N (see [118] for a recent discussion of this algebra).

Acknowledgements

It is a pleasure to thank Chris Akers, Roberto Emparan, Netta Engelhardt, Daniel Harlow, Matt Headrick, Philipp Hoehn, Adam Levine, Martin Sasieta, Gautam Satishchandran, Jon Sorce and Nico Valdes-Meller for illuminating discussions. I also especially want to thank Netta Engelhardt for comments on an earlier draft of the manuscript, Matt Headrick for extensive discussions leading to Theorem 1, and Nico Valdes-Meller for many useful discussions. This research is supported by the John Templeton Foundation via the Black Hole Initiative. The author is grateful to the MIT Center for Theoretical Physics for general support.

References

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

[2] S. Ryu and T. Takayanagi, *Aspects of Holographic Entanglement Entropy*, *JHEP* **0608** (2006) 045, [hep-th/0605073].

[3] V. E. Hubeny, M. Rangamani, and T. Takayanagi, *A Covariant holographic entanglement entropy proposal*, *JHEP* **0707** (2007) 062, [arXiv:0705.0016].

[4] B. Czech, J. L. Karczmarek, F. Nogueira, and M. Van Raamsdonk, *The Gravity Dual of a Density Matrix*, *Class.Quant.Grav.* **29** (2012) 155009, [arXiv:1204.1330].

[5] A. C. Wall, *Maximin Surfaces, and the Strong Subadditivity of the Covariant Holographic Entanglement Entropy*, *Class.Quant.Grav.* **31** (2014), no. 22 225007, [arXiv:1211.3494].

[6] M. Headrick, V. E. Hubeny, A. Lawrence, and M. Rangamani, *Causality & holographic entanglement entropy*, *JHEP* **12** (2014) 162, [arXiv:1408.6300].

[7] N. Engelhardt and A. C. Wall, *Quantum Extremal Surfaces: Holographic Entanglement Entropy beyond the Classical Regime*, *JHEP* **01** (2015) 073, [arXiv:1408.3203].

[8] A. Almheiri, X. Dong, and D. Harlow, *Bulk Locality and Quantum Error Correction in AdS/CFT*, *JHEP* **04** (2015) 163, [arXiv:1411.7041].

[9] 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].

[10] D. L. Jafferis, A. Lewkowycz, J. Maldacena, and S. J. Suh, *Relative entropy equals bulk relative entropy*, arXiv:1512.06431.

[11] D. Harlow, *The Ryu-Takayanagi Formula from Quantum Error Correction*, *Commun. Math. Phys.* **354** (2017), no. 3 865–912, [arXiv:1607.03901].

[12] X. Dong, D. Harlow, and A. C. Wall, *Reconstruction of Bulk Operators within the Entanglement Wedge in Gauge-Gravity Duality*, *Phys. Rev. Lett.* **117** (2016), no. 2 021601, [arXiv:1601.05416].

[13] J. Cotler, P. Hayden, G. Salton, B. Swingle, and M. Walter, *Entanglement Wedge Reconstruction via Universal Recovery Channels*, arXiv:1704.05839.

[14] W. Donnelly and S. B. Giddings, *Observables, gravitational dressing, and obstructions to locality and subsystems*, *Phys. Rev. D* **94** (2016), no. 10 104038, [arXiv:1607.01025].

[15] C. Chowdhury, V. Godet, O. Papadoulaki, and S. Raju, *Holography from the Wheeler-DeWitt equation*, *JHEP* **03** (2022) 019, [arXiv:2107.14802].

[16] V. Balasubramanian, D. Marolf, and M. Rozali, *Information Recovery From Black Holes*, *Gen. Rel. Grav.* **38** (2006) 1529–1536, [hep-th/0604045].

[17] D. Marolf, *Asymptotic flatness, little string theory, and holography*, *JHEP* **03** (2007) 122, [hep-th/0612012].

[18] D. Marolf, *Unitarity and Holography in Gravitational Physics*, *Phys. Rev. D* **79** (2009) 044010, [arXiv:0808.2842].

[19] D. Marolf, *Holography without strings?*, *Class. Quant. Grav.* **31** (2014) 015008, [arXiv:1308.1977].

[20] C. Chowdhury, O. Papadoulaki, and S. Raju, *A physical protocol for observers near the boundary to obtain bulk information in quantum gravity*, *SciPost Phys.* **10** (2021), no. 5 106, [arXiv:2008.01740].

[21] A. Laddha, S. G. Prabhu, S. Raju, and P. Shrivastava, *The Holographic Nature of Null Infinity*, *SciPost Phys.* **10** (2021), no. 2 041, [arXiv:2002.02448].

[22] A. Komar, *Construction of a complete set of independent observables in the general theory of relativity*, *Phys. Rev.* **111** (Aug, 1958) 1182–1187.

[23] P. G. Bergmann and A. B. Komar, *Poisson brackets between locally defined observables in general relativity*, *Phys. Rev. Lett.* **4** (Apr, 1960) 432–433.

[24] S. B. Giddings, D. Marolf, and J. B. Hartle, *Observables in effective gravity*, *Phys. Rev. D* **74** (2006) 064018, [hep-th/0512200].

[25] J. Tambornino, *Relational Observables in Gravity: a Review*, *SIGMA* **8** (2012) 017, [arXiv:1109.0740].

[26] D. Marolf, *Comments on Microcausality, Chaos, and Gravitational Observables*, *Class. Quant. Grav.* **32** (2015), no. 24 245003, [arXiv:1508.00939].

[27] I. Khavkine, *Local and gauge invariant observables in gravity*, *Class. Quant. Grav.* **32** (2015), no. 18 185019, [arXiv:1503.03754].

[28] C. Goeller, P. A. Hoehn, and J. Kirklin, *Diffeomorphism-invariant observables and dynamical frames in gravity: reconciling bulk locality with general covariance*, arXiv:2206.01193.

[29] E. Bahiru, A. Belin, K. Papadodimas, G. Sarosi, and N. Vardian, *State-dressed local operators in the AdS/CFT correspondence*, arXiv:2209.06845.

[30] E. Bahiru, A. Belin, K. Papadodimas, G. Sarosi, and N. Vardian, *Holography and Localization of Information in Quantum Gravity*, arXiv:2301.08753.

[31] R. Schoen and S. T. Yau, *Proof of the positive mass theorem. ii*, *Comm. Math. Phys.* **79** (1981), no. 2 231–260.

[32] E. Witten, *A Simple Proof of the Positive Energy Theorem*, *Commun. Math. Phys.* **80** (1981) 381.

[33] X. Wang, *The Mass of Asymptotically Hyperbolic Manifolds*, *Journal of Differential Geometry* **57** (2001), no. 2 273 – 299.

[34] P. T. Chrusciel and M. Herzlich, *The mass of asymptotically hyperbolic Riemannian manifolds*, *arXiv Mathematics e-prints* (Oct., 2001) math/0110035, [math/0110035].

[35] H. Roos, *Independence of local algebras in quantum field theory*, *Commun. Math. Phys.* **16** (1970) 238–246.

[36] D. Buchholz, *PRODUCT STATES FOR LOCAL ALGEBRAS*, *Commun. Math. Phys.* **36** (1974) 287–304.

[37] S. Doplicher and R. Longo, *Standard and split inclusions of von Neumann algebras*, *Invent. Math.* **75** (1984) 493–536.

[38] D. Buchholz, C. D'Antoni, and K. Fredenhagen, *The universal structure of local algebras*, *Communications in Mathematical Physics* **111** (1987), no. 1 123 – 135.

[39] C. J. Fewster, *The split property for quantum field theories in flat and curved spacetimes*, in *Abhandlungen aus dem Mathematischen Seminar der Universität Hamburg*, vol. 86, pp. 153–175, Springer, 2016.

[40] H. Geng, A. Karch, C. Perez-Pardavila, S. Raju, L. Randall, M. Riojas, and S. Shashi, *Inconsistency of islands in theories with long-range gravity*, *JHEP* **01** (2022) 182, [arXiv:2107.03390].

[41] A. Almheiri, N. Engelhardt, D. Marolf, and H. Maxfield, *The entropy of bulk quantum fields and the entanglement wedge of an evaporating black hole*, *JHEP* **12** (2019) 063, [arXiv:1905.08762].

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

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

[44] Y. Fourès-Bruhat, *Théorème d'existence pour certains systèmes d'équations aux dérivées partielles non linéaires*, *Acta Mathematica* **88** (1952), no. 0 141–225.

[45] Y. Choquet-Bruhat and R. P. Geroch, *Global aspects of the Cauchy problem in general relativity*, *Commun. Math. Phys.* **14** (1969) 329–335.

[46] S. W. Hawking and G. F. R. Ellis, *The large scale stucture of space-time*. Cambridge University Press, Cambridge, England, 1973.

[47] Y. Choquet-Bruhat, *General Relativity and the Einstein Equations*. Oxford Mathematical Monographs. Oxford University Press, United Kingdom, 2009.

[48] W. Donnelly and L. Freidel, *Local subsystems in gauge theory and gravity*, *JHEP* **09** (2016) 102, [arXiv:1601.04744].

[49] R. Beig and P. T. Chruściel, *Shielding linearized gravity*, *Phys. Rev. D* **95** (2017), no. 6 064063, [arXiv:1701.00486].

[50] D. Harlow and H. Ooguri, *Symmetries in quantum field theory and quantum gravity*, *Commun. Math. Phys.* **383** (2021), no. 3 1669–1804, [arXiv:1810.05338].

[51] R. Schoen and S. T. Yau, *Proof that the bondi mass is positive*, *Phys. Rev. Lett.* **48** (Feb, 1982) 369–371.

[52] L. Andersson, M. Cai, and G. J. Galloway, *Rigidity and Positivity of Mass for Asymptotically Hyperbolic Manifolds*, *Annales Henri Poincare* **9** (2008) 1–33, [math/0703259].

[53] S. B. Giddings and A. Kinsella, *Gauge-invariant observables, gravitational dressings, and holography in AdS*, *JHEP* **11** (2018) 074, [arXiv:1802.01602].

[54] S. Giddings and S. Weinberg, *Gauge-invariant observables in gravity and electromagnetism: black hole backgrounds and null dressings*, *Phys. Rev. D* **102** (2020), no. 2 026010, [arXiv:1911.09115].

[55] D. Harlow and J.-q. Wu, *Algebra of diffeomorphism-invariant observables in Jackiw-Teitelboim gravity*, *JHEP* **05** (2022) 097, [arXiv:2108.04841].

[56] S. B. Giddings and J. Perkins, *Perturbative quantum evolution of the gravitational state and dressing in general backgrounds*, arXiv:2209.06836.

[57] P. T. Chruściel, J. Isenberg, and D. Pollack, *Gluing initial data sets for general relativity*, *Phys. Rev. Lett.* **93** (Aug, 2004) 081101.

[58] N. Engelhardt and Å. Folkestad, *General bounds on holographic complexity*, *JHEP* **01** (2022) 040, [arXiv:2109.06883].

[59] R. Geroch, *Energy extraction**, *Annals of the New York Academy of Sciences* **224** (1973), no. 1 108–117.

[60] P. Jang and R. Wald *J. Math. Phys.* **18** (1977) 41.

[61] P. T. Chrusciel and W. Simon, *Towards the classification of static vacuum space-times with negative cosmological constant*, *J. Math. Phys.* **42** (2001) 1779–1817, [gr-qc/0004032].

[62] C. W. Misner and D. H. Sharp, *Relativistic equations for adiabatic, spherically symmetric gravitational collapse*, *Phys. Rev.* **136** (Oct, 1964) B571–B576.

[63] S. W. Hawking, *Gravitational radiation in an expanding universe*, *Journal of Mathematical Physics* **9** (1968), no. 4 598–604, [https://doi.org/10.1063/1.1664615].

[64] S. A. Hayward, *Inequalities relating area, energy, surface gravity and charge of black holes*, *Phys. Rev. Lett.* **81** (1998) 4557–4559, [gr-qc/9807003].

[65] H. Bray, S. Hayward, M. Mars, and W. Simon, *Generalized inverse mean curvature flows in spacetime*, *Commun. Math. Phys.* **272** (2007) 119–138, [gr-qc/0603014].

[66] J. T. Jebsen, *On the general spherically symmetric solutions of Einstein's gravitational equations in vacuo*, *General Relativity and Gravitation* **37** (Dec., 2005) 2253–2259.

[67] G. D. Birkhoff and R. E. Langer, *Relativity and modern physics*. 1923.

[68] N. Voje Johansen and F. Ravndal, *On the discovery of Birkhoff's theorem*, *Gen. Rel. Grav.* **38** (2006) 537–540, [physics/0508163].

[69] K. Schleich and D. M. Witt, *A simple proof of Birkhoff's theorem for cosmological constant*, *J. Math. Phys.* **51** (2010) 112502, [arXiv:0908.4110].

[70] G. Huisken and T. Ilmanen, *The Inverse Mean Curvature Flow and the Riemannian Penrose Inequality*, *Journal of Differential Geometry* **59** (2001), no. 3 353 – 437.

[71] H. L. Bray, *Proof of the Riemannian Penrose Inequality Using the Positive Mass Theorem*, *Journal of Differential Geometry* **59** (2001), no. 2 177 – 267.

[72] H. L. Bray and D. A. Lee, *On the Riemannian Penrose inequality in dimensions less than 8*, *Duke Math. J.* **148** (2009) 81–106, [arXiv:0705.1128].

[73] L. L. de Lima and F. Girao, *Positive mass and penrose type inequalities for asymptotically hyperbolic hypersurfaces*, 2012.

[74] Y. Ge, G. Wang, J. Wu, and C. Xia, *A penrose inequality for graphs over kottler space*, 2013.

[75] V. Husain and S. Singh, *Penrose inequality in anti–de Sitter space*, *Phys. Rev. D* **96** (2017), no. 10 104055, [arXiv:1709.02395].

[76] Å. Folkestad, *Penrose Inequality as a Constraint on the Low Energy Limit of Quantum Gravity*, *Phys. Rev. Lett.* **130** (2023), no. 12 121501, [arXiv:2209.00013].

[77] S. Aminneborg, I. Bengtsson, D. Brill, S. Holst, and P. Peldan, *Black holes and wormholes in (2+1)-dimensions*, *Class. Quant. Grav.* **15** (1998) 627–644, [gr-qc/9707036].

[78] D. Brill, *Black holes and wormholes in (2+1)-dimensions*, *Lect. Notes Phys.* **537** (2000) 143, [gr-qc/9904083].

[79] J. Corvino, *Scalar curvature deformation and a gluing construction for the einstein constraint equations*, *Communications in Mathematical Physics* **214** (2000) 137–189.

[80] J. Corvino and R. M. Schoen, *On the asymptotics for the vacuum Einstein constraint equations*, *J. Diff. Geom.* **73** (2006), no. 2 185–217, [gr-qc/0301071].

[81] P. Hintz, *The linearized Einstein equations with sources*, arXiv:2306.07715.

[82] G. W. Gibbons and S. W. Hawking, *Cosmological event horizons, thermodynamics, and particle creation*, *Phys. Rev. D* **15** (May, 1977) 2738–2751.

[83] V. Chandrasekaran, R. Longo, G. Penington, and E. Witten, *An algebra of observables for de Sitter space*, *JHEP* **02** (2023) 082, [arXiv:2206.10780].

[84] J. Frauendiener, *On the Penrose inequality*, *Phys. Rev. Lett.* **87** (2001) 101101, [gr-qc/0105093].

[85] H. L. Bray and J. L. Jauregui, *Time flat surfaces and the monotonicity of the spacetime Hawking mass*, *Commun. Math. Phys.* **335** (2015), no. 1 285–307, [arXiv:1310.8638].

[86] S. Fischetti and T. Wiseman, *A Bound on Holographic Entanglement Entropy from Inverse Mean Curvature Flow*, *Class. Quant. Grav.* **34** (2017), no. 12 125005, [arXiv:1612.04373].

[87] S. Czimek and I. Rodnianski, *Obstruction-free gluing for the Einstein equations*, arXiv:2210.09663.

[88] G. W. Gibbons and S. W. Hawking, *Action integrals and partition functions in quantum gravity*, *Phys. Rev. D* **15** (1977) 2752–2756.

[89] T. Shiromizu, K.-i. Nakao, H. Kodama, and K.-i. Maeda, *Can large black holes collide in de sitter space-time? an inflationary scenario of an inhomogeneous universe*, *Phys. Rev. D* **47** (Apr, 1993) R3099–R3102.

[90] T. Shiromizu, K. Izumi, K. Lee, and D. Soligon, *Maximum size of black holes in our accelerating Universe*, *Phys. Rev. D* **106** (2022), no. 8 084014, [arXiv:2207.10202].

[91] M. Min-Oo, *Scalar curvature rigidity of certain symmetric spaces*, 2002.

[92] S. Brendle, F. C. Marques, and A. Neves, *Deformations of the hemisphere that increase scalar curvature*, *Inventiones mathematicae* **185** (2011), no. 1 175–197.

[93] V. A. Toponogov, *Evaluation of the length of a closed geodesic on a convex surface*, *Dokl. Akad. Nauk SSSR* **124** (1959) 282–284.

[94] J. Corvino, M. Eichmair, and P. Miao, *Deformation of scalar curvature and volume*, *Mathematische annalen* **357** (2013), no. 2 551–584.

[95] P. Miao and L. fai Tam, *Scalar curvature rigidity with a volume constraint*, 2011.

[96] F. C. Marques and A. Neves, *Rigidity of min-max minimal spheres in three-manifolds*, *Duke Mathematical Journal* **161** (nov, 2012).

[97] P. Sweeney Jr, *New counterexamples to min-oo's conjecture via tunnels*, *arXiv preprint arXiv:2308.03184* (2023).

[98] S. Brendle and F. C. Marques, *Scalar curvature rigidity of geodesic balls in $S\hat{n}$*, *arXiv e-prints* (May, 2010) arXiv:1005.2782, [arXiv:1005.2782].

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

[100] V. Balasubramanian, A. Lawrence, J. M. Magan, and M. Sasieta, *Microscopic origin of the entropy of black holes in general relativity*, arXiv:2212.02447.

[101] A. Almheiri, T. Hartman, J. Maldacena, E. Shaghoulian, and A. Tajdini, *Replica Wormholes and the Entropy of Hawking Radiation*, *JHEP* **05** (2020) 013, [arXiv:1911.12333].

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

[103] S. Leutheusser and H. Liu, *Emergent times in holographic duality*, *Phys. Rev. D* **108** (2023), no. 8 086020, [arXiv:2112.12156].

[104] E. Witten, *Gravity and the crossed product*, *JHEP* **10** (2022) 008, [arXiv:2112.12828].

[105] S. Leutheusser and H. Liu, *Subalgebra-subregion duality: emergence of space and time in holography*, arXiv:2212.13266.

[106] V. Chandrasekaran, G. Penington, and E. Witten, *Large N algebras and generalized entropy*, arXiv:2209.10454.

[107] K. Jensen, J. Sorce, and A. Speranza, *Generalized entropy for general subregions in quantum gravity*, arXiv:2306.01837.

[108] G. Penington and E. Witten, *Algebras and States in JT Gravity*, arXiv:2301.07257.

[109] S. Ali Ahmad and R. Jefferson, *Crossed product algebras and generalized entropy for subregions*, arXiv:2306.07323.

[110] D. K. Kolchmeyer, *von Neumann algebras in JT gravity*, *JHEP* **06** (2023) 067, [arXiv:2303.04701].

[111] M. S. Klinger and R. G. Leigh, *Crossed Products, Extended Phase Spaces and the Resolution of Entanglement Singularities*, arXiv:2306.09314.

[112] E. Gesteau, *Large N von Neumann algebras and the renormalization of Newton's constant*, arXiv:2302.01938.

[113] J. Kudler-Flam, S. Leutheusser, and G. Satishchandran, *Generalized Black Hole Entropy is von Neumann Entropy*, arXiv:2309.15897.

[114] E. Colafranceschi, X. Dong, D. Marolf, and Z. Wang, *Algebras and Hilbert spaces from gravitational path integrals: Understanding Ryu-Takayanagi/HRT as entropy without invoking holography*, arXiv:2310.02189.

[115] E. Witten, *Algebras, Regions, and Observers*, arXiv:2303.02837.

[116] A. R. Brown, H. Gharibyan, G. Penington, and L. Susskind, *The Python's Lunch: geometric obstructions to decoding Hawking radiation*, *JHEP* **08** (2020) 121, [arXiv:1912.00228].

[117] N. Engelhardt, G. Penington, and A. Shahbazi-Moghaddam, *A World without Pythons would be so Simple*, arXiv:2102.07774.

[118] N. Engelhardt and H. Liu, *Algebraic ER=EPR and Complexity Transfer*, arXiv:2311.04281.

Chapter 7

Canonical Purification of Evaporating Black holes

Abstract: We show that the canonical purification of an evaporating black hole after the Page time consists of a short, connected, Lorentzian wormhole between two asymptotic boundaries, one of which is unitarily related to the radiation. This provides a quantitative and general realization of the predictions of ER=EPR in an evaporating black hole after the Page time; this further gives a standard AdS/CFT calculation of the entropy of the radiation (without modifications of the homology constraint). Before the Page time, the canonical purification consists of two disconnected, semiclassical black holes. From this, we construct two bipartite entangled holographic CFT states, with equal (and large) amount of entanglement, where the semiclassical dual of one has a connected ERB and the other does not. From this example, we speculate that measures of multipartite entanglement may offer a more complete picture into the emergence of spacetime.

7.1 Introduction

The recent developments on the black hole information problem were catalyzed by a holographic calculation of the entropy of an AdS black hole evaporating into a reservoir [1, 2]. Central to these results was the appearance of a novel quantum extremal surface (QES) χ [3], a stationary point of

$$S_{\text{gen}}[\chi] = \frac{\text{Area}[\chi]}{4G_N} + S_{\text{vN}}[\chi], \tag{7.1}$$

where the second term above computes the von Neumann entropy on one side of the surface χ. Contrary to expectations that QESs appear only in the vicinity of their classical counterparts (which compute von Neumann entropy in the classical setting [4–6]), QESs crucially can appear even in spacetimes that have no classical extremal surfaces – such as black holes formed from collapse. See [7] for a recent review.

The von Neumann entropy is expected to play a central role in the emergence of spacetime (see work starting with [8–12]). The closely related, but not identical, conjecture of ER=EPR proposes that (bipartite) entanglement between two parties can be geometrized by a spacetime connecting them [13–15]. (This was originally proposed as a way to make progress on the monogamy argument of the firewall problem [16–18] by encoding the black hole in the radiation itself, see [14, 19–21] among others.) This expectation is realized in many examples in AdS/CFT, most prominently in the case of the thermofield double on two identical copies of a holographic CFT:

$$|\text{TFD}\rangle = \frac{1}{\sqrt{Z}} \sum_n e^{-\beta E_n/2} |\tilde{n}\rangle |n\rangle . \tag{7.2}$$

For sufficiently high temperatures (and thus entanglement), this state is dual to a static black hole [22] with an Einstein-Rosen bridge (ERB) connecting the asymptotic boundaries. This is despite the fact that each individual term in the sum must be disconnected whenever a geometric dual exists. Since forming the sum creates both entanglement and connection, it appears that entanglement is a crucial ingredient in connectedness.

There are several subtleties in the statements of ER=EPR and spacetime emergence from entanglement. When is the connection semiclassical? Without an independent definition of a "highly quantum" wormhole, what is a sharp (falsifiable) formulation of ER=EPR?[1] In the specific context of AdS/CFT in the limit where the bulk is semiclassical, if a given bipartite CFT state (with a semiclassical dual) has a sufficiently large amount of entanglement , is it enough to guarantee that a semiclassical spacetime that connects them is emergent? Addressing these questions is a critical stepping stone towards understanding spacetime as an emergent phenomenon.

In light of these unknowns and the novel developments starting with [1, 2], let us ask the following: does the new QES after the Page time shed light on the subtleties of ER=EPR and how or whether entanglement builds spacetime?

A specific prediction of ER=EPR is that an old black hole should feature a semiclassical connection between the interior and the radiation. In the context of the recent developments

[1] Note that there is import in ER=EPR even in the absence of such a precise definition: if there is any sense in which the black hole interior is "connected" to the radiation, it is possible to act on the black hole interior by acting on the radiation. We thank J. Maldacena for discussion on this point.

on the holography of evaporating black holes, this question naively presents a puzzle: the island [23] is not connected to the radiation via an ERB, suggesting a potential conflict with ER=EPR.[2] This is somewhat ameliorated by the doubly holographic model (see literature starting with [23] in the context of the Page curve [24]), but the latter requires the bulk matter to be holographic with a classical bulk dual, which limits its range of applicability. Other connections include the Lewkowycz-Maldacena [25] style justifications for the novel QES via spacetime replica wormholes [26, 27], where the latter may be a Euclidean avatar of ER=EPR; these non-factorizing geometries naturally present a new and exciting set of challenges [28–30]. Yet another possibility is turning on gravity in the reservoir – without requiring double holography (see work starting with [31, 32]) – and subsequently testing whether an ERB forms dynamically. In [32], it was shown that black holes evaporating into gravitating baths in certain toy models feature a Hawking-Page-like [33] transition after the Page time; of particular relevance is the work of [34, 35], which started out with boundary conditions for two gravitating universes in JT gravity in the state (7.2) and showed that a Euclidean path integral preparation gave a dominant connected geometry for the computation of the von Neumann entropy. Yet another approach to connect the black hole to the radiation was made in [36], relying crucially on an end-of-the-world brane to cap off the black hole spacetime at $r > 0$. This allowed an additional spacetime representing a toy model of the radiation to be glued to the brane to build connection.

As advocated in [37], we would like to work fully in the Lorentzian bulk, in broad generality, and via operational definitions. Recall that the expectation expressed in [13] is not necessarily that a semiclassical connection exists in the original state, but that the application of a (high complexity) unitary to the radiation should result in a semiclassical wormhole connecting the old black hole to the radiation; that is, since the spacetime is not already semiclassically connected, a unitary on the radiation should render it so. We will therefore be concerned with finding such a unitary map, independently of whether the bath is gravitating or not. Our procedure will be entirely Lorentzian and fully generalizable to any number of dimensions, choice of UV completion (under assumption of the QES formula), and reasonable generic matter.

We give a precise and general procedure that shows that there exists a unitary acting only on the radiation of an old black hole that converts the entanglement between an old black hole and its radiation into a semiclassical ERB. In the context of a two-sided black hole, a simple way to realize this expectation is by merging the radiation with the left black hole. However, prior to the discovery of the novel QES, this perspective was highly nontrivial in the context of a single-sided black hole, as there was no clear geometric division between the "black hole" system and the subset of the interior that could be modified by high complexity operations on the radiation.

The novel QES elucidates this picture: after the Page time, there is a region behind the horizon that can be modified by acting exclusively on the microscopic state of the radiation ρ_{RAD}. This is a key ingredient: the existence of a nontrivial QES after the Page time means that a Cauchy slice of the entanglement wedge of a single-sided black hole is extendible (even

[2]There is a sense in which the interface between the CFT and the bath provides a connection between the black hole and the radiation, but this is not a connection through a gravitational spacetime, and the connection can anyway be severed by turning on reflecting boundary conditions.

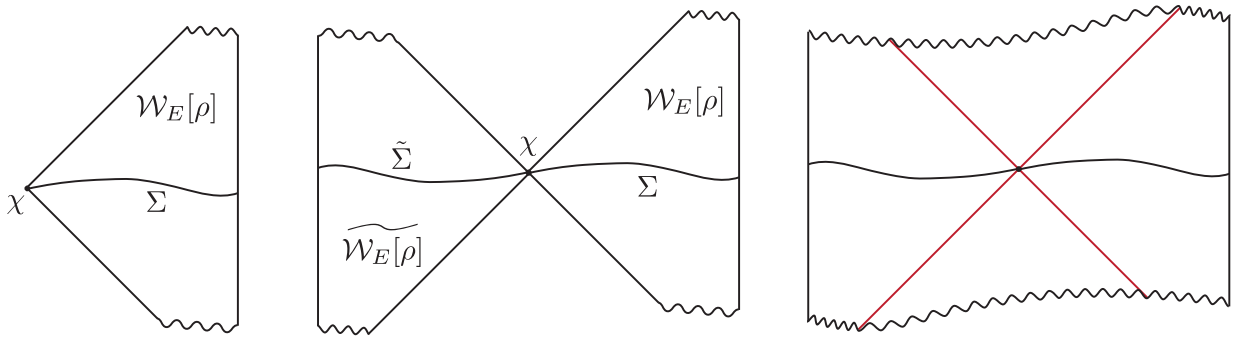


Figure 7.1: To the left we see the entanglement wedge of ρ, and in the center we see $\mathcal{W}_E[\rho]$ glued to its CPT conjugate across the QES/HRT surface χ. To the right we display the final evolution of the data on $\Sigma \cup \tilde{\Sigma}$, giving the spacetime dual to $|\sqrt{\rho}\rangle$. Shockwaves (red) are present when including quantum corrections.

though the boundary Cauchy slice is inextendible). One may hope to apply an explicit map:

$$\rho_{\text{RAD}} \to U \rho_{\text{RAD}} U^\dagger, \tag{7.3}$$

whose bulk dual introduces a boundary behind the QES, realizing precisely the post-Page time expectations that entanglement builds spacetime. Since the novel QES is a generic phenomenon for old black holes [38] (possibly even away from AdS [38–41]), it may be possible to give an algorithmic prescription for the map U that realizes ER=EPR in general old black holes.

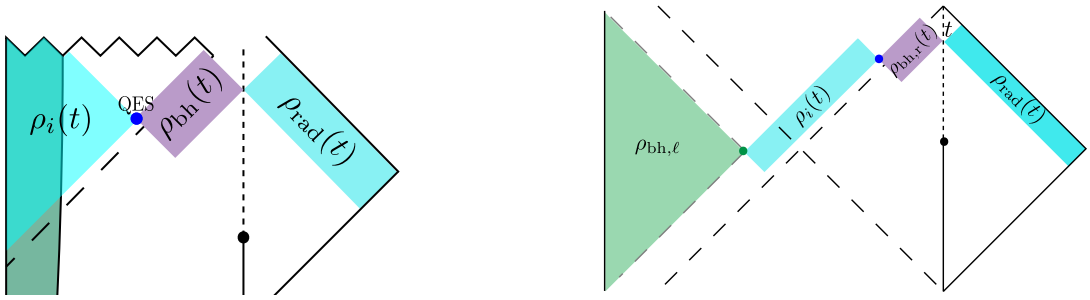
We give a prescription that shows that such a unitary exists using the so-called canonical purification[3]. A mixed state ρ can be purified by doubling the Hilbert space on which it acts $\mathcal{H} \to \mathcal{H} \otimes \mathcal{H}$, and essentially entangling ρ with its CPT conjugate. A familiar example of this procedure is the purification of the Gibbs ensemble via the thermofield double state. As is standard in the literature, we shall refer to the canonically purified state as $\left|\sqrt{\rho}\right\rangle$, and to the trace of $\left|\sqrt{\rho}\right\rangle$ over the original Hilbert space (on which ρ acts) as $\widetilde{\rho}$. This construction will be reviewed at greater length in Sec. 7.2.

The canonical purification has a well-understood geometric dual. Given a mixed CFT state ρ with a classical bulk dual and entanglement wedge $\mathcal{W}_E[\rho]$, it was proposed in [42] that $\widetilde{\rho}$ is dual to $\widetilde{\mathcal{W}_E[\rho]}$, the CPT-conjugate of $\mathcal{W}_E[\rho]$, and the spacetime dual to the full canonically purified state $|\sqrt{\rho}\rangle$ is constructed by gluing $\mathcal{W}_E$ to $\widetilde{\mathcal{W}_E}$ along the HRT surface [5]. See Fig. 7.1. [43] gave a path integral argument for this proposal (and further developed it in the context of the reflected entropy); the proposal was generalized to include bulk quantum corrections in [44]. When quantum corrections are included, the bulk state is canonically purified as well, and the spacetime incurs a shock at the gluing. The geometric dual to $|\sqrt{\rho}\rangle$ will also be reviewed in detail in Sec. 7.2.

We are now in position to explain our main construction. Consider an AdS black hole coupled to a reservoir, as described in [1, 2]. There are two or three "boundary" subsystems

[3]Our prescription is an indirect proof of existence: the canonical purification of the black hole maps unitarily to the radiation, but we do not know how to explicitly write down this unitary.

Figure 7.2: The three relevant non-gravitational systems in the microscopic picture of the evaporating two-sided black hole at a moment of time. The blue circles supports holographic CFTs, while the orange plane supports the reservoir system.



(a) (b)

Figure 7.3: The relevant subregions and states for evaporating one-sided (a) and two-sided (b) black holes in AdS.

of interest, depending on the whether the black hole is one- or two-sided [4]: the microscopic state of the radiation, with state ρ_{RAD}, whose entropy follows the unitary Page curve [24], and the state of the remaining black hole ρ_{BH}; if there are two boundaries, we shall call the remaining black hole states $\rho_{\mathrm{BH,L}}$ and $\rho_{\mathrm{BH,R}}$. For clarity we shall always take the right boundary to be coupled to the reservoir. This is illustrated in Fig. 7.2. On the bulk side, we have three or four bulk regions: the non-gravitating, coarse-grained reservoir rad, whose state we shall denote ρ_{rad}; the entanglement wedge of the old black hole, $\mathcal{W}_E[\rho_{\mathrm{BH}}]$, (or $\mathcal{W}_E[\rho_{\mathrm{BH,R}}]$ for two boundaries), whose state we shall denote ρ_{bh} (or $\rho_{\mathrm{bh,r}}$); the island [23] i, whose state we shall denote as ρ_i; and if relevant, the left black hole $\mathcal{W}_E[\rho_{\mathrm{BH,L}}]$, whose state is denoted by $\rho_{\mathrm{bh},\ell}$. These are illustrated in Fig. 7.3.

We build the canonical purification $|\sqrt{\rho_{\mathrm{BH}}(t)}\rangle$ of the black hole subsystem, working for now with a one-sided black hole for pedagogical clarity. Before the Page time, at $t_1 < t_{\mathrm{P}}$, the QES is empty, and so gluing $\mathcal{W}_E[\rho_{\mathrm{BH}}(t_1)]$ to its CPT conjugate across the QES amounts to simply introducing a second copy of the spacetime.[5]

[4] We can similarly consider more boundaries; the story is qualitatively unchanged. Of course, if there are more than two CFT subsystems, there is an expectation that some multipartite entanglement will be important.

[5] $\widetilde{\mathcal{W}_E[\rho_{\mathrm{BH}}]}$ is introduced to replace the bath system as the purifier of ρ_{BH}, so the bath system is removed rather than doubled.

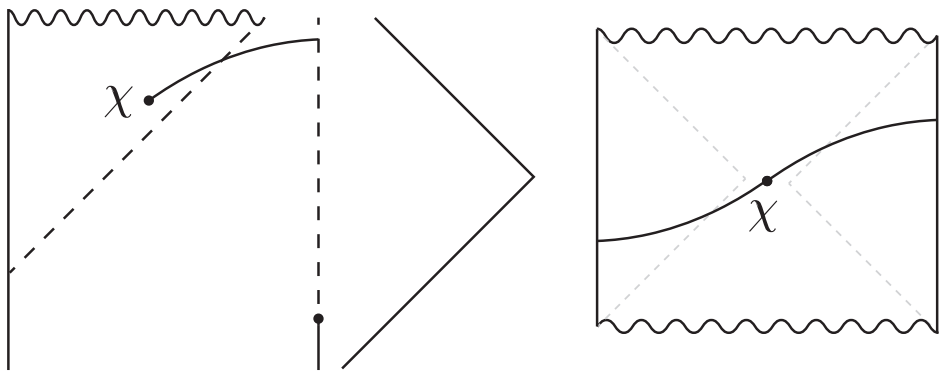


Figure 7.4: On the left we see the QES χ of an evaporating black hole for some time $t > t_P$. On the right, we see the canonical purification, dual to the state $|\sqrt{\rho_{BH}(t)}\rangle$.

After the Page time, the dominant QES is nontrivial: CPT conjugation around the QES yields a *connected* spacetime. [6] Because the QES is close to the event horizon, this is a relatively short wormhole, reminiscent of (but not identical to) AdS-Schwarzschild. This is illustrated in Fig. 7.4. Since the QES for $\rho_{\rm BH}$ is identical to that of $\widetilde{\rho_{\rm BH}}$, the Page curve for the black hole immediately implies a Page curve for its canonical purification; since the von Neumann entropy is a unitary invariant (and purifications are unitarily-related), this also implies a Page curve for the radiation. Let us emphasize this point: by employing this unitary map, we circumvent any modification of the homology constraint in the QES formula: the standard QES formula applied to the canonical purification obeys the Page curve, which is a unitary invariant. The ERB that forms in the canonical purification after the Page time is a more conventional and more general manifestation of the connectivity that appears in the doubly holographic model. Because $\rho_{\rm RAD}$ and the canonical purification of $\rho_{\rm BH}$ are both purifications of $\rho_{\rm BH}$, they are related by a unitary. Thus this immediately shows that there exists a unitary acting exclusively on the post-Page radiation, which results in a connected spacetime.

We emphasize that while this story is consistent with earlier expectations that a black hole maximally mixed with its radiation should have a manifestation as an AdS-Schwarzschild-like geometry, this realization of that expectation is crucially reliant on the more recent understanding of the phenomenon of QESs far from classical extremal surfaces.

It may be tempting at this point to conclude that the canonical purification is a good candidate for a precise realization of the 'entanglement builds spacetime' paradigm: if the canonical purification of a density matrix with sufficient entanglement has a semiclassical bulk dual, then that bulk dual is connected.

Here, however, we recall the case of the young single-sided black hole, whose canonical purification as described above is disconnected. Because the Page curve is non-monotonic, we may choose $t_1 < t_P$ and $t_2 > t_P$ such that $S_{\rm vN}[\rho_{\rm BH}(t_1)] = S_{\rm vN}[\rho_{\rm BH}(t_2)]$. However $|\sqrt{\rho_{\rm BH}(t_2)}\rangle$ gives a connected geometry, while $|\sqrt{\rho_{\rm BH}(t_1)}\rangle$ has a disconnected bulk dual *even though they have the same von Neumann entropy.* This serves as a concrete example of a holographic semiclassical spacetime satisfying the standard QES prescription (with no modification of

[6] This gluing procedure bears some resemblance to [45], but the similarity is superficial.

the homology constraint) in which the amount of fine-grained entropy cannot diagnose the emergence of spacetime connecting the bipartite subsystems. Let us emphasize this point: while nonstandard generalizations of the QES formula that modify the homology constraint – i.e. the "island" formula – may require a more liberal interpretation of spacetime emergence (e.g. in an additional dimension [23] or via replica wormholes [26, 27]), here we have a conventional setup in which we may apply the standard QES formula (with the disconnected dominant topology in the replica trick), and we find that a large amount of bipartite entanglement cannot guarantee an emergent connected spacetime.

To summarize, our main technique relies on the following observation: that the canonical purification of an entanglement wedge with a nontrivial QES is connected, while the canonical purification of an entanglement wedge with an empty QES is disconnected. Our main purpose here consists of applications of this observation in a context where nontrivial QESs were previously unexpected (e.g. single-sided black holes formed from collapse). In those contexts, this property of the canonical purification allows us to (1) give an argument for the existence of the unitary U_{RAD} realizing ER=EPR, and (2) show that bipartite entanglement may fail to build spacetime even in bipartite states with semiclassical bulk duals.

Furthermore, we show that obvious refinements to spacetime from bipartite entanglement conjecture, from reflected entropy to mutual information, complexity, or proximity to the maximally mixed state (as measured by $S_{\text{thermal}} - S_{\text{vN}}$) cannot be solely responsible for spacetime emergence. This requires a refinement of the general expectation (e.g. [46]) that a sufficient von Neumann entropy of bipartite holographic states necessarily builds spacetime between the subsystems.

Under the assumption of entanglement wedge nesting and reconstruction, our results are robust against any unitary acting on the radiation: amount of entanglement is not a sufficient criterion for a semiclassical connection between the black hole and the (unitarily modified) radiation. It is simple to see why: any Cauchy slice of the entanglement wedge of the young black hole is inextendible. Without modifying $\rho_{\text{BH}}(t_1)$ in some way, it is impossible to create a spacetime connection between it and any other spacetime.[7]

This has somewhat different implications for ER=EPR and the specific question of spacetime emergence. There is at least one interpretation of ER=EPR that maintains its applicability to the black hole information problem and is consistent with our results: that a factorized unitary $U_{\text{RAD}} \otimes U_{\text{BH}}$ applied to the young black hole and its radiation $|\Psi(t < t_P)\rangle$ can yield a connected geometry.[8] This requires U_{BH} to change the topology of $\mathcal{W}_E[\rho_{\text{BH}}]$, but there is no obvious obstacle to that. However, this interpretation poses a puzzle for spacetime emergence independent of the information problem considerations of ER=EPR. To understand this latter phenomenon, we would like to identify the property of a given quantum state (rather than some unitary equivalent) that generates spacetime. In the case of a young black hole, we have a bipartite state $|\sqrt{\rho(t < t_P)}\rangle$ whose marginals have conventional semiclassical bulk duals with $\mathcal{O}(G_N^{-1})$ bipartite entanglement, and yet that entanglement fails to build spacetime between them. In the case of the canonically purified old black hole $|\sqrt{\rho(t > t_P)}\rangle$, the same amount of bipartite entanglement successfully builds a connecting spacetime. It is immaterial that a factorized unitary could possibly be applied to make

[7] Unless one adds additional spacetime dimensions, as in [23], or the semiclassical picture for ρ_{BH} is invalid.

[8] It is of course trivial that a nonfactorizing unitary can accomplish this.

$|\sqrt{\rho(t < t_P)}\rangle$ connected: our task, to understand what gives rise to an emergent spacetime in a given state, remains unsolved. In fact, this example makes it clear that whatever quantity is responsible, it cannot be a unitary invariant. This, along with other considerations that we will discuss in Sec. 7.4, is suggestive that *multipartite* entanglement may play a crucial role in distinguishing between even *bipartite* states with connected and disconnected semiclassical duals. We conclude with a discussion in Sec. 7.5 of the shortfalls of various natural refinements and remaining possibilities, including the possible role of reconstruction within a code subspace of typical microstates.

7.2 Holographic Canonical Purification

Let us begin with a brief review of the canonical purification of some mixed state ρ, for convenience chosen to be in a diagonal basis:

$$\rho = \sum_i p_i |\rho_i\rangle\langle\rho_i|. \tag{7.4}$$

We may obtain a pure state by doubling the Hilbert space and essentially duplicating the conjugated state by sending bras to kets:

$$|\sqrt{\rho}\rangle = \sum_i \sqrt{p_i}\, |\rho_i\rangle|\tilde{\rho}_i\rangle, \tag{7.5}$$

where $|\tilde{\rho}_i\rangle$ is the CPT conjugate of $|\rho_i\rangle$. We will refer to the trace of $|\sqrt{\rho}\rangle$ over the original system as $\widetilde{\rho}$ for clarity.

A very familiar example of this procedure is the map from the Gibbs ensemble to the thermofield double state:

$$\rho_\beta = \frac{1}{Z}\sum e^{-\beta E_n}|n\rangle\langle n| \;\rightarrow |\text{TFD}\rangle \equiv \left|\sqrt{\rho_\beta}\right\rangle = \frac{1}{\sqrt{Z}}\sum e^{-\beta E_n/2}|n\rangle|\tilde{n}\rangle. \tag{7.6}$$

The canonical purification may be thought of as the generalization of this procedure to more general, non-thermal states. The mathematically rigorous procedure, including the proof of existence and uniqueness, is simply an implementation of the GNS construction [47, 48].

In the classical bulk regime, the holographic dual of this purification was proposed [42] to be the CPT-conjugation of $\mathcal{W}_E[\rho]$ across the HRT surface. To be precise, we take any Cauchy slice of $\mathcal{W}_E[\rho]$ with its gravitational initial data (Σ, h_{ab}, K_{ab}) as well as initial data for any matter fields. In an abuse of notation, we will collectively refer to this entire set of initial data of $\mathcal{W}_E[\rho]$ as Σ. We then CPT-conjugate Σ across the HRT surface, generating a maximally extended Cauchy slice for a complete, possibly multi-boundary spacetime. [9] This can be thought of as a gluing of Σ to its CPT conjugate $\widetilde{\Sigma}$ across the HRT surface χ. The initial data is then evolved via the standard Cauchy problem to generate the rest of the spacetime. This is illustrated in Fig. 7.1.

[9] The spacetime will be multi-boundary whenever the HRT surface in question is nontrivial for a complete connected boundary. The procedure however can also be implemented for subregions.

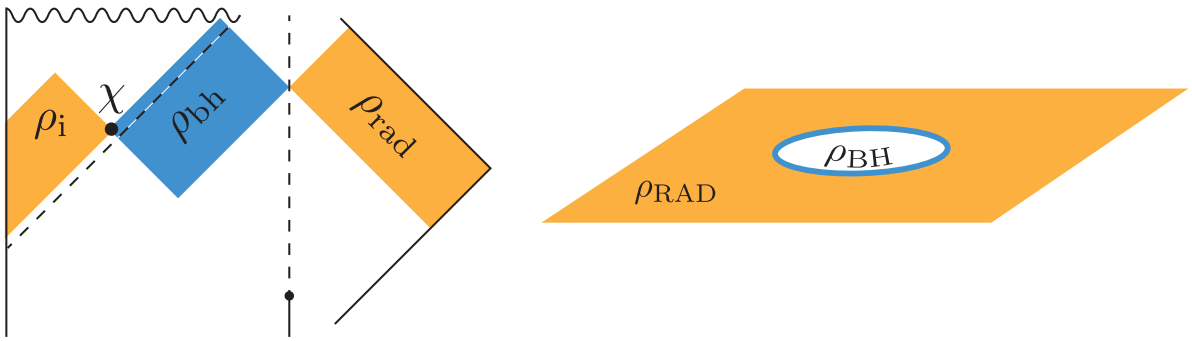


Figure 7.5: On the left we see the bulk picture of an AdS black hole evaporating into a reservoir, with χ the QES for some time $t > t_P$. On the right, we see the microscopic picture at the fixed time t.

The construction is straightforward largely due to the fact that the HRT surface by definition has vanishing mean curvature $K^a = 0$, which simplifies the junction conditions [42, 49] of Σ with $\widetilde{\Sigma}$. Under the inclusion of quantum corrections, the classical extremality condition $K^a = 0$ is replaced by quantum extremality. The quantum corrected generalization of the holographic dual to the canonical purification naturally requires CPT conjugation around the QES (and canonical purification of the bulk state) [44]. A QES, however, does not generically satisfy $K^a = 0$. Consequently, gluing Σ to its CPT conjugate incurs a shock that corrects for the mismatch between $K^a[\chi]$ on approach from Σ and from $\widetilde{\Sigma}$. The precise form of the shock, derived in [44], is:[10]

$$T_{kk} = \frac{1}{\pi}\delta(\lambda)\frac{\delta S_{\rm out}[V(\lambda)]}{\delta V(\lambda)}, \tag{7.7}$$

where k^a is a null vector orthogonal to χ, $S_{\rm out}$ the von Neumann entropy of the bulk state outside of the QES, λ the affine parameter of k^a, and $V(\lambda)$ a function determining the location of a slice of the null congruence generated by k^a – see [44] for details. The protocol, including the expected location of the shocks, is illustrated in Fig. 7.1.

Note that by definition, the canonical purification leaves the original state ρ unaltered. Likewise, its bulk dual makes no modification to $\mathcal{W}_E[\rho]$. Thus any canonical purification of the boundary dual to an evaporating black hole will leave the black hole entanglement wedge unchanged.

7.3 ER=EPR from the New QES

As a central point of ER=EPR, [13] conjectured that the Hawking radiation of an old black hole connects to the black hole interior through a wormhole. While the wormhole may not be geometric, the proposal was that it can be made semiclassical by acting on the Hawking radiation with a unitary – e.g. collapsing it into a second black hole [14].

[10] We must multiply by a factor of two, since we are CPT conjugating about a QES rather than a quantum minimar, leading the two shocks of [44] to coincide.

As explained in the introduction, we take an asymptotically (say, one-sided) AdS black hole, which we shall initially take to be evaporating into a reservoir. Let $|\Psi(t)\rangle$ be the full state of the system on $\mathcal{H}_{\rm BH} \otimes \mathcal{H}_{\rm RAD}$ at boundary time $t > t_P$ and denote the reduced states as $\rho_{\rm BH}(t)$ and $\rho_{\rm RAD}(t)$. See Fig. 7.5.

The canonical purification $|\sqrt{\rho_{\rm BH}(t)}\rangle \in \mathcal{H}_{\rm BH} \otimes \mathcal{H}_{\rm BH}$ after the Page time is a new two-sided connected spacetime, which is CPT symmetric about $\chi(t)$; the bulk quantum state $|\psi(t)\rangle$ is the canonical purification of $\rho_{\rm bh}(t)$, the bulk quantum state in $\mathcal{W}_E[\rho_{\rm BH}(t)]$. How would we diagnose connectivity in this case if we do not have access to the global geometry of spatial slices?

Nonlinear quantities such as the mutual information have historically been proposed as appropriate diagnostic tools for spacetime connectivity [8]. For instance, in a purely classical bulk, the mutual information of two subregions is $\mathcal{O}(1/G_N)$ only if their entanglement wedge is connected. These ideas immediately break down once bulk entanglement can make appreciable contributions to $S_{\rm gen}[\chi]$, as we will see in later sections.

It is fairly simple to give a specialized sufficient condition for connectivity of a class of low-complexity states: simply couple the two sides together via some Gao-Jafferis-Wall-style protocol [50] (in JT gravity as in [51]). If the QES is approximately on the horizon, by introducing this coupling, we can make the wormhole traversable; this effect is easily detected by the near-boundary limit of bulk correlators [51]. In higher dimensions, under the assumption that a similar protocol as in [50] can be made to work in general, the same diagnostic may be used as a sufficient condition. For the states that we consider in this paper this is actually enough to unambiguously diagnose connectivity. If we execute this protocol for our post-Page time canonical purification, we will indeed recover connectivity for this spacetime.

As an aside, we may execute this protocol whenever the spacetime has no Python's lunch [52]: if the outermost QES is in fact the minimal QES. In that case, the intuition would be that the outermost (and in this case, minimal) QES lives on the bifurcation surface whenever there is no infalling matter (and gravitational waves) [42]; this matter may in principle be removed by acting with simple sources [42, 53], however quantum backreaction introduces a number of subtleties discussed in [54].

Returning to the construction at hand, since any two purifications of $\rho_{\rm BH}(t)$ are unitarily related,[11] there exists a unitary $U_{\rm RAD}$ with support only on $\mathcal{H}_{\rm RAD}$ that satisfies

$$U_{\rm RAD} \otimes \mathbb{I}_{\rm BH} |\Psi(t)\rangle = \left|\sqrt{\rho_{\rm BH}(t)}\right\rangle . \tag{7.8}$$

Here $U_{\rm RAD}$ can be seen as a quantum computation on the radiation that gives a semiclassical geometrization of the entanglement between the radiation and the black hole. Effectively, $U_{\rm RAD}$ collapses the Hawking radiation into a black hole and then shortens the resulting wormhole as much as is possible when only acting on the radiation. Since $U_{\rm RAD}$ has no

[11] For this to be true, the two Hilbert spaces should have the same dimension. One might wonder if $\mathcal{H}_{\rm RAD}$ is too large to be unitarily related to $\mathcal{H}_{\rm BH}$. However, this should not cause concern: most of the Hilbert space $\mathcal{H}_{\rm RAD}$ is not explored by the Hawking radiation – we only permit modes into $\mathcal{H}_{\rm RAD}$ that could fit into $\mathcal{H}_{\rm BH}$ in the first place. Thus, we can divide $\mathcal{H}_{\rm RAD} = \mathcal{H}_{\rm ENT} \otimes \mathcal{H}_{\rm AUX}$ with $|\mathcal{H}_{\rm ENT}| = |\mathcal{H}_{\rm BH}|$, and act with a unitary on the radiation to get a state that factorizes on $\mathcal{H}_{\rm ENT} \otimes \mathcal{H}_{\rm AUX}$, so that the factor $\mathcal{H}_{\rm ENT}$ contains all entanglement with the black hole.

support on $\mathcal{H}_{\rm BH}$, the entanglement wedge of BH is left unchanged. So is the entanglement spectrum of the black hole.

The procedure works equally well for two-sided black holes. With two CFTs on $\mathcal{H}_{\rm BH_L} \otimes \mathcal{H}_{\rm BH_R}$, with the system on $\mathcal{H}_{\rm BH_R}$ coupled to $\mathcal{H}_{\rm RAD}$, $|\sqrt{\rho_{\rm BH_R}(t)}\rangle$ geometrizes the entanglement between $\rm BH_R$ and $\rm BH_L \cup RAD$.[12] Now the unitary relating the full state $|\Psi(t)\rangle \in \mathcal{H}_{\rm BH_L} \otimes \mathcal{H}_{\rm BH_R} \otimes \mathcal{H}_{\rm RAD}$ to the canonical purification has support only on $\mathcal{H}_{\rm BH_L} \otimes \mathcal{H}_{\rm RAD}$ and effectively throws the radiation into the left black hole and shortens the wormhole as much as possible while leaving the right black hole alone.

Let us emphasize a few aspects of our construction. First, the canonical purification of the (right) black hole is not identical to the thermofield double and can never be made into the TFD as long as we only act on the radiation (and left black hole). Second, while it might be intuitive that some action on the Hawking radiation, like collapsing it into a black hole, can create a semiclassical geometric connection to the original black hole, there has to our knowledge not been any explicit demonstration of a unitary implementing this in broad generality. Our protocol relies crucially on the recently discovered post Page time QES. In fact, more surprisingly, we will show in Sec. 7.4 that in the absence of such a QES, it is not possible to construct a wormhole connecting an unmodified $\mathcal{W}_E[\rho_{\rm BH}]$ to the radiation even when the two are highly entangled, showing that in the absence of significant modifications, a strict interpretation of ER=EPR will not hold.

7.3.1 A Simple AdS/CFT Dictionary for Unitary Invariants

Because $|\sqrt{\rho_{\rm BH}(t)}\rangle$ is unitarily related to $|\Psi(t)\rangle$ via a unitary with no support on BH, $S_{\rm vN}[\rho_{\rm RAD}(t)]$ can be computed in the CPT conjugated spacetime. In particular, computing $S_{\rm vN}[\rho_{\rm RAD}(t)]$ in $|\sqrt{\rho_{\rm BH}(t)}\rangle$ requires no modification of the homology constraint of the QES: $|\sqrt{\rho_{\rm BH}(t)}\rangle$ is a standard instance of AdS/CFT. The entropy of the complement of BH is computed by $S_{\rm gen}[\chi[t]]$, and the state on the complement of BH is unitarily related to $\rho_{\rm RAD}(t)$; thus $S_{\rm gen}[\chi[t]] = S_{\rm vN}[\rho_{\rm RAD}(t)]$. More generally, any quantity that is invariant under local unitaries will be identical in the two states, giving a way of computing unitary invariants without novel modifications.[13]

The standard AdS/CFT picture of the radiation presented by $|\sqrt{\rho_{\rm BH}}\rangle$ provides a simple geometrization of the difference in the experience of an infalling observer before and after the Page time. If we turn on a sufficiently early local unitary in $\widetilde{\rho_{\rm BH}}[t > t_P]$, an infalling observer will encounter the energy shock resulting from this unitary. This is illustrated in Fig. 7.6. We naturally expect that acting on the radiation can modify the experience of the infalling observer after the Page time; the dual of $|\sqrt{\rho_{\rm BH}}\rangle$ precisely geometrizes this expectation, yielding the same setup as in Marolf-Wall [56]. Indeed, it is geometrically clear that there is no localized operator acting on $\widetilde{\rho_{\rm BH}}[t < t_P]$ that can modify the experience of an infalling observer prior to the Page time. See Fig. 7.6.

[12] Note that in this case we may also choose to canonically purify the state on $\mathcal{H}_{\rm BH_L} \otimes \mathcal{H}_{\rm BH_R}$, resulting in $|\sqrt{\rho_{\rm BH_R\ BH_L}(t)}\rangle$. The same line of reasoning applies.

[13] It would be interesting to understand the import on the work of [55] regarding implications of relaxing the homology constraint.

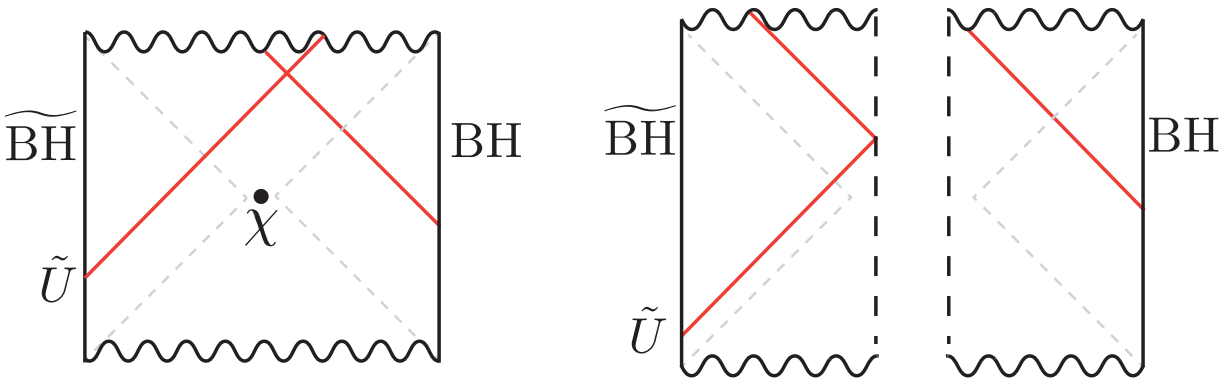


Figure 7.6: Canonical purifications after (left) and before (right) the Page time. After the Page time, a signal sent from $\widetilde{\text{BH}}$ via some local unitary $\tilde{U}$ can effect an infalling observer from BH. Before the Page time, this cannot happen.

7.3.2 Concrete Example: JT Gravity with Conformal Matter

To alleviate any concerns about potential sick behavior of the canonical purification, let us now construct fully explicit canonical purifications of the evaporating one-sided black hole in JT gravity coupled to conformal matter. Indeed, we will see that the average null energy condition holds and that the QES is acausally separated from the conformal boundary. We first summarize the salient results from [2, 57].

The Evaporating Black Hole in JT Gravity

We work with JT gravity [58–60] coupled to a CFT

$$\begin{aligned} I_{\text{JT}} &= I_0[g_{ab}] + I_g[g_{ab},\phi] + I_{\text{CFT}}[g_{ab}], \\ I_0 &= \frac{\phi_0}{16\pi G}\left[\int_M R + 2\int_{\partial M} K\right], \\ I_g &= \frac{1}{16\pi G}\left[\int_M \phi(R+2) + 2\int_{\partial M}\phi(K-1)\right], \end{aligned} \tag{7.9}$$

which on-shell yields a locally AdS_2 metric with the following dilaton equations of motion:

$$-\nabla_a\nabla_b\phi + g_{ab}\nabla^2\phi - g_{ab}\phi = 8\pi G_N T_{ab}, \tag{7.10}$$

where T_{ab} is the CFT stress tensor. We work in the semiclassical approximation where we replace T_{ab} with $\langle T_{ab}\rangle$, and omit brackets from now on.

We take the boundary ∂M of the AdS_2 spacetime to lie at a finite cutoff, with boundary conditions

$$h_{uu} = \frac{1}{\epsilon^2}, \qquad \phi|_{\partial M} = \frac{\bar{\phi}_r}{\epsilon}, \tag{7.11}$$

for constants $\bar{\phi}_r$ and $\epsilon \ll 1$, where h_{uu} is the induced metric on the boundary, and u the boundary time. We will make use of Poincare coordinates,

$$\mathrm{d}s^2 = -\frac{4\mathrm{d}x^+\mathrm{d}x^-}{(x^+ - x^-)^2} = \frac{-\mathrm{d}t^2 + \mathrm{d}z^2}{z^2}, \qquad x^\pm = t \pm z, \tag{7.12}$$

and describe the boundary location by the function $f(u)$, which gives its Poincare time as function of boundary time: $t|_{\partial M} = f(u)$. Solving (7.11) to leading order in ϵ then gives $z|_{\partial M} = \epsilon f'(u)$.

A two-sided static JT black hole at temperature T and with energy E is given by

$$\phi = 2\bar{\phi}_r \frac{1 - (\pi T)^2 x^+ x^-}{x^+ - x^-}, \qquad E = \frac{\pi \bar{\phi}_r}{4G_N} T^2. \tag{7.13}$$

In [2] a black hole at temperature T_0 was coupled to a bath CFT_2 living on a half-line by turning on absorbing boundary conditions at the AdS boundary for $t \geq 0$. Instantly turning on absorbing boundary conditions leads to the injection of a positive energy shockwave into the bulk that raises the temperature of the black hole to $T_1 > T_0$ at $t = 0^+$, and after that the effective temperature falls off as the black hole evaporates. The resulting bulk stress tensor is

$$T_{x^-x^-}(x^-) = E_S \delta(x^-) - \frac{c}{24\pi}\{f^{-1}(x^-), x^-\}\theta(x^-), \qquad T_{x^+x^+} = T_{x^+x^-} = 0, \tag{7.14}$$

where E_S is the injected shockwave energy, c the central charge of the bulk CFT, $\{\cdot,\cdot\}$ the Schwarzian derivative, and θ a step function. After determining the stress tensor, the backreaction on the dilaton can be found using the fact that the dilaton equations of motion can be directly integrated when $T_{x^+x^-} = 0$ [60].

Now, to explicitly solve for the boundary trajectory $f(u)$, it is useful to have the spacetime energy $E(u)$, which is given by:

$$E(u) = \theta(-u)E_0 + \theta(u)E_1 e^{-ku}, \qquad k = \frac{cG_N}{3\bar{\phi}_r}, \tag{7.15}$$

where $E_0 = \frac{\pi\bar{\phi}_r}{4G_N}T_0^2$ and $E_1 = E_0 + E_S = \frac{\pi\bar{\phi}_r}{4G_N}T_1^2$. (7.15) together with (7.56) in the appendix can be used to solve for $f(u)$.

The Construction

Now let us turn to the case at hand: the explicit construction of the canonical purification after the Page time, which here means $u \sim \mathcal{O}(k^{-1})$. We can either consider the same setup as in [2], meaning we work with a two-sided black hole evaporating on one side, or we can replace the left conformal boundary with an (unflavored) end-of-the-world brane, so that our spacetime is one-sided, as in [61]. The effect of this modification is that the early time QES becomes the empty surface, which is now homologous to the right conformal boundary. The late time QES is unchanged, giving thus the same late time canonical purification.[14]

[14] We assume the brane has sufficient tension so that, at $t = 0$, the location of the brane approaches that of the would-be physical conformal boundary defined by (7.11).

To construct the canonical purification at some fixed boundary time u_∂, we do the following: (1) take a spatial slice Σ anchored at the late time QES and the boundary at $t_\partial = f(u_\partial)$ and compute the dilaton and CFT stress tensor on Σ, (2) glue Σ to its CPT conjugate slice $\tilde{\Sigma}$ across the QES, and (3) evolve the new initial data $\Sigma \cup \tilde{\Sigma}$ with some choice of boundary conditions.

Any choice of boundary conditions will give the same spacetime in the Wheeler-de-Witt patch $D[\Sigma \cup \tilde{\Sigma}]$, and so the geometry encoded in the canonical purification at time t_∂, which encodes the entanglement structure between the black hole and the Hawking radiation at that time, does not care about boundary conditions. However, it is interesting to see a more complete development of the spacetime, and so we will choose to evolve the state beyond $D[\Sigma \cup \tilde{\Sigma}]$ with reflecting boundary conditions. The latter means that there is no flux of energy across the conformal boundary:

$$T_{x^-x^-}|_{\partial M} - T_{x^+x^+}|_{\partial M} = 0. \tag{7.16}$$

This seemingly presents a problem: the data on Σ is just the data inherited from the evaporating black hole, so for any $u_\partial > 0$ and choice of Σ, the left hand side of (7.16) is nonzero. One way to deal with this is to turn off the transparent boundary conditions over a small time window $2\delta u$ around u_∂, which leads to the injection of an energy shock and imperfect reflectivity over the time window $[u_\partial - \delta u, u_\partial + \delta u]$. We can work in the limit $\delta u \to 0$ so that the injected shock becomes a delta function. In this case, evolving T_{ab} with reflecting boundary conditions translates into finding a distributional solution of $\nabla_a T^{ab} = 0$ that agrees with the stress tensor in $D[\Sigma] \cup D[\tilde{\Sigma}]$, and which satisfies reflecting boundary conditions in the distributional sense:

$$\int_{u_0-\delta u}^{u_0+\delta u} \mathrm{d}u \left[T_{x^-x^-}(u) - T_{x^+x^+}(u)\right]|_{\partial M} = 0, \qquad \forall u_0 \in \mathbb{R}, \delta u > 0. \tag{7.17}$$

Initial data

Let now $u_\partial \sim \mathcal{O}(k^{-1})$ be after the Page time. We begin by finding the initial data to be evolved in Poincare coordinates. After this we transition to global coordinates, where we do the full evolution.

The late time QES and the boundary time in question is strictly to the future of the shockwave arising from turning on the coupling to the bath, so the stress tensor is just given by the Schwarzian term in (7.14). As shown in [2], for $u \sim \mathcal{O}(k^{-1})$, we have

$$\{f^{-1}(u), u\} = \frac{1}{2(u - t_\infty)^2}\left(1 + \mathcal{O}(k^2 e^{-ku})\right), \tag{7.18}$$

where t_∞ is the Poincare time at which the dilaton boundary terminates to the future in the original evaporating black hole spacetime. It can be checked that the approximation (7.18) is valid for all x^- covered by a spatial slice running from the boundary to the QES, and so our stress tensor initial data on the interior of Σ, or more precisely in $D[\Sigma]$, reads

$$T_{x^-x^-}|_{D[\Sigma]} \approx -\frac{c}{48\pi}\frac{1}{(x^- - t_\infty)^2}. \tag{7.19}$$

As we alluded to earlier, additional shocks will be present at $\partial\Sigma$ in the canonical purification, i.e. at the QES and at the conformal boundary. We will return to these shocks in a moment.

The calculation of the dilaton in $D[\Sigma]$ is somewhat more involved, and carried out in the appendix. The final result is

$$\begin{aligned}\phi(x^+,x^-)|_{D[\Sigma]} = \frac{\bar{\phi}_r}{2(x^+-x^-)}\Big[&(2t_\infty - x^- - x^+)\mathcal{C} - k(x^+-x^-)\\ &+ k(2t_\infty - x^+ - x^-)\log\left(\frac{t_\infty - x^-}{t_\infty}\right)\Big],\end{aligned} \tag{7.20}$$

where we have neglected terms of order $\mathcal{O}\left([t_\infty - x^\pm]^2\right)$ in the square brackets, which are suppressed since $t_\infty - x^\pm$ is non-perturbatively small on the whole of Σ. The constant $\mathcal{C}$ is

$$\mathcal{C} = \frac{4}{t_\infty} + k(2+\gamma), \tag{7.21}$$

where the $\mathcal{O}(1)$ constant γ is defined in the appendix.

Changing coordinates

We have thus far determined the initial data on $D[\Sigma]$. Let us now transition to global coordinates, where we will work out the shocks and evolve our initial data. We define global coordinates through

$$\begin{aligned}\mu(t+\lambda) &= \frac{\sin(\tau-\tau_0)}{\cos(\tau-\tau_0)+\sin\rho}, \qquad \rho\in\left(-\frac{\pi}{2},\frac{\pi}{2}\right),\\ \mu z &= \frac{\cos\rho}{\cos(\tau-\tau_0)+\sin\rho},\end{aligned} \tag{7.22}$$

where $\lambda,\tau_0\in\mathbb{R}$ and $\mu>0$ are constants parametrizing isometries that we can choose freely to obtain the most convenient coordinates. This brings the metric to the form

$$\mathrm{d}s^2 = \frac{1}{\cos^2\rho}\left(-\mathrm{d}\tau^2+\mathrm{d}\rho^2\right) = \frac{-\mathrm{d}w^+\mathrm{d}w^-}{\cos^2\left(\frac{w^+-w^-}{2}\right)}, \qquad w^\pm = \tau\mp\rho, \tag{7.23}$$

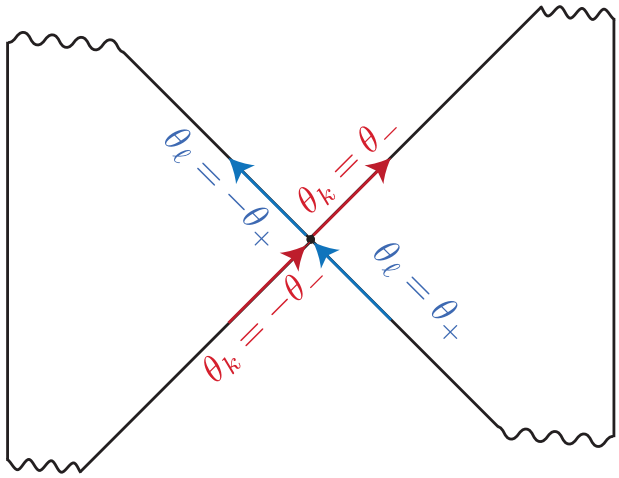
with the right conformal boundary is at $\rho=\frac{\pi}{2}$.

For convenience we now want to pick λ,μ,τ_0 so that our initial data slice Σ can be taken to be the $\tau=0$ slice, and with the QES at $\rho=0$. The solution is worked out in the appendix. The result is

$$\chi\equiv\mu(t_\infty+\lambda) = -\frac{1}{4}+\frac{5e^{-ku_\partial/2}k}{8\pi T_1}+\mathcal{O}(k^2), \tag{7.24}$$

$$\tau_0 = \frac{\pi}{2}-\frac{ke^{-ku_\partial/2}}{\pi T_1}+\mathcal{O}(k^2), \tag{7.25}$$

$$\mu = \frac{1}{t_\infty - f(u_\partial)}\left(\frac{3}{4}-\frac{3ke^{ku_\partial/2}}{8\pi T_1}+\mathcal{O}(k^2)\right). \tag{7.26}$$



$$[\theta_\ell] = \theta_\ell|_{\text{future}} - \theta_\ell|_{\text{past}} = -2\theta_+$$
$$[\theta_k] = \theta_k|_{\text{future}} - \theta_k|_{\text{past}} = 2\theta_-$$

Figure 7.7: Illustration of the discontinuities of the null expansions across the QES, where $\theta_\pm$ are the expansions in the right wedge.

Determining the shocks

Next, let us work out the shocks present at the QES caused by the discontinuous derivatives of the dilaton after gluing Σ to $\tilde{\Sigma}$. In the appendix, we derive the null junction conditions in JT gravity and show that a null junction S with a future-directed generator k^a (unrelated to the constant k) induces a stress tensor shock

$$T_{ab}\ell^a\ell^b = -\frac{1}{8\pi G_N}\delta(\lambda)[\ell^a\nabla_a\phi], \tag{7.27}$$

where $\ell^a = \left(\frac{\mathrm{d}}{\mathrm{d}\lambda}\right)^a$ is the null tangent of a future-directed (not necessarily affine) geodesic with $k\cdot\ell = -1$ and $\lambda = 0$ on S. The bracket denotes the discontinuity across the junction in the direction from past to future.

Consider now a specific set of null-vectors:

$$\tilde{\ell}^\mu = (\partial_{w^+})^\mu, \qquad \tilde{k}^\mu = 2\cos\left(\frac{w^+ - w^-}{2}\right)^2 (\partial_{w^-})^\mu, \tag{7.28}$$

and let us focus on the junction generated by $\tilde{k}^a$. See Fig. 7.7. For this normalization, $\lambda = w^+$. Defining now the expansion of the QES as we approach it from $D[\Sigma]$ to be

$$\tilde{\theta}_+ \equiv \tilde{\ell}^a\nabla_a\phi|_{\text{QES},\Sigma} = \partial_{w^+}\phi|_{\text{QES},\Sigma}, \tag{7.29}$$

we have

$$T_{w^+w^+}|_{\text{shock}} = -\frac{1}{8\pi G_N}\delta(w^+)[\theta_{\tilde{\ell}}] = \frac{\tilde{\theta}_+}{4\pi G_N}\delta(w^+) \equiv a_+\delta(w^+). \tag{7.30}$$

The sign of $[\theta_{\tilde{\ell}}]$ is explained in Fig. 7.7. A completely analogous computation gives that

$$T_{w^-w^-}|_{\text{QES shock}} = -\frac{\tilde{\theta}_-}{4\pi G_N}\delta(w^-) \equiv a_-\delta(w^-), \qquad \tilde{\theta}_- = \partial_{w^-}\phi|_{\text{QES},\Sigma}. \tag{7.31}$$

Again the sign is explained in Fig. 7.7. Computing the dilaton derivatives at $w^+ = w^- = 0$, we find to leading order

$$\begin{aligned} a_+ &= -\frac{c}{12\pi} + \mathcal{O}(k\log k), \\ a_- &= \frac{c}{12\pi}\frac{e^{-u_\partial k/2}}{t_\infty k} + \mathcal{O}(\log k). \end{aligned} \tag{7.32}$$

We see that $T_{w^+w^+}$ has negative null energy, but it is easy to check that the ANEC holds for any $b > 0$.

Evolving the stress-tensor

We have now obtained a convenient coordinate system where the QES lies at $(\tau, \rho) = (w^+, w^-) = (0, 0)$, and where Σ can be taken to be the slice $(\tau = 0, \rho \geq 0)$. Let us now change coordinates to prepare for evolution of the data. Combining (7.79), (7.19), and (7.24), we get

$$\begin{aligned} g(w^- \geq 0) \equiv T_{w^-w^-}|_{D[\Sigma]} &= \left(\frac{\partial w^-}{\partial x^-}\right)^{-2} T_{x^-x^-}|_{D[\Sigma]} \\ &= -\frac{c}{48\pi}\frac{1}{\left[\chi + \chi\sin(w^- - \tau_0) + \cos(w^- - \tau_0)\right]^2}. \end{aligned} \tag{7.33}$$

We can now finally write down the initial data on the full Cauchy slice:[15] CPT conjugation acts simply on our new coordinates,

$$\text{CPT}(w^+, w^-) = (-w^+, -w^-), \tag{7.34}$$

and so to glue Σ to its CPT conjugate, all we need to do for the stress tensor is to extend it symmetrically about $\rho = 0$. For the function $g(w^-)$, we must extend g symmetrically about $w^- = 0$. In toto we have

$$\begin{aligned} T_{w^-w^-}|_{\Sigma\cup\tilde{\Sigma}} &= g(w^-) + a_-\delta(w^-), \\ T_{w^+w^+}|_{\Sigma\cup\tilde{\Sigma}} &= a_+\delta(w^-). \end{aligned} \tag{7.35}$$

As explained earlier, we now want a distributional evolution that (1) agrees with (7.35) on $D[\Sigma] \cup D[\tilde{\Sigma}]$ and (2) has reflecting boundary conditions in the distributional sense. Finding the solution is straight forward and carried out in the appendix. Stress tensor conservation $\nabla_a T^{ab} = 0$ together with conformality ($T^a{}_a = 0$) is enough to completely solve for T_{ab} given some initial data. The solution is simply given by

$$T_{w^+w^+} = T_{w^+w^+}(w^+), \qquad T_{w^+w^+} = T_{w^-w^-}(w^-), \qquad T_{w^+w^-} = 0. \tag{7.36}$$

The only complication is working out how reflecting boundary conditions are implemented. The result on the domain $|w^\pm| \leq \pi$ is given in the appendix.

[15] Of course, it is natural to take Σ to be the $\tau = 0$ slice, in which case we set $-w^+ = w^- = \rho$ above, but it is convenient to work directly with $w^\pm$ coordinates.

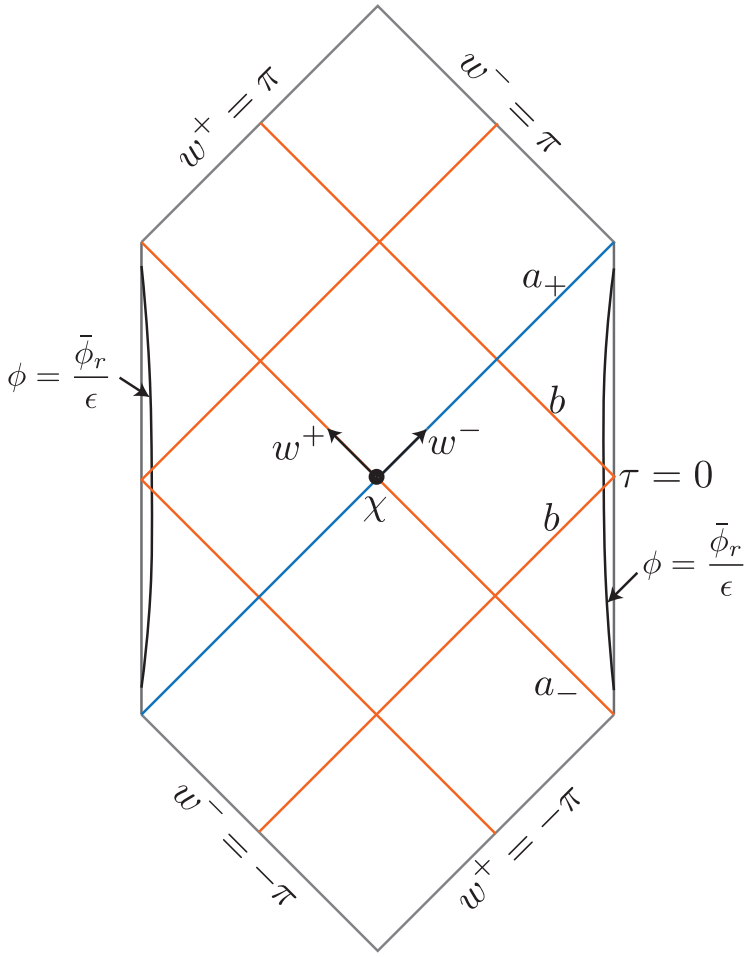


Figure 7.8: Canonical purification of the evaporating black hole in JT gravity. The orange lines show positive null energy shockwaves, while the blue line is a negative null energy shock.

$$\begin{aligned} T_{w^-w^-}(w^-) &= b\delta\left(w^- - \frac{\pi}{2}\right) + b\delta\left(w^- + \frac{\pi}{2}\right) + a_+\delta(w^- + \pi) + a_+\delta(w^- - \pi) + a_-\delta(w^-), \\ &\quad + g(w^-)\theta\left(w^- - \frac{\pi}{2}\right)\theta\left(\frac{\pi}{2} - w^-\right) \\ T_{w^+w^+}(w^+) &= b\delta\left(w^+ - \frac{\pi}{2}\right) + b\delta\left(w^+ + \frac{\pi}{2}\right) + a_-\delta(w^+ + \pi) + a_-\delta(w^+ - \pi) + a_+\delta(w^+), \\ &\quad + \theta\left(w^+ - \frac{\pi}{2}\right) g(w^+ - \pi) + \theta\left(-\frac{\pi}{2} - w^+\right) g(w^+ + \pi), \end{aligned} \tag{7.37}$$

The solution depends on the constant b that gives the injected energy shock caused by turning off absorbing boundary conditions. Its value depends on the CFT dynamics and how the coupling is turned off. However, from previous studies [2, 57] we know that $b > 0$.

We illustrate the shock profiles in Fig. 7.8. Note that we do not care about $|w^\pm| > \pi$ for the following reason: we only need to keep track of the spacetime that is dual to the physical conformal boundary, which is the union of all the WdW patches of the physical conformal boundary. Now, if the QES is to be acausally separated from the boundary, for which $b > 0$

is sufficient (see the appendix), then the physical conformal boundary must terminate to the future and past at some $|\tau| \leq \frac{\pi}{2}$. In null coordinates this translates to the fact that the part of spacetime we are interested in is contained in the region $|w^{\pm}| \leq \pi$.

Evolving the dilaton

Changing to global coordinates, the general solution for the dilaton (7.55) is

$$\begin{aligned}
\phi(w^+, w^-) &= \frac{c_1 + c_2(f_+ + f_-) + c_3 f_+ f_- + 8\pi G_N(\hat{I}_+ + \hat{I}_-)}{f_+ - f_-} \\
\hat{I}_+(w^+, w^-) &= \int_{-\varepsilon}^{w^+} \mathrm{d}s\, [1 - \sin(s - \tau_0)]\, [f_+(s) - f_+]\, [f_+(s) - f_-]\, T_{w^+w^+}(s), \\
\hat{I}_-(w^+, w^-) &= -\int_{\varepsilon}^{w^-} \mathrm{d}s\, [1 + \sin(s - \tau_0)]\, [f_-(s) - f_+]\, [f_-(s) - f_-]\, T_{w^-w^-}(s),
\end{aligned} \tag{7.38}$$

where implicit arguments in $f_\pm$ always mean $f_+ \equiv f_+(w^+)$, $f_- \equiv f_-(w^-)$. We have here chosen the reference point $(w^+, w^-) = (-\varepsilon, \varepsilon)$ in the integrals to lie slightly to the right of the QES inside $D[\Sigma]$, with $\varepsilon > 0$ arbitrarily small.

Getting the final value of the dilaton is now a matter of (1) changing to global coordinates in the dilaton initial data (7.20), (2) matching onto (7.38) to extract the coefficients c_i, and (3) using (7.37) to compute the integrals $\hat{I}_+$ and $\hat{I}_-$. The integrals split up into δ-function contributions and contributions from the continuous part of the stress tensor. The δ-function contributions effectively add jumps to the coefficients c_i as we cross shocks, so we redefine c_i to contain these steps. Obtaining the final dilaton is straightforward but somewhat involved, and so is carried out in the appendix. The final result for $w^- > 0$ reads

$$\begin{aligned}
\phi(w^+, w^-) &= \phi_0(w^+, w^-) + \phi_m(w^+, w^-), \\
\phi_0(w^+, w^-) &= \frac{c_1 + c_2(f_+ + f_-) + c_3 f_+ f_-}{f_+ - f_-}, \\
\phi_m(w^+, w^-) &= \frac{1}{f_+ - f_-}\Big[H_1(w^+, w^-)\theta\left(\frac{\pi}{2} - w^-\right) + H_2\left(w^+, w^-\right)\theta\left(w^- - \frac{\pi}{2}\right) \\
&\qquad + \theta\left(-w^+ - \frac{\pi}{2}\right) H_3(w^+, w^-) + \theta\left(w^+ - \frac{\pi}{2}\right) H_4(w^+, w^-)\Big],
\end{aligned} \tag{7.39}$$

where ϕ_0 is a homogenous solution (away from the shocks), while ϕ_m describes the deviation from a static black hole solution. The piecewise constant coefficients c_i and the functions H_i are given in the appendix.

7.4 Puzzles Before the Page Time

Let us now examine the canonical purification of a young single-sided AdS black hole evaporating into a reservoir; to begin with, we will focus on single-sided black holes (we will consider multiple boundaries in Sec. 7.4.3). Before the Page time, the dominant QES defining the entanglement wedge of ρ_{BH} is simply $\varnothing$: Cauchy slices of $W_E[\rho_{\mathrm{BH}}]$ are inextendible. As

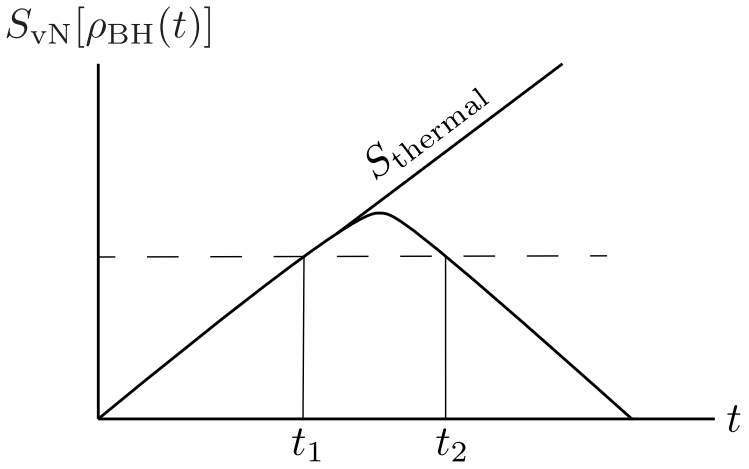


Figure 7.9: Choice of times to canonically purify.

discussed in the introduction, the inextendibility of $W_E[\rho_{\rm BH}]$ has two important consequences: first, if the pre-Page Hawking radiation is collapsed into a black hole, it will not be connected to the original black hole via a semiclassical ERB of the same dimensionality as the original black hole.[16] This follows immediately from inextendibility of Cauchy slices of the pre-Page entanglement wedge $\mathcal{W}_E[\rho_{\rm BH}]$. Second, a sufficient amount of bipartite entanglement cannot imply bulk connectedness of a bipartite state[17] even under existence of a semiclassical dual bulk satisfying the standard QES formula. These two points show that the strictest interpretation of ER=EPR that does not allow modifications of $\rho_{\rm BH}$ cannot be correct: we must allow for factorized unitaries. In Sec. 7.5, we will discuss the possibility that multipartite entanglement may also play a role, even for bipartite states.

7.4.1 Entanglement Entropy is Not Enough

Let us remind the reader of the argument for point (1): that the amount of bipartite entanglement is not a sufficient criterion for the emergence of spacetime between bipartite states with holographic semiclassical duals. As discussed in the introduction (since the dominant QES is $\varnothing$), $\left|\sqrt{\rho_{\rm BH}[t<t_P]}\right\rangle$ is dual to a two-boundary geometry, which is just two disconnected copies of the original one-sided geometry. The bulk state is just $\left|\sqrt{\rho_{\rm bh}}\right\rangle$, the canonical purification of the bulk state in $\mathcal{W}_E[\rho_{\rm BH}]$.

Consider now two boundary times t_1 and t_2 such that $t_1 < t_{\rm P} < t_2$ and

$$S_{\rm vN}[\rho_{\rm BH}(t_1)] = S_{\rm vN}[\rho_{\rm BH}(t_2)]. \tag{7.40}$$

[16] This does not exclude the possibility that collapsing the Hawking radiation into a black hole *and* acting on the original black hole in some nontrivial way could create an ERB.

[17] Let us emphasize this point: while it is clear that a tripartite state will likely require some tripartite entanglement measure, we work with *bipartite* states, and we still find that the von Neumann entropy falls short.

This choice is illustrated in Fig. 7.9. Canonically purifying $\rho_{\rm BH}(t_1)$ yields two disconnected spacetimes whereas canonically purifying $\rho_{\rm BH}(t_2)$ yields a single connected spacetime. We may pick t_1 and t_2 to be close to $t_{\rm P}$ so that $S_{\rm vN}[\rho_{\rm BH}(t)]$ at both times is large, as in Fig. 7.9.

When the QES of a connected component of $\mathscr{I}$ is $\varnothing$, it is impossible to add additional bulk regions (connected to $\mathscr{I}$) without modifying the entanglement wedge $\mathcal{W}_E[\rho_{\mathscr{I}}]$ because Cauchy slices of the entanglement wedge are inextendible. While the formula

$$S_{\rm gen} = \frac{\text{Area}[\chi]}{4G} + S_{\rm vN}[\rho_{\text{out}[\chi]}] \tag{7.41}$$

generally contains an ambiguity in the relative contribution between area and entropy, in the absence of a nontrivial QES, there is no choice of UV cutoff for which there is a surface term. It may in principle be possible to diagnose a nonzero area of a QES purely from the CFT, see [62] for an investigation for certain classes of states, although there are a number of subtleties, e.g. [32].

We may take this to be a more general criterion: if the QES of any complete connected component $\mathscr{I}$ is empty, then the bulk dual of the reduced density matrix $\rho_{\mathscr{I}}$ is not semiclassically connected to the remaining spacetime. In particular, under this definition, a spacetime with just two connected asymptotic boundaries will be disconnected whenever the entanglement wedge of each boundary is inextendible. Note that this is not a necessary condition, as evidenced by two components of two different thermofield double states.

As a quantitative example (and to further illustrate the disconnect between entanglement and spacetime emergence), we will build the canonical purification of a (two-sided) black hole before the Page time in JT gravity in Sec. 7.4.3. A single-sided example may be constructed by including an (unflavored) end-of-the-world brane, as in [61].

Returning to $|\sqrt{\rho_{\rm BH}[t < t_P]}\rangle$ and $|\sqrt{\rho_{\rm BH}[t > t_P]}\rangle$, we obtain an immediate counterexample to conjectures claiming that von Neumann entropy necessarily builds semiclassical spacetime: here, the amount of entanglement is clearly not correlated with spacetime connectivity. The spacetime can be connected – as in the post-Page time – or disconnected – as before the Page time – with the same amount of bipartite entanglement between the two sides. This gives a concrete illustration that bipartite entanglement is not always sufficient for spacetime emergence in a bipartite state, even in standard AdS/CFT. In fact, since our total state is pure, the von Neumann entropy is proportional to the reflected entropy and the mutual information, so these quantities cannot diagnose connectivity either. Note that our examples rely crucially on bulk matter entanglement: in the strictly classical case where the HRT formula applies so that $S_{\rm gen}[\varnothing] = 0$, mutual information can always be used to diagnose connectivity.

A potential complaint at this stage is that while the von Neumann entropy may be identical in the two cases, only the old black hole is maximally mixed (or more correctly, "thermally mixed"). We may thus speculate that connectedness requires our state to have a particular entanglement structure – for example, the black hole $\rho_{\rm BH}$ must be maximally mixed *in addition* to having a large amount of entropy. In other words, $S_{\rm vN}[\rho_{BH}]$ should be large and $S_{\rm vN}[\rho_\beta] - S_{\rm vN}[\rho_{\rm BH}(t)]$ should be small, where ρ_β is the thermal state at the same energy and charges as $\rho_{\rm BH}$. This modified proposal would then identify connected geometries with states that are (approximately) maximally (thermally) mixed. Prima facie, this appears to diagnose the difference in the canonical purifications before and after the Page time.

However, this turns out to be a red herring. In our example we can easily make $S_{\rm vN}[\rho_\beta] - S_{\rm vN}[\rho_{\rm BH}(t)]$ large after the Page time without altering connectedness. By throwing in a large mass excitation in a pure state from the conformal boundary (this effect can be magnified e.g. by adding a large number of bulk fields), we can increase $S_{\rm vN}[\rho_\beta]$, but since this operation can be implemented by acting with a unitary with support only on $\mathcal{H}_{\rm BH}$, $S_{\rm vN}[\rho_{\rm BH}(t)]$ is unchanged. One may worry that this operation can modify the QES in some way that alters the connectivity. This concern, however, is unfounded: since $S_{\rm vN}[\rho_{\rm BH}(t)]$ and $S_{\rm gen}[\varnothing]$ are unchanged by the addition of a pure state, we know that even after the additional excitation,

$$S_{\rm vN}[\rho'_{\rm BH}(t)] < S_{\rm gen}[\varnothing] \tag{7.42}$$

where $\rho'_{\rm BH}(t)$ is the state after the addition of the pure state. This immediately implies that even if the dominant QES shifts as a result of the additional matter (and the large number of bulk fields), it remains a nontrivial compact surface and most importantly does not revert to the empty set. Thus the connectivity of the canonically purified old black hole is robust against such modifications. Moving the state away from thermality does not alter its connectivity. [18]

7.4.2 Complexity is Not Enough

Another possibility for discerning the feature that is responsible for the disparity in spacetime connectivity between the two states is computational complexity: it has previously been suggested in work starting with [64–67] that complexity is also responsible in part of the emergence of the black hole interior. The pre-Page canonical purification is highly complex (more than one scrambling time after black hole formation), a fact that may be seen geometrically from the existence of a nonminimal QES in the interior of the entanglement wedge [52]. Could it be that exponential complexity is responsible for lack of connection? The answer appears to be no, as there certainly exist states with exponential complexity (which are also submaximally mixed), which should by any reasonable definition of connectivity be considered connected. The union of two boundaries of a three-boundary wormhole, illustrated in Fig. 7.10, is connected to the third boundary, not maximally mixed with it, and has non-minimal QESs inherited from the bifurcation surfaces. In the opposite direction, we also expect that at very early times $\rho_{\rm BH}(t)$ is not exponentially complex, but $|\sqrt{\rho_{\rm BH}(t)}\rangle$ is still disconnected. Thus a criterion based on complexity classes fails to diagnose the difference between connected and disconnected spacetimes.

7.4.3 An Instructive Counterexample

In light of the failure of bipartite entanglement (and complexity) to build spacetime in $|\sqrt{\rho_{\rm BH}(t < t_P)}\rangle$, we must ask what, then, actually builds spacetime. To address this question, we must understand the salient difference between the dual CFT states $|\sqrt{\rho_{\rm BH}(t)}\rangle$ before and after the Page time.

We will not answer this question in this section (or in this article), but rather provide an additional example that could serve as an arena to investigate these issues in more detail in

[18] See [63] for earlier work on holography of partially mixed states in the purely classical limit.

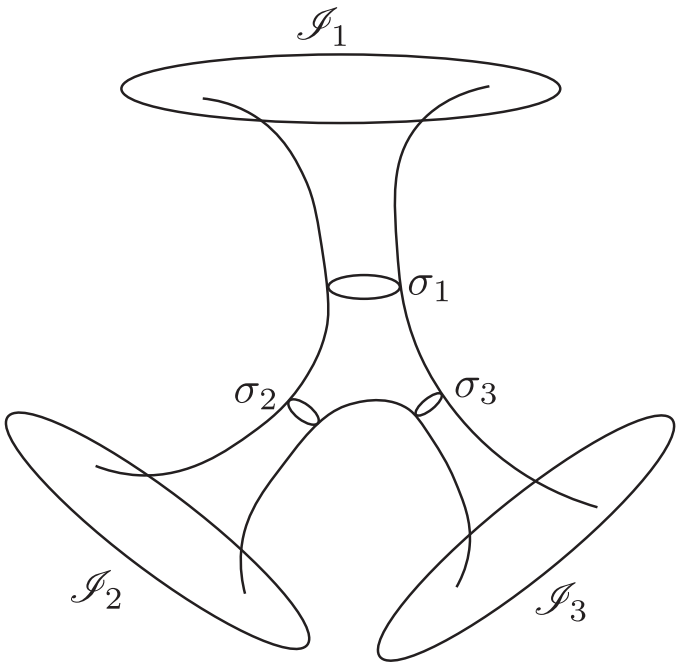


Figure 7.10: A spatial slice of three-boundary wormhole. Assuming for example that σ_1 and $\sigma_2 \cup \sigma_3$ are QESs with respect to $\mathscr{I}_1$, with $\sigma_2 \cup \sigma_3$ the minimal one, we see that a python's lunch is present, and the region bounded by the σ_i is exponentially hard to reconstruct from $\mathscr{I}_1$.

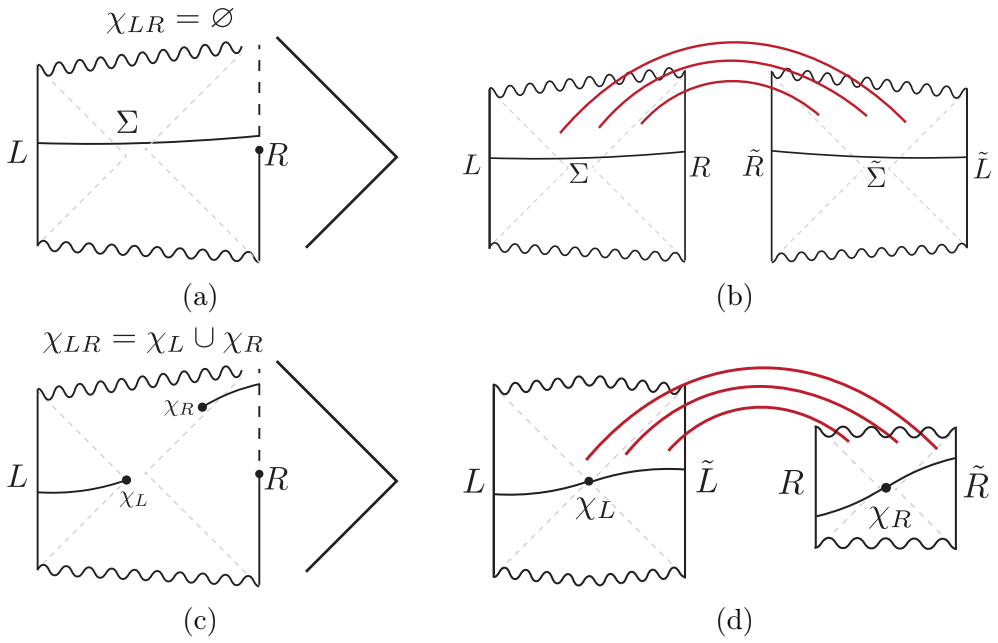


Figure 7.11: In (a) we see an evaporating two-sided black hole with a Cauchy slice Σ anchored at $t < t_P$, and in (b) we show the spacetime dual to $|\sqrt{\rho_{\text{BH,LR}}(t)}\rangle$. (c) and (d) shows the same for $t > t_P$. Red lines indicate entanglement of the bulk matter fields.

the future. We will outline the construction of this example in general and then compute it explicitly in JT gravity coupled to conformal matter.

For this construction, we work with a two-sided black hole evaporating into a reservoir as in [2], though we will not restrict ourselves to two dimensions just yet. At any time, we may execute the same procedure as before, tracing out the reservoir and canonically purifying the state $\rho_{\text{BH,LR}}$, which is the joint state of the left and right boundaries.

Prior to the Page time, the QES of $\rho_{\text{BH,LR}}(t)$ is the empty set. Canonical purification of the state generates a new two-boundary spacetime *not connected* to the original spacetime via a semiclassical wormhole. This spacetime is illustrated in Fig. 7.11a.

Let us now examine this four-boundary, two component spacetime. Label the two original boundaries L and R, and the newly generated ones $\widetilde{L}$ and $\widetilde{R}$. Before the Page time, L is connected to R, and $\tilde{L}$ to $\tilde{R}$. After the Page time however, we have that L is connected to $\tilde{L}$, and R to $\tilde{R}$. It would be interesting to study various forms of multipartite entanglement between $R, \tilde{R}, L, \tilde{L}$ to see if different properties of multipartite entanglement may be at play here;[19] what amount and type of entanglement builds spacetime, or is entanglement in fact a red herring and there is some other property of the state that is responsible for spacetime connectivity?

Let us now construct this spacetime quantitatively in two dimensions. In JT gravity it is straightforward to compute the canonically purified spacetime, and we carry it out in the appendix. Carrying out the purification at boundary Poincare time $t_\partial = f(u_\partial)$ for $u_\partial \sim \mathcal{O}(1)$,

[19] A similar setup in SYK and JT gravity was studied in [68].

the stress tensor reads, using the same conventions as in Sec. 7.3,

$$\begin{aligned} T_{x^-x^-} &= \theta(t_\partial - x^-)h(x^-) + b\delta(x^- - t_\partial), \\ T_{x^+x^+} &= \theta(t_\partial - x^+)h(x^+) + b\delta(x^+ - t_\partial), \end{aligned} \tag{7.43}$$

where

$$h(x^-) = E_S\delta(x^-) - \frac{c}{24\pi}\theta(x^-)\frac{2(\pi T_1)^2}{[1-(\pi T_1 x^-)^2]^2}, \tag{7.44}$$

where $b > 0$ is a constant depending on the microscopics of the CFT and how the absorbing boundary conditions are turning off – just like in Sec. 7.3. The dilaton reads

$$\phi = \frac{c_1 + c_2(x^+ + x^-) + c_3 x^+ x^-}{x^+ - x^-} + 8\pi G_N \frac{\theta(t_\partial - x^-)F(x^-) - \theta(t_\partial - x^+)F(x^+)}{x^+ - x^-}, \tag{7.45}$$

where the piecewise constant coefficents c_i and the function F are listed in the appendix.

7.5 Discussion

We have explicitly constructed (1) a precise procedure that generates a connected spacetime from an old black hole and its radiation without modifying the black hole system – assuming the validity of the QES formula, and (2) an example of a disconnected semiclassical spacetime whose CFT dual is a highly entangled bipartite state. We have ruled out the possibility that this counterexample could be dismissed by appealing to thermality of the old black hole state or by refining the ER=EPR proposal to refer to other entanglement measures such the reflected entropy. In particular, we have shown that no operation that leaves $\rho_{\rm BH}$ unchanged can consistently construct spacetime connectivity when the bipartite entanglement is large.

ER=EPR with bipartite entanglement: If a bipartite state $|\psi\rangle_{AB}$ has a sufficiently large von Neumann entropy, and ERBs are indeed sourced by bipartite entanglement, then there exist unitaries U_A and U_B with support strictly on A and B, respectively, such that $U_A \otimes U_B |\psi\rangle_{AB}$ has a connected semiclassical dual. In light of the earlier discussions, if this were to be true, then there would have to exist a unitary $U_{\rm BH}$ such that the state

$$\rho'_{\rm BH} = U_{\rm BH}\rho_{\rm BH}(t < t_P)U_{\rm BH}^\dagger \tag{7.46}$$

has a non-empty QES. Then with the action of $U_{\rm RAD}$, a Cauchy slice of the entanglement wedge of BH in the state $U_{\rm BH} \otimes I_{\rm RAD} |\psi\rangle$ could in principle be embedded in a larger spatial slice forming a wormhole. As a constraint on the unitaries in question, they will need to modify the topology of the dominant QES.

Multipartite entanglement for connectivity: It is clear that for multipartite states $|\psi\rangle_{\mathscr{I}_1\cdots\mathscr{I}_n}$ (or e.g. subregions), multipartite entanglement measures are necessarily for a proper understanding of the emergent connectivity of the bulk geometry – e.g. tripartite entanglement for a three-boundary wormhole. However, our counterexample is a strictly bipartite state, where we would have expected bipartite entanglement to be the definitive

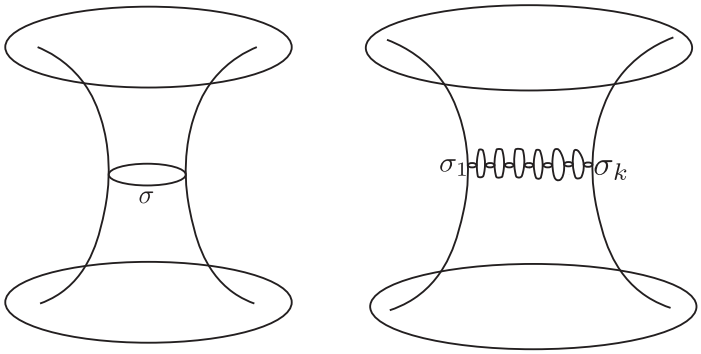


Figure 7.12: Spatial slices of two different classical two-sided wormholes, both dual to states with the same amount of entanglement between the two sides, i.e. Area$[\sigma]$ = Area$[\sigma_1 \cup \ldots \cup \sigma_k]$, where $k \sim \mathcal{O}(G_N^{-1})$. The right picture represents a potential toy model for a highly quantum ERB.

determinating quantity. Nevertheless, a key difference between $\rho_{\text{BH}}(t < t_P)$ and $\rho_{\text{BH}}(t > t_P)$ is that the former is heavily correlated with itself while the latter is not. It is in principle possible that there is some measure of these internal correlations of a system – e.g. multipartite entanglement measures – that may be responsible for the connectivity of the spacetime when the bipartite entanglement turns out to be too coarse. This appears to have some similarity with the results of [69], which showed that duals to the entanglement wedge cross-section require holographic states to have tripartite rather than primarily bipartite entanglement; however, it is not obviously the same phenomenon: the arguments of [69] relied on the reflected entropy and entanglement of purification at $\mathcal{O}(G_N^{-1})$ vs. $\mathcal{O}(1)$, whereas in our case these quantities are $\mathcal{O}(G_N^{-1})$ in both the connected and disconnected phases. Moreover, [69] relied primarily on subregions in their argument, and it is clear that subregions must have at least some non-bipartite entanglement. In our case the state is truly bipartite on two separate boundaries.

Wormhole condensation: [70] constructed a toy model for Hawking radiation where the Hawking quanta were modeled by small classical individual ERBs connected to the original black hole. This models a highly quantum wormhole as a very small ERB. The model is studied at discrete timesteps between the emission of Hawking quanta, remaining agnostic about the topology-changing process of ERB creation. It would be interesting to see if this type of toy model also could shed light on the type of quantities that can differentiate between the "quantum" and classical wormholes. For instance, we may consider two different two-sided spacetimes with the same $\mathcal{O}(G_N^{-1})$ amount of entanglement, where one has an HRT surface with one connected component of area $A \sim \mathcal{O}(1)$, while the other has an HRT surface with $\sim \mathcal{O}(G_N^{-1})$ connected components, each having an area of order $\mathcal{O}(G_N)$, but with the total area of all components adding up to A. See Fig. 7.12. The latter is a toy model of the entanglement before it is merged into semiclassical form. It would be interesting to see which observables are sensitive to this fragmentation of the total area.

Code Subspace Dependence: The fact that the entanglement wedge $\mathcal{W}_E[\rho_{\rm BH}]$ cannot be extended prior to the Page time shows, as argued above, that there is no way to leave $\rho_{\rm BH}$ unaltered while acting purely on the radiation to create a semiclassical wormhole connecting the two. However, let us pause now for a reminder that while $\mathcal{W}_E[\rho_{\rm BH}]$ is the dual to $\rho_{\rm BH}$, reconstruction of the bulk is quantum error correcting [11] and requires the specification of a nontrivial code subspace.

As shown in [71] and further elucidated in [72], state-independent reconstruction in general covers only a subset of the entanglement wedge when quantum corrections are taken into account. In particular, [71] argued that it is the entanglement wedge of the maximally mixed state in the code subspace that determines the extent of state-independent reconstruction – the so-called reconstruction wedge [72].

If we define the black hole system as the reconstruction wedge of $\rho_{\rm BH}$ in a large code subspace consisting of all states with a smooth horizon, we will in fact find that when the black hole is within the adiabatic regime, the wedge is bounded by a nontrivial QES. In this case, the canonical purification of the maximally mixed state of $\mathcal{H}_{\rm code}$ is connected. The potential relevance of the choice of code subspace and typicality for ER=EPR had been previously noted in [73], which also clarified the arrow of implication from spacetime connectivity to a high amount of entanglement (conversely with our construction, which illustrates a subtlety into the arrow of implication from EPR to ER).

Experience of an infalling observer: The experience of an observer falling through the black hole after the Page time can be causally affected by turning on some local unitary in the canonical purification. This gives an explicit realization of action on the island via the radiation, and specifically realizes the map between Marolf-Wall [56] and the relation between the radiation and the black hole interior. Note that by extension, this is also an explicit mapping from the factorization problem due to replica wormholes in the island picture to the standard factorization problem in AdS/CFT [56, 74–76] of multiboundary geometries in AdS/CFT.

In light of these observations, it is tempting to think of the map $U_{\rm RAD}$ from $\rho_{\rm RAD}$ to $\widetilde{\rho_{\rm BH}}$ as a decoding unitary. The state $\widetilde{\rho_{\rm BH}}$ is simple, and thus $U_{\rm RAD}$ may be thought of as mapping a high complexity state to a simply reconstructible state. However, it is not a particularly useful 'decoding' map: the map $U_{\rm RAD}$ is dependent on which operators we choose to turn on in the island. Nevertheless, it would be interesting to better understand the properties of this map.

Acknowledgments

It is a pleasure to thank Sebastian Fischetti, Sergio Hernández-Cuenca, Arjun Kar, Juan Maldacena, and Arvin Shahbazi-Moghaddam for comments on an earlier draft of this manuscript. We are also grateful to Jan de Boer, Sebastian Fischetti, Ben Freivogel, Sergio Hernández-Cuenca, Daniel Jafferis, Arjun Kar, Sam Leutheusser, Henry Lin, Hong Liu, Don Marolf, Geoff Penington, Lisa Randall, Arvin Shahbazi-Moghaddam, Shreya Vardhan, Erik Verlinde, and Herman Verlinde for valuable conversations. This work is supported in part by the MIT department of physics. NE is supported in part by NSF grant no. PHY-2011905, by the

U.S. Department of Energy Early Career Award DE-SC0021886, by the U.S. Department of Energy grant DE-SC0020360 (Contract 578218), by the John Templeton Foundation and the Gordon and Betty Moore Foundation via the Black Hole Initiative, and by funds from the MIT department of physics. The work of ÅF is also supported in part by an Aker Scholarship and by NSF grant no. PHY-2011905. This work was initially started at the KITP in the Gravitational Holography workshop, supported in part by the National Science Foundation under Grant No. PHY-1748958.

7.6 Appendix

7.6.1 JT gravity junction conditions

The JT equations of motion with matter reads $\tilde{G}_{ab}(\phi) = 8\pi G_N T_{ab}$, where

$$\tilde{G}_{ab} = -\nabla_a\nabla_b\phi + g_{ab}\nabla^2\phi - g_{ab}\phi \tag{7.47}$$

is the effective dilaton "Einstein tensor". Consider now gluing two spacetimes M^+ and M^- along a codimension-1 null junction S generated by a null vector k^a. Let M^+ be to the future of the junction and M^- to the past, and define the discontinuity of some quantity A across the junction as

$$[A] = A_+ - A_-. \tag{7.48}$$

Let now $\ell^a = \left(\frac{\mathrm{d}}{\mathrm{d}\lambda}\right)^a$ be a rigging null field with $k\cdot\ell = -1$ on S. Firing geodesics along ℓ^a, the parameter λ can be viewed to a scalar field on spacetime in a neighbourhood of the junction, chosen so that $\lambda = 0$ on S. Using this, we can in a neighbourhood of the junction write

$$\phi = \theta(\lambda)\phi^+ + \theta(-\lambda)\phi^-. \tag{7.49}$$

Assume that $[\phi] = 0$. This is the analogue of the first junction condition and is required in order to have a distributional Einstein tensor. We then find that

$$\begin{aligned} \nabla_a\phi &= \theta(\lambda)\nabla_a\phi^+ + \theta(-\lambda)\nabla_a\phi^-, \\ \nabla_b\nabla_a\phi &= \nabla_{(b}\nabla_{a)}\phi = \delta(\lambda)\nabla_{(b}\lambda[\nabla_{a)}\phi] + \text{non-singular}. \end{aligned} \tag{7.50}$$

Thus the effective Einstein tensor gets a singular part

$$\tilde{G}_{ab}|_{\text{sing}} = \delta(\lambda)\left(-\nabla_{(a}\lambda[\nabla_{b)}\phi] + g_{ab}\nabla^c\lambda[\nabla_c\phi]\right). \tag{7.51}$$

By definition we have $\mathrm{d}x^\mu = \ell^\mu\mathrm{d}\lambda$. Contracting this with k_μ we get

$$k_\mu\mathrm{d}x^\mu = -\mathrm{d}\lambda = -\nabla_\mu\lambda\mathrm{d}x^\mu \quad \Rightarrow \quad \nabla_a\lambda = -k_a, \tag{7.52}$$

giving that

$$\tilde{G}_{ab} = \delta(\lambda)\left(k_{(a}[\nabla_{b)}\phi] - g_{ab}[k^c\nabla_c\phi]\right), \tag{7.53}$$

or

$$8\pi G_N T_{ab}\ell^a\ell^b = -\delta(\lambda)\left[\ell^a\nabla_a\phi\right]. \tag{7.54}$$

7.6.2 Computing the Post Page Time Canonical Purification in JT Gravity

Computing the dilaton on Σ

We want the dilaton on a late time slice Σ running between the QES and the conformal boundary. The approach to compute this is the same as in [2], except we have to relax the assumption $|t_\infty - x^+| \ll |t_\infty - x^-|$, which holds near the QES but not near the conformal boundary, and thus not on Σ in general. It is however still true that $\frac{t_\infty - x^\pm}{t_\infty}$ is non-perturbatively small. To make the derivation clearer, we retrace many steps from [2].

The general solution for the dilaton coupled to matter with $T_{x^+x^-} = 0$ reads

$$\begin{aligned}
\phi &= \frac{c_1 + c_2(x^+ + x^-) + c_3 x^+ x^- + 8\pi G_N (I^+ + I^-)}{x^+ - x^-}, \\
I^+ &= \int_{x_0^+}^{x^+} \mathrm{d}s (s - x^+)(s - x^-) T_{x^+x^+}(s), \\
I^- &= -\int_{x_0^-}^{x^-} \mathrm{d}s (s - x^+)(s - x^-) T_{x^-x^-}(s).
\end{aligned} \tag{7.55}$$

The boundary trajectory $f(u)$, which will be useful below, can be found through an analysis of the energy of the spacetime, whose value and rate of change is given by

$$E(u) = -\frac{\bar{\phi}_r}{8\pi G_N}\{f(u), u\}, \tag{7.56}$$

$$\partial_u E(u) = f'(u)^2 \left[T_{x^-x^-}(u) - T_{x^+x^+}(u)\right]. \tag{7.57}$$

Equation (7.57) together with the stress tensor (7.14) lets us solve for $E(u)$. Next, this solution (given in (7.15)) together with (7.56) can be used to solve for $f(u)$.

Now let us work out the dilaton. First, from (7.13), (7.14) and (7.55), we have that the dilaton to the future of the shock reads

$$\frac{x^+ - x^-}{2\bar{\phi}_r}\phi = 1 - (\pi T_1)^2 x^+ x^- + \frac{k}{2}\int_0^{x^-} \mathrm{d}s (s - x^+)(s - x^-)\{f^{-1}(s), s\}, \tag{7.58}$$

giving

$$\frac{(x^+ - x^-)^2}{2\bar{\phi}_r}\partial_{x^\pm}\phi = \mp\left[1 - (\pi T_1)^2 (x^\mp)^2 + \frac{1}{2}k\int_0^{x^-} \mathrm{d}s (s - x^\mp)^2 \{f^{-1}(s), s\}\right]. \tag{7.59}$$

Let us now approach the endpoint of the conformal boundary at fixed $x^- = t_\infty$, meaning $x^+ \to t_\infty$. This endpoint is characterized by $\phi = 0$, and so in order for $\partial_{x^+}\phi$ to stay finite there, we must have

$$I_\infty \equiv \int_0^{t_\infty} \mathrm{d}s (s - x^-)^2 \{f^{-1}(s), s\} = \frac{2}{k}\left[(\pi T_1 t_\infty)^2 - 1\right]. \tag{7.60}$$

Let us for later convenience note that

$$\pi T_1 = \frac{1}{t_\infty}\left[1 + \frac{k}{4}t_\infty\right] + \mathcal{O}(k^2), \tag{7.61}$$

giving

$$\begin{aligned} 1 - (\pi T_1)^2(x^\pm)^2 + \frac{k}{2}I_\infty &= (\pi T_1)^2\left[t_\infty - (x^\pm)^2\right] \\ &= \frac{1}{t_\infty^2}\left[1 + t_\infty\frac{k}{2}\right](t_\infty - x^\pm)2t_\infty + \mathcal{O}\left((t_\infty - x^\pm)^2\right) + \mathcal{O}(k^2) \\ &= [2 + kt_\infty]\,\frac{t_\infty - x^\pm}{t_\infty} + \mathcal{O}\left((t_\infty - x^\pm)^2\right) + \mathcal{O}(k^2). \end{aligned} \tag{7.62}$$

From now on we implicitly drop terms of order $\mathcal{O}(k^2)$ and $\mathcal{O}\left((t_\infty - x^\pm)^2\right)$.

Let us now first compute $\partial_+\phi$. Using (7.18) we have

$$\begin{aligned} \partial_- \int_0^{x^-} \mathrm{d}s(s - x^-)^2\{f^{-1}(s), s\} &= \mathcal{O}(1) + \int^{x^-} \mathrm{d}s\frac{x^- - s}{(s - t_\infty)^2} \\ &= -\gamma_1 - \log\left(\frac{t_\infty - x^-}{t_\infty}\right) + \mathcal{O}\left(t_\infty - x^-\right) \end{aligned} \tag{7.63}$$

where $-\gamma_1$ is just some unknown constant which we have parametrized by γ_1 for later convenience (note that here and in the following, $\mathcal{O}(1)$ means $\mathcal{O}\left([t_\infty - x^\pm]^0\right)$). We then have

$$\begin{aligned} \int_0^{x^-} \mathrm{d}s(s - x^-)^2\{f^{-1}(s), s\} = I_\infty + \gamma_1(t_\infty - x^-) + (t_\infty - x^-)\log\left(\frac{t_\infty - x^-}{t_\infty}\right) \\ + \mathcal{O}\left(\left[t_\infty - x^-\right]^2\right). \end{aligned} \tag{7.64}$$

Inserting (7.64) and (7.62) into (7.59) then gives

$$\partial_{x^+}\phi \approx -\frac{\bar{\phi}_r}{(x^+ - x^-)^2}\left\{[4 + kt_\infty(2 + \gamma_1)]\,\frac{t_\infty - x^-}{t_\infty} + k(t_\infty - x^-)\log\left(\frac{t_\infty - x^-}{t_\infty}\right)\right\}. \tag{7.65}$$

Next we turn to $\partial_{x^-}\phi$. Expanding $\int_0^{x^-} \mathrm{d}s(s - x^+)^2\{f^{-1}(s), s\}$ about $x^+ = t_\infty$ in (7.59), we get

$$\frac{(x^+ - x^-)^2}{2\bar{\phi}_r}\partial_-\phi = 1 - (\pi T_1)^2(x^+)^2 + \frac{1}{2}k\mathcal{I} \tag{7.66}$$

where

$$\begin{aligned} \mathcal{I} &= I_\infty + \mathcal{I}_0 + (x^+ - t_\infty)\mathcal{I}_1 + (x^+ - t_\infty)^2\mathcal{I}_2, \\ \mathcal{I}_0 &= -\int_{x^-}^{t_\infty} \mathrm{d}s(s - t_\infty)^2\{f^{-1}(s), s\}, \\ \mathcal{I}_1 &= 2\int_0^{x^-} \mathrm{d}s(t_\infty - s)\{f^{-1}(s), s\}, \\ \mathcal{I}_2 &= \int_0^{x^-} \mathrm{d}s\{f^{-1}(s), s\}. \end{aligned} \tag{7.67}$$

Using now again the late time Schwarzian, we get the leading behavior

$$\begin{aligned}\mathcal{I}_0 &= -\frac{1}{2}(t_\infty - x^-)\\ \mathcal{I}_1 &\sim \mathcal{O}(1) + \int^{x^-} \mathrm{d}s \frac{1}{(t_\infty - s)} = -\log\left(\frac{t_\infty - x^-}{t_\infty}\right) + \mathcal{O}(1)\\ \mathcal{I}_2 &\sim \mathcal{O}(1) + \int^{x^-} \mathrm{d}s \frac{1}{2(s - t_\infty)^2} = \frac{1}{2(t_\infty - x^-)} + \mathcal{O}(1)\end{aligned} \tag{7.68}$$

We want to keep the $\mathcal{O}(1)$ term in $\mathcal{I}_1$, but we do not need the $\mathcal{O}(1)$ term in $\mathcal{I}_2$ since it contributes as $\mathcal{O}\left([t_\infty - x^+]^2\right)$ to $\mathcal{I}$, which is higher order than what we work at. Let us define γ_2 through

$$\mathcal{I}_1 = -\frac{1}{2}\log\left(\frac{t_\infty - x^-}{t_\infty}\right) - \gamma_2 + \mathcal{O}(t_\infty - x^-). \tag{7.69}$$

Then, assembling everything, we find

$$\begin{aligned}\frac{(x^+ - x^-)^2}{\bar{\phi}_r}\partial_- \phi = [4 + k t_\infty (2 + \gamma_2)]\,\frac{t_\infty - x^+}{t_\infty} + k\frac{(x^+ - t_\infty)^2}{2(t_\infty - x^-)} + k(t_\infty - x^+)\log\left(\frac{t_\infty - x^-}{t_\infty}\right)\\ - \frac{k}{2}(t_\infty - x^-)\end{aligned} \tag{7.70}$$

Integrating (7.66) and (7.70) gives

$$\begin{aligned}\frac{\phi}{\bar{\phi}_r} &= A(x^-) + \frac{t_\infty - x^-}{x^+ - x^-}\left[\mathcal{C}_1 + k\log\left(\frac{t_\infty - x^-}{t_\infty}\right)\right]\\ \frac{\phi}{\bar{\phi}_r} &= B(x^+) + \frac{k}{2}\log\left(\frac{t_\infty - x^-}{t_\infty}\right) + \frac{t_\infty - x^+}{x^+ - x^-}\left[\mathcal{C}_2 + k\log\left(\frac{t_\infty - x^-}{t_\infty}\right)\right]\end{aligned} \tag{7.71}$$

for some unknown functions A, B, where we define

$$\mathcal{C}_i \equiv \frac{4 + k t_\infty (2 + \gamma_i)}{t_\infty}. \tag{7.72}$$

Taking the difference we find

$$0 = A(x^-) - B(x^+) + \frac{k}{2}\log\left(\frac{t_\infty - x^-}{t_\infty}\right) + \frac{(\mathcal{C}_1 - \mathcal{C}_2)t_\infty + \mathcal{C}_1 x^+ - \mathcal{C}_2 x^-}{x^+ - x^-} \tag{7.73}$$

This equation is inconsistent if we cannot break up each term into one depending purely on x^+ and one depending only on x^-, and so we must have that $\mathcal{C}_1 = \mathcal{C}_2 \equiv \mathcal{C}$, leading to $\gamma_1 = \gamma_2$. Then we get the solution

$$\begin{aligned}B(x^+) &= \text{constant} \equiv B\\ A(x^-) &= B - \frac{k}{2}\log\left(\frac{t_\infty - x^-}{t_\infty}\right) + \frac{4 + 2k}{t_\infty}\end{aligned} \tag{7.74}$$

and so we get

$$\phi = \frac{\bar{\phi}_r}{x^+ - x^-}\left[\mathcal{C}(t_\infty - x^+) + B(x^+ - x^-) + \frac{k}{2}(2t_\infty - x^+ - x^-)\log\left(\frac{t_\infty - x^-}{t_\infty}\right)\right]. \quad (7.75)$$

Now, plugging this back into the $+-$ component of the dilaton equations of motion, which reads

$$\partial_{x^+x^-}\phi + \frac{2\phi}{(x^+ - x^-)^2} = 0, \quad (7.76)$$

we find that we must have

$$B = \frac{\mathcal{C} - k}{2}. \quad (7.77)$$

giving finally (7.20). As a double check, we find that when plugging (7.20) back into the dilaton equations of motion, we get that the solution is indeed sourced by (7.19).

Transforming to global coordinates

(7.22) and some basic algebra gives[20]

$$x^\pm = -\lambda + \mu^{-1} f_\pm(w^\pm), \qquad f_\pm(z) \equiv \tan(z - \tau_0) \pm \frac{1}{\cos(z - \tau_0)}, \quad (7.78)$$

$$\frac{\partial w^\pm}{\partial x^\pm} = \mu\left[1 \mp \sin\left(w^\pm - \tau_0\right)\right]. \quad (7.79)$$

As indicated in the main text, we want to solve

$$\tau(z_*, t_*) = \tau(0, t_\partial) = \rho(z_*, t_*) = 0, \quad (7.80)$$

for λ, μ, τ_0, where (t_*, z_*) is the QES position in Poincare coordinates.

Through (7.22), (7.80) translates to

$$\begin{aligned}(t_* + \lambda)\mu &= -\tan\tau_0, \\ [f(u_\partial) + \lambda]\mu &= -\frac{\sin\tau_0}{1 + \cos\tau_0}, \\ z_*\mu &= \frac{1}{\cos\tau_0}.\end{aligned} \quad (7.81)$$

Next, we know the QES location to leading order is given by

$$\begin{aligned}x_*^+ &= t_\infty + \frac{1}{3}\left[t_\infty - f(u_\partial)\right], \\ x_*^- &= t_\infty - \frac{8\pi T_1 e^{-kf^{-1}(x_*^-)/2}}{3k}\left[t_\infty - f(u_\partial)\right].\end{aligned} \quad (7.82)$$

[20] Solving for $x^\pm(w^\pm)$ from (7.22) also gives another branch, but with our chosen value of τ_0 the given one will be the one where $x^\pm(w^\pm)$ is continuous in the right Poincare patch.

Solving now (7.81) and inserting (7.82), we find

$$\chi \equiv \mu(t_\infty + \lambda) = -\frac{1}{4} + \frac{5e^{-kf^{-1}(x_*^-)/2}k}{8\pi T_1} + \mathcal{O}(k^2), \tag{7.83}$$

$$\tau_0 = \frac{\pi}{2} - \frac{ke^{-kf^{-1}(x_*^-)}}{\pi T_1} + \mathcal{O}(k^2), \tag{7.84}$$

$$\mu = \frac{1}{t_\infty - f(u_\partial)}\left(\frac{3}{4} - \frac{3ke^{kf^{-1}(x_*^-)/2}}{8\pi T_1} + \mathcal{O}(k^2)\right). \tag{7.85}$$

Next, from [2] we have at the QES that

$$f^{-1}(x_*^-(u)) \approx u - \frac{e^{ku/2}}{2\pi T_1}\log\left(\frac{8\pi T_1 e^{-ku/2}}{3k}\right), \tag{7.86}$$

which gives that

$$e^{kf^{-1}(x_*^{-1})/2} = e^{ku_\partial/2}\left[1 + \mathcal{O}(k\log k)\right], \tag{7.87}$$

giving finally (7.24), (7.25) and (7.26).

The stress-tensor initial boundary value problem

Consider general initial data for the CFT stress tensor in global AdS on the $\tau = 0$ slice:

$$\begin{aligned} T_{w^-w^-}|_{\tau=0} &= g(\rho) = g(w^-)|_\Sigma, \\ T_{w^+w^+}|_{\tau=0} &= h(-\rho) = h(w^+)|_\Sigma. \end{aligned} \tag{7.88}$$

Before we even specify boundary conditions, the fact that solutions $T_{w^-w^-}, T_{w^+w^+}$ are functions of only w^- and w^+, respectively, imposes the following boundary values

$$\begin{aligned} T_{w^-w^-}|_{\partial M_R} &= g(w^-) = g\left(\tau + \frac{\pi}{2}\right), && \tau \in \left[-\frac{\pi}{2}, 0\right], \\ T_{w^-w^-}|_{\partial M_L} &= g(w^-) = g\left(\tau - \frac{\pi}{2}\right), && \tau \in \left[0, \frac{\pi}{2}\right], \\ T_{w^+w^+}|_{\partial M_R} &= h(w^+) = h\left(\tau - \frac{\pi}{2}\right), && \tau \in \left[0, \frac{\pi}{2}\right], \\ T_{w^+w^+}|_{\partial M_L} &= h(w^+) = h\left(\tau + \frac{\pi}{2}\right), && \tau \in \left[-\frac{\pi}{2}, 0\right], \end{aligned} \tag{7.89}$$

where $\partial M_{L/R}$ are the left/right conformal boundaries. This is obtained simply by tracing null rays from the initial data to the boundary, holding the relevant stress-tensor component fixed. Now let us consider reflection boundary conditions

$$\left(T_{w^-w^-} - T_{w^+w^+}\right)|_{\partial M} = 0. \tag{7.90}$$

The combination of (7.89) and (7.90) now determines the boundary values of $T_{w^\pm w^\pm}$ for all $\tau \in \left[-\frac{\pi}{2}, \frac{\pi}{2}\right]$:

$$\begin{aligned} T_{w^-w^-}|_{\partial M_R} = T_{w^+w^+}|_{\partial M_R} &= \theta(-\tau)g\left(\tau + \frac{\pi}{2}\right) + \theta(\tau)h\left(\tau - \frac{\pi}{2}\right) + b_R\delta(\tau), \\ T_{w^-w^-}|_{\partial M_L} = T_{w^+w^+}|_{\partial M_L} &= \theta(\tau)g\left(\tau - \frac{\pi}{2}\right) + \theta(-\tau)h\left(\tau + \frac{\pi}{2}\right) + b_L\delta(\tau), \end{aligned} \tag{7.91}$$

Note that we have discontinuities at $\tau = 0$, which stems from the fact that our initial data does not satisfy (7.90) at $\tau = 0$ for general g and h. Due to the discontinuous stress tensors at $\tau = 0$ there is no way to rule out a δ-function shock at $\tau = 0$, and so this must be allowed. Altering the boundary conditions in AdS causes the injection of energy [2], and since we are considering the scenario where we turn off absorbing boundary condition very fast to the future and the past, we will have that b_R, b_L are nonzero. The specific value of b_L and b_R will depend on the specifics of the theory and the state, and so they cannot be determined at our level of analysis. However, in order to have that the spacetime energy is constant, we must have that that the strength w^+ and w^- shocks on each boundary is the same, which we used in (7.91).

Rewriting these expressions in terms of $w^\pm$ we finally obtain the full solution for all $|w^\pm| \leq \pi$:

$$\begin{aligned}
T_{w^-w^-} &= \theta\left(-w^- + \frac{\pi}{2}\right) g\left(w^-\right) + \theta\left(w^- - \frac{\pi}{2}\right) h\left(w^- - \pi\right) + b_R \delta\left(w^- - \frac{\pi}{2}\right), \quad w^- \in [0, \pi] \\
T_{w^-w^-} &= \theta\left(w^- + \frac{\pi}{2}\right) g\left(w^-\right) + \theta\left(-w^- - \frac{\pi}{2}\right) h\left(w^- + \pi\right) + b_L \delta\left(w^- + \frac{\pi}{2}\right), \quad w^- \in [-\pi, 0] \\
T_{w^+w^+} &= \theta\left(w^+ + \frac{\pi}{2}\right) h\left(w^+\right) + \theta\left(-w^+ - \frac{\pi}{2}\right)\ g\left(w^+ + \pi\right) + b_R \delta\left(w^+ + \frac{\pi}{2}\right), \quad w^+ \in [-\pi, 0] \\
T_{w^+w^+} &= \theta\left(-w^+ + \frac{\pi}{2}\right) h\left(w^+\right) + \theta\left(w^+ - \frac{\pi}{2}\right) g\left(w^+ - \pi\right) + b_L \delta\left(w^+ - \frac{\pi}{2}\right), \quad w^+ \in [0, \pi] .
\end{aligned} \tag{7.92}$$

In the domain covered by $|w^\pm| \leq \pi$, this represents the distributional evolution that agrees (1) agrees with our initial data everywhere in the domain of the dependence, and (2) has reflecting boundary conditions in the weak sense. Note also that in the case relevant for us, where Σ has CPT symmetry about the QES, we must have $b_L = b_R$. Combining (7.35) with (7.92) gives (7.37).

Evolving the dilaton

Thanks to the CPT symmetry, it is sufficient to restrict to the region in the future and spacelike to the right of the QES, which is covered by $w^- \geq 0$. Thus, in what follows we always work in the range $0 \leq w^- \leq \pi$.

Changing to global coordinates in (7.20) gives the dilaton in $D[\Sigma]$ (i.e. for $w^+ < 0, w^- > 0$) on the form:

$$\begin{aligned}
\phi(w^+, w^-)|_{D[\Sigma]} = \frac{\bar{\phi}_r}{2(f_+ - f_-)} \Big[&(2\chi - f_- - f_+)\mathcal{C} - k(f_+ - f_-) \\
&+ k(2\chi - f_+ - f_-) \log\left(\frac{\chi - f_-}{t_\infty \mu}\right) \Big],
\end{aligned} \tag{7.93}$$

Now we compute $\hat{I}^-(w^+, w^-)$ for $0 < w^- < \frac{\pi}{2}$. The integral can be done analytically and

reads

$$\begin{aligned} 8\pi G_N \hat{I}^- &= -8\pi G_N \int_\varepsilon^{w^-} \mathrm{d}s \left[1+\sin(s-\tau_0)\right]\left[f_-(s)-f_+(w^+)\right]\left[f_-(s)-f_-(w^-)\right] g(s) \\ &= \frac{1}{2}k\bar{\phi}_r\left[\frac{2}{k\delta\tau_0}+(\chi+f_--f_+)+(2\chi-f_+-f_-)\log\left(\frac{(\chi-f_-)k\delta\tau_0}{2}\right)\right] \\ &\equiv H_1(w^+,w^-), \end{aligned} \tag{7.94}$$

where we introduced the notation

$$\tau_0 = \frac{\pi}{2} - k\delta\tau_0 + \mathcal{O}(k^2). \tag{7.95}$$

$\hat{I}^+$ vanishes on $D[\Sigma]$, and so we can write (7.93)

$$\begin{aligned} \phi &= \phi_0 + \phi_m, \\ \phi_m &\equiv \frac{H_1(w^+,w^-)}{f_+-f_-}, \\ \phi_0 &\equiv \phi - \phi_m = \frac{c_1^{(0)}+c_2^{(0)}(f_++f_-)+c_3^{(0)}f_+f_-}{f_+-f_-}, \end{aligned} \tag{7.96}$$

with

$$\begin{aligned} c_1^{(0)} &= \bar{\phi}_r\left[\mathcal{C}\chi - \frac{1}{\delta\tau_0} - \frac{k}{2}\chi - k\chi\log\left(\frac{k\delta\tau_0\mu}{2}\right)\right], \\ c_2^{(0)} &= -\frac{1}{2}\bar{\phi}_r\left[\mathcal{C} - k\log\left(\frac{k\delta\tau_0\mu}{2}\right)\right], \\ c_3^{(0)} &= 0. \end{aligned} \tag{7.97}$$

Now let us evolve out of $D[\Sigma]$. Let's focus first on the part of the dilaton sourced by continuous stress-energy, focusing on the δ-function contributions afterwards. We can ignore the ε for this, as they are only there to make sure we are not lying exactly on top of a shock.

Thanks to the step-function in (7.37), $\hat{I}^-$ for $\frac{\pi}{2} < w^- < \pi$:

$$\begin{aligned} 8\pi G_N \hat{I}^-|_{\text{cont}} &= -8\pi G_N \int_0^{\pi/2} \mathrm{d}s \left[1+\sin(s-\tau_0)\right]\left[f_-(s)-f_+(w^+)\right]\left[f_-(s)-f_-(w^-)\right] g(s) \\ &= \frac{k\bar{\phi}_r}{2}\Big[\frac{2}{k\delta\tau_0} + \frac{-1+f_+f_- - (1+f_++f_-)\chi+\chi^2}{1+\chi} \\ &\qquad\qquad + (2\chi-f_+-f_-)\log\left(\frac{k\delta\tau_0(1+\chi)}{2}\right)\Big] \\ &= H_2(w^+,w^-). \end{aligned} \tag{7.98}$$

Next lets turn to the continuous part of $\hat{I}^+$, which becomes nonzero either when $\hat{w}^+ \geq \frac{\pi}{2}$

or $\hat{w}^+ \leq -\frac{\pi}{2}$. Consider first the latter range. We find, ignoring $\mathcal{O}(k^2)$ corrections,

$$\begin{aligned}
8\pi G_N \hat{I}^+|_{\text{cont}} &= 8\pi G_N \int_{-\frac{\pi}{2}}^{w^+} \mathrm{d}s\, [1-\sin(s-\tau_0)] \left[f_+(s)-f_+(w^+)\right] \left[f_+(s)-f_-(w^-)\right] g(s+\pi) \\
&= \frac{\bar{\phi}_r k}{2} \left[(f_+ + f_- - 2\chi) \log\left|\chi - \tan\frac{s}{2}\right| - \frac{(f_- - \chi)(f_+ - \chi)}{\chi\left(\chi\cos\frac{s}{2} - \sin\frac{s}{2}\right)} - \tan\frac{s}{2} \right] \Bigg|_{s=-\frac{\pi}{2}}^{s=w^+} \\
&\equiv H_3(w^+, w^-).
\end{aligned} \tag{7.99}$$

Next, in the range $w^+ \geq \frac{\pi}{2}$ we find

$$\begin{aligned}
\hat{I}^+|_{\text{cont}} &= \int_{\frac{\pi}{2}}^{w^+} \mathrm{d}s\, [1-\sin(s-\tau_0)] \left[f_+(s)-f_+(w^+)\right] \left[f_+(s)-f_-(w^-)\right] g(s-\pi) \\
&= \frac{\bar{\phi}_r k}{2} \left[(f_+ + f_- - 2\chi) \log\left|\chi - \tan\frac{s}{2}\right| - \frac{(f_- - \chi)(f_+ - \chi)}{\chi\left(\chi\cos\frac{s}{2} - \sin\frac{s}{2}\right)} - \tan\frac{s}{2} \right] \Bigg|_{s=\frac{\pi}{2}}^{s=w^+} \\
&\equiv H_4(w^+, w^-).
\end{aligned} \tag{7.100}$$

Next we turn to the contribution of the shocks. Let us start with $\hat{I}^-$. We get a jump as we cross $w^- = \frac{\pi}{2}$:

$$\begin{aligned}
8\pi G_N \hat{I}^-|_{\text{shock}} &= -8\pi G_N \times \\
&\int_{\epsilon}^{w^-} \mathrm{d}s\, [1+\sin(s-\tau_0)] \left[f_-(s)-f_+(w^+)\right] \left[f_-(s)-f_-(w^-)\right] b\delta\left(s-\frac{\pi}{2}\right) \\
&= -8\pi G_N b\theta\left(w^- - \frac{\pi}{2}\right)(1+f_+)(1+f_-),
\end{aligned} \tag{7.101}$$

where we neglect non-perturbatively small corrections. Next, consider w^+. Here we have three ranges of interest. By a similar computation as above we have

$$\hat{I}^+\left(w^+ < -\frac{\pi}{2}\right)|_{\text{shock}} = 8\pi G_N b(1+f_+)(1+f_-), \tag{7.102}$$

$$\hat{I}^+\left(0 < w^+ < \frac{\pi}{2}\right)|_{\text{shock}} = 16\pi G_N a_+ f_+ f_-, \tag{7.103}$$

$$\hat{I}^+\left(\frac{\pi}{2} < w^+ < \pi\right)|_{\text{shock}} = 16\pi G_N a_+ f_+ f_- + 8\pi G_N b(f_+ - 1)(f_- - 1) \tag{7.104}$$

These contributions to ϕ can most conveniently be included by modifying the c_1, c_2, c_3 coefficients from earlier to be step functions:

$$\begin{aligned}
c_1 &= c_1^{(0)} + 8\pi G_N b\left[-\theta\left(w^- - \frac{\pi}{2}\right) + \theta\left(-\frac{\pi}{2} + w^+\right) + \theta\left(w^+ - \frac{\pi}{2}\right)\right], \\
c_2 &= c_2^{(0)} + 8\pi G_N b\left[-\theta\left(w^- - \frac{\pi}{2}\right) + \theta\left(-\frac{\pi}{2} + w^+\right) - \theta\left(w^+ - \frac{\pi}{2}\right)\right], \\
c_3 &= c_3^{(0)} + 8\pi G_N b\left[-\theta\left(w^- - \frac{\pi}{2}\right) + \theta\left(-\frac{\pi}{2} + w^+\right) + \theta\left(w^+ - \frac{\pi}{2}\right)\right] + 16\pi G_N a_+ \theta(w^+).
\end{aligned} \tag{7.105}$$

This completes the determination of the dilaton.

Bounding b

Let us now find the leading order bound on b by demanding that the QES is causally disconnected from the conformal boundary. This means that we need that the boundary endpoints of the physical conformal boundary lies in the interval $|\tau| \leq \frac{\pi}{2}$. Consider the future boundary, which we thus require to lie in the region $\frac{\pi}{2} \leq w^-$, $-\pi < w^+ < \frac{\pi}{2}$. Evaluating the dilaton on the boundary where $f_+ = f_-$, we get to leading order that

$$\begin{aligned}(f_+ - f_-)\phi|_{\partial M_R} &= c_1 + 2c_2 f_+ + c_3 f_+^2 + H_2 \\ &= -8\pi G_N (1+f_+)^2 - 4\pi T_1 \bar{\phi}_r e^{-u_\partial/2} f_+ + \mathcal{O}(k \log k),\end{aligned} \tag{7.106}$$

where $w^+ = \tau - \frac{\pi}{2}$. In order for the physical boundary to terminate in the future some $\tau \leq \pi/2$ we need

$$-8\pi G_N b \left[1 + f_+ (\tau - \pi/2)\right]^2 - 4 t_\infty^{-1} \bar{\phi}_r e^{-u_\partial/2} f_+ (\tau - \pi/2) = 0 \tag{7.107}$$

has a solution in the range $0 < \tau < \frac{\pi}{2}$. The above can be simplified to

$$\tan^2 \tau = \frac{c e^{-u_\partial k} T_1}{6\pi b^2 k^2 t_\infty} \left(b e^{u_\partial k/2} k + \frac{c}{24\pi t_\infty} \right). \tag{7.108}$$

This only has a solution if

$$b > -\frac{c}{24\pi} \frac{e^{-u_\partial k/2}}{t_\infty k} = -2a_-. \tag{7.109}$$

A similar computation for $\tau < 0$ gives the slightly stronger bound

$$b > -\frac{c}{30\pi} \frac{e^{-u_\partial k/2}}{t_\infty k}. \tag{7.110}$$

7.6.3 Computing the Pre Page Time Canonical Purification in JT Gravity

Let us now build the canonical purification before the Page time, purifying at some specific boundary time $u = u_\partial$. For analytical convenience we work at times of order $u_\partial \sim \mathcal{O}(1)$, but the general picture is also valid shortly before the Page time.

The stress tensor on the domain of dependence for our initial data slice Σ is given by (7.14). Since we do not keep track of $O(k^2)$ corrections, we can use the early time expression

$$f(u) = \frac{1}{\pi T_1} \tanh(\pi T_1 u). \tag{7.111}$$

We can see this by noting that (1) the expression $\hat{f}(u) \equiv (\pi T_1)^{-1} \tanh\left[\frac{2\pi T_1}{k}(1 - e^{-ku/2})\right]$ solves the differential equation for $f(u)$ to order $O(k^2)$,[21] and (2) up to $O(k^2)$ corrections, (7.111) and $\hat{f}(u)$ gives the same value for $\{f^{-1}(x), x\}$. Inverting for f^{-1} and computing the Schwarzian derivative, we get (7.44)

$$T_{x^- x^-}|_{D[\Sigma]} = E_S \delta(x^-) - \frac{c}{24\pi} \theta(x^-) \frac{2(\pi T_1)^2}{\left[1 - (\pi T_1 x^-)^2\right]^2} \equiv h(x^-). \tag{7.112}$$

[21] The differential equation determining $f(u)$ is (7.56)=(7.15).

We want to canonically purify the state at $t = t_\partial = f(u_\partial)$. The fact that $T_{x^\pm x^\pm}$ is constant along lines of constant $x^\pm$ implies that

$$\begin{aligned} T_{x^- x^-}(t_1 < t < t_\partial)|_{\partial M} &= h(t), \\ T_{x^+ x^+}(t_\partial < t < t_2)|_{\partial M} &= 0, \end{aligned} \tag{7.113}$$

where $t_2(/t_1)$ is the future-most (/past-most) boundary time in causal contact with Σ. Imposing reflecting boundary conditions gives

$$T_{x^- x^-}|_{\partial M} = T_{x^+ x^+}|_{\partial M} = \theta(t_\partial - t)h(t) + b\delta(t - t_\partial), \qquad t \in [t_1, t_2], \tag{7.114}$$

which fixes the solution

$$\begin{aligned} T_{x^- x^+} &= \theta(t_\partial - x^-)h(x^-) + b\delta(x^- - t_\partial), \\ T_{x^+ x^+} &= \theta(t_\partial - x^+)h(x^+) + b\delta(x^+ - t_\partial), \end{aligned} \tag{7.115}$$

where Fig. X indicates the domain where this solution is fixed. As was the case after the Page time, b is the amplitude of the shock caused by turning of absorbing boundary conditions.

Now we compute the dilaton. Choosing our reference point for the stress tensor integrals at $x^\pm = t_\partial \pm \varepsilon$ for some small positive ε, we have

$$\begin{aligned} I^+(x^+, x^-) &= \int_{t_\partial + \varepsilon}^{x^+} \mathrm{d}s (s - x^+)(s - x^-) \left[\theta(t_\partial - s)h(s) + b\delta(s - t_\partial)\right] \\ &= b(t_\partial - x^+)(t_\partial - x^-)\theta(x^+ - t_\partial) - \theta(-x^+)x^+ x^- E_S - \theta(t_\partial - x^+)F(x^+) \end{aligned} \tag{7.116}$$

where

$$\begin{aligned} F(x^+) &= -\frac{c}{24\pi} \int_{\max\{x^+, 0\}}^{t_\partial} \mathrm{d}s (s - x^+)(s - x^-)\theta(s) \frac{2(\pi T_1)^2}{\left[1 - (\pi T_1 s)^2\right]^2} \\ &= -\frac{c}{24\pi} \left[\frac{x^+ + x^- - s\left(1 + \pi^2 T_1^2 x^+ x^-\right)}{\pi^2 T_1^2 s^2 - 1} \right. \\ &\qquad \left. + \left(\pi T_1 x^+ x^+ - \frac{1}{\pi T_1}\right) \arctan\left(\pi T_1 s\right) \right] \Bigg|_{s = \max\{x^+, 0\}}^{s = t_\partial}. \end{aligned} \tag{7.117}$$

Similarly we get

$$I^-(x^+, x^-) = -b\theta(x^- - t_\partial)(t_\partial - x^+)(t_\partial - x^-) + \theta(-x^-)E_S x^+ x^- + \theta(t_\partial - x^-)F(x^-). \tag{7.118}$$

In total, this gives a dilaton

$$\begin{aligned} \phi &= \frac{c_1 + c_2(x^+ + x^-) + c_3 x^+ x^-}{x^+ - x^-} + \phi_m, \\ \phi_m &= 8\pi G_N \frac{\theta(t_\partial - x^-)F(x^-) - \theta(t_\partial - x^+)F(x^+)}{x^+ - x^-}, \end{aligned} \tag{7.119}$$

with the piecewise constant coefficients c_i

$$\begin{aligned} c_1 &= 2\bar{\phi}_r + 8\pi G_N b t_\partial^2 \left[\theta(x^+ - t_\partial) - \theta(x^- - t_\partial)\right] \\ c_2 &= -8\pi G_N t_\partial \left[\theta(x^+ - t_\partial) - \theta(x^- - t_\partial)\right] \\ c_3 &= -2\bar{\phi}_r(\pi T_1)^2 + 8\pi G_N \left[b\theta(x^+ - t_\partial) - b\theta(x^- - t_\partial) - E_S\theta(-x^+) + E_S\theta(-x^-))\right], \end{aligned} \tag{7.120}$$

where the coefficients in $D[\Sigma]$ is fixed by the knowledge that the homogeneous part of the solution at our reference point is that of (7.13) with temperature $T = T_1$.

References

[1] G. Penington, *Entanglement Wedge Reconstruction and the Information Paradox*, arXiv:1905.08255.

[2] A. Almheiri, N. Engelhardt, D. Marolf, and H. Maxfield, *The entropy of bulk quantum fields and the entanglement wedge of an evaporating black hole*, *JHEP* **12** (2019) 063, [arXiv:1905.08762].

[3] N. Engelhardt and A. C. Wall, *Quantum Extremal Surfaces: Holographic Entanglement Entropy beyond the Classical Regime*, *JHEP* **01** (2015) 073, [arXiv:1408.3203].

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

[5] V. E. Hubeny, M. Rangamani, and T. Takayanagi, *A Covariant holographic entanglement entropy proposal*, *JHEP* **0707** (2007) 062, [arXiv:0705.0016].

[6] T. Faulkner, A. Lewkowycz, and J. Maldacena, *Quantum corrections to holographic entanglement entropy*, *JHEP* **1311** (2013) 074, [arXiv:1307.2892].

[7] A. Almheiri, T. Hartman, J. Maldacena, E. Shaghoulian, and A. Tajdini, *The entropy of Hawking radiation*, *Rev. Mod. Phys.* **93** (2021), no. 3 035002, [arXiv:2006.06872].

[8] M. Van Raamsdonk, *Comments on quantum gravity and entanglement*, arXiv:0907.2939.

[9] M. Van Raamsdonk, *Building up spacetime with quantum entanglement*, *Gen. Rel. Grav.* **42** (2010) 2323–2329, [arXiv:1005.3035]. [Int. J. Mod. Phys.D19,2429(2010)].

[10] B. Czech, J. L. Karczmarek, F. Nogueira, and M. Van Raamsdonk, *The Gravity Dual of a Density Matrix*, *Class.Quant.Grav.* **29** (2012) 155009, [arXiv:1204.1330].

[11] A. Almheiri, X. Dong, and D. Harlow, *Bulk Locality and Quantum Error Correction in AdS/CFT*, *JHEP* **04** (2015) 163, [arXiv:1411.7041].

[12] X. Dong, D. Harlow, and A. C. Wall, *Reconstruction of Bulk Operators within the Entanglement Wedge in Gauge-Gravity Duality*, *Phys. Rev. Lett.* **117** (2016), no. 2 021601, [arXiv:1601.05416].

[13] J. Maldacena and L. Susskind, *Cool horizons for entangled black holes*, arXiv:1306.0533.

[14] M. Van Raamsdonk, *Evaporating Firewalls*, *JHEP* **11** (2014) 038, [arXiv:1307.1796].

[15] E. Verlinde and H. Verlinde, *Passing through the Firewall*, arXiv:1306.0515.

[16] A. Almheiri, D. Marolf, J. Polchinski, and J. Sully, *Black Holes: Complementarity or Firewalls?*, arXiv:1207.3123.

[17] A. Almheiri, D. Marolf, J. Polchinski, D. Stanford, and J. Sully, *An Apologia for Firewalls*, arXiv:1304.6483.

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

[19] E. Verlinde and H. Verlinde, *Black Hole Entanglement and Quantum Error Correction*, arXiv:1211.6913.

[20] L. Susskind, *Black Hole Complementarity and the Harlow-Hayden Conjecture*, arXiv:1301.4505.

[21] K. Papadodimas and S. Raju, *Remarks on the necessity and implications of state-dependence in the black hole interior*, *Phys. Rev.* **D93** (2016), no. 8 084049, [arXiv:1503.08825].

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

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

[24] D. N. Page, *Expected entropy of a subsystem*, *Phys. Rev. Lett.* **71** (1993) 1291–1294, [http://arXiv.org/abs/gr-qc/9305007].

[25] A. Lewkowycz and J. Maldacena, *Generalized gravitational entropy*, *JHEP* **1308** (2013) 090, [arXiv:1304.4926].

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

[27] A. Almheiri, T. Hartman, J. Maldacena, E. Shaghoulian, and A. Tajdini, *Replica Wormholes and the Entropy of Hawking Radiation*, *JHEP* **05** (2020) 013, [arXiv:1911.12333].

[28] S. R. Coleman, *Black Holes as Red Herrings: Topological Fluctuations and the Loss of Quantum Coherence*, *Nucl. Phys. B* **307** (1988) 867–882.

[29] S. B. Giddings and A. Strominger, *Loss of Incoherence and Determination of Coupling Constants in Quantum Gravity*, *Nucl. Phys. B* **307** (1988) 854–866.

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

[31] H. Geng, A. Karch, C. Perez-Pardavila, S. Raju, L. Randall, M. Riojas, and S. Shashi, *Information Transfer with a Gravitating Bath*, *SciPost Phys.* **10** (2021), no. 5 103, [arXiv:2012.04671].

[32] L. Anderson, O. Parrikar, and R. M. Soni, *Islands with gravitating baths: towards ER = EPR*, *JHEP* **21** (2020) 226, [arXiv:2103.14746].

[33] S. Hawking and D. N. Page, *Thermodynamics of Black Holes in anti-De Sitter Space*, *Commun.Math.Phys.* **87** (1983) 577.

[34] V. Balasubramanian, A. Kar, and T. Ugajin, *Entanglement between two gravitating universes*, arXiv:2104.13383.

[35] A. Miyata and T. Ugajin, *Entanglement between two evaporating black holes*, arXiv:2111.11688.

[36] V. Balasubramanian, A. Kar, O. Parrikar, G. Sárosi, and T. Ugajin, *Geometric secret sharing in a model of Hawking radiation*, *JHEP* **01** (2021) 177, [arXiv:2003.05448].

[37] D. Marolf and H. Maxfield, *Observations of Hawking radiation: the Page curve and baby universes*, *JHEP* **04** (2021) 272, [arXiv:2010.06602].

[38] R. Bousso and A. Shahbazi-Moghaddam, *Island Finder and Entropy Bound*, *Phys. Rev. D* **103** (2021), no. 10 106005, [arXiv:2101.11648].

[39] Y. Chen, V. Gorbenko, and J. Maldacena, *Bra-ket wormholes in gravitationally prepared states*, *JHEP* **02** (2021) 009, [arXiv:2007.16091].

[40] T. Hartman, Y. Jiang, and E. Shaghoulian, *Islands in cosmology*, *JHEP* **11** (2020) 111, [arXiv:2008.01022].

[41] M. Van Raamsdonk, *Comments on wormholes, ensembles, and cosmology*, arXiv:2008.02259.

[42] N. Engelhardt and A. C. Wall, *Coarse Graining Holographic Black Holes*, *JHEP* **05** (2019) 160, [arXiv:1806.01281].

[43] S. Dutta and T. Faulkner, *A canonical purification for the entanglement wedge cross-section*, *JHEP* **03** (2021) 178, [arXiv:1905.00577].

[44] R. Bousso, V. Chandrasekaran, and A. Shahbazi-Moghaddam, *From black hole entropy to energy-minimizing states in QFT*, *Phys. Rev. D* **101** (2020), no. 4 046001, [arXiv:1906.05299].

[45] J. Polchinski and A. Strominger, *A possible resolution of the black hole information puzzle*, *Phys. Rev. D* **50** (1994) 7403–7409, [http://arXiv.org/abs/hep-th/9407008].

[46] K. Jensen and J. Sonner, *Wormholes and entanglement in holography*, *Int. J. Mod. Phys. D* **23** (2014), no. 12 1442003, [arXiv:1405.4817].

[47] I. Gelfand and N. Neumark, *On the imbedding of normedrings into the diskof operators in hilbert space*, *Rec. Math. [Mat Sbornik] N.S.* **12(54)** (1943), no. 2 197–217.

[48] I. E. Segal, *Irreducible representations of operator algebras*, *Bull. Amer. Math. Soc.* **53** (02, 1947) 73–88.

[49] C. Barrabes and W. Israel, *Thin shells in general relativity and cosmology: The Lightlike limit*, *Phys. Rev.* **D43** (1991) 1129–1142.

[50] P. Gao, D. L. Jafferis, and A. Wall, *Traversable Wormholes via a Double Trace Deformation*, arXiv:1608.05687.

[51] J. Maldacena, D. Stanford, and Z. Yang, *Diving into traversable wormholes*, *Fortsch. Phys.* **65** (2017), no. 5 1700034, [arXiv:1704.05333].

[52] A. R. Brown, H. Gharibyan, G. Penington, and L. Susskind, *The Python's Lunch: geometric obstructions to decoding Hawking radiation*, *JHEP* **08** (2020) 121, [arXiv:1912.00228].

[53] N. Engelhardt, G. Penington, and A. Shahbazi-Moghaddam, *A World without Pythons would be so Simple*, arXiv:2102.07774.

[54] N. Engelhardt, G. Penington, and A. Shahbazi-Moghaddam, *Finding Pythons in Unexpected Places*, arXiv:2105.09316.

[55] H. Geng, A. Karch, C. Perez-Pardavila, S. Raju, L. Randall, M. Riojas, and S. Shashi, *Inconsistency of Islands in Theories with Long-Range Gravity*, arXiv:2107.03390.

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

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

[58] C. Teitelboim, *Gravitation and Hamiltonian Structure in Two Space-Time Dimensions*, *Phys. Lett. B* **126** (1983) 41–45.

[59] R. Jackiw, *Lower Dimensional Gravity*, *Nucl. Phys. B* **252** (1985) 343–356.

[60] A. Almheiri and J. Polchinski, *Models of AdS_2 backreaction and holography*, *JHEP* **11** (2015) 014, [arXiv:1402.6334].

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

[62] A. Belin and S. Colin-Ellerin, *Bootstrapping Quantum Extremal Surfaces I: The Area Operator*, arXiv:2107.07516.

[63] A. Goel, H. T. Lam, G. J. Turiaci, and H. Verlinde, *Expanding the Black Hole Interior: Partially Entangled Thermal States in SYK*, *JHEP* **02** (2019) 156, [arXiv:1807.03916].

[64] L. Susskind, *Computational Complexity and Black Hole Horizons*, *Fortsch. Phys.* **64** (2016) 24–43, [arXiv:1402.5674]. [Addendum: Fortsch.Phys. 64, 44–48 (2016)].

[65] L. Susskind, *Addendum to Computational Complexity and Black Hole Horizons*, arXiv:1403.5695.

[66] L. Susskind, *Entanglement is not enough*, *Fortsch. Phys.* **64** (2016) 49–71, [arXiv:1411.0690].

[67] D. Stanford and L. Susskind, *Complexity and Shock Wave Geometries*, *Phys. Rev. D* **90** (2014), no. 12 126007, [arXiv:1406.2678].

[68] T. Numasawa, *Four coupled SYK models and Nearly AdS_2 gravities: Phase Transitions in Traversable wormholes and in Bra-ket wormholes*, arXiv:2011.12962.

[69] C. Akers and P. Rath, *Entanglement Wedge Cross Sections Require Tripartite Entanglement*, *JHEP* **04** (2020) 208, [arXiv:1911.07852].

[70] C. Akers, N. Engelhardt, and D. Harlow, *Simple holographic models of black hole evaporation*, *JHEP* **08** (2020) 032, [arXiv:1910.00972].

[71] P. Hayden and G. Penington, *Learning the Alpha-bits of Black Holes*, *JHEP* **12** (2019) 007, [arXiv:1807.06041].

[72] C. Akers, S. Leichenauer, and A. Levine, *Large Breakdowns of Entanglement Wedge Reconstruction*, *Phys. Rev. D* **100** (2019), no. 12 126006, [arXiv:1908.03975].

[73] H. Verlinde, *ER = EPR revisited: On the Entropy of an Einstein-Rosen Bridge*, arXiv:2003.13117.

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

[75] M. Guica and D. L. Jafferis, *On the construction of charged operators inside an eternal black hole*, *SciPost Phys.* **3** (2017), no. 2 016, [arXiv:1511.05627].

[76] D. Harlow and D. Jafferis, *The Factorization Problem in Jackiw-Teitelboim Gravity*, *JHEP* **02** (2020) 177, [arXiv:1804.01081].

Chapter 8

General Bounds on Holographic Compexity

Abstract: We prove a positive volume theorem for asymptotically AdS spacetimes: the maximal volume slice has nonnegative vacuum-subtracted volume, and the vacuum-subtracted volume vanishes if and only if the spacetime is identically pure AdS. Under the Complexity=Volume proposal, this constitutes a positive holographic complexity theorem. The result features a number of parallels with the positive energy theorem, including the assumption of an energy condition that excludes false vacuum decay (the AdS weak energy condition). Our proof is rigorously established in broad generality in four bulk dimensions, and we provide strong evidence in favor of a generalization to arbitrary dimensions. Our techniques also yield a holographic proof of Lloyd's bound for a class of bulk spacetimes. We further establish a partial rigidity result for wormholes: wormholes with a given throat size are more complex than AdS-Schwarzschild with the same throat size.

8.1 Introduction

The interpretation of aspects of spacetime geometry in terms of quantum information theoretic quantities – from (quantum) extremal surfaces [1–9] to holographic circuit complexity [10–20] has been a critical aspect of recent progress towards a complete holographic description of the black hole interior. The Complexity=Volume dictionary entry [10], which is our focus in this article, was initially motivated in part by the potential of the maximal volume slice to probe the deep black hole interior. The proposal relates the circuit complexity $\mathcal{C}$ of preparing the CFT state $|\psi(\tau)\rangle$ from some reference state $|R\rangle$ and the bulk maximal volume slice Σ_τ anchored at τ: [1]

$$\mathcal{C}(|\psi(\tau)\rangle) = \max_{\Sigma_\tau} \frac{\text{vol}[\Sigma_\tau]}{G_N \ell} \equiv \mathcal{C}_V, \tag{8.1}$$

where ℓ is a reference length scale which we take to be the AdS radius L, $\text{vol}[\Sigma_\tau]$ is the cutoff-regulated volume of the hypersurface Σ_τ, and the maximum is taken over all bulk timeslices anchored at τ in the dual spacetime (M, g). Here we have given the name $\mathcal{C}_V$ to the geometric quantity for convenience. This proposal has led to numerous insights on spacetime emergence in the deep interior, from the late-time behavior of wormhole volumes [10, 20], to the connection between bulk depth and momentum [21–24], to the late-time breakdown of classical GR [25], among others (e.g. [26]).

Several aspects of the CV proposal remain ambiguous: the appropriate reference length scale ℓ, a complexity distance measure,[2] and a choice of reference state $|R\rangle$. It is natural to consider the vacuum $|0\rangle$ as reference state, but since $\mathcal{C}_V(|0\rangle)$ is non-vanishing this is not feasible. However, we could instead identify $\mathcal{C}(|\psi\rangle, |0\rangle)$ with the so-called complexity of formation $\mathcal{C}_F(|\psi\rangle)$ [15, 18], which is just the vacuum-subtracted version of $\mathcal{C}_V$:

$$\mathcal{C}_F(|\psi(\tau)\rangle) \equiv \mathcal{C}_V(|\psi(\tau)\rangle) - \mathcal{C}_V(|0\rangle) = \max_{\Sigma_\tau} \frac{\text{vol}[\Sigma_\tau] - n\,\text{vol}[\Sigma_{\text{AdS}}]}{G_N \ell}, \tag{8.2}$$

where n is the number of disjoint conformal boundaries on which the CFT lives. For sufficiently fast matter falloffs this quantity is finite as near-boundary cutoff is removed (see Sec. 8.2).

It is immediately clear that this interpretation of $\mathcal{C}$ requires a highly nontrivial geometric consistency check:

$$\mathcal{C}_F(|\psi\rangle) \geq 0, \tag{8.3}$$

where equality holds if and only if Σ_τ is identically a maximal volume slice of pure AdS. The positivity of $\mathcal{C}_F$ has been investigated in particular spacetimes [18, 28] and perturbations thereof [29, 30].

The statement (8.3) is reminiscent of the positive energy theorem [31–37], which in rough terms states that the mass of a spacetime with a well-behaved and complete initial data set satisfying the dominant energy condition is nonnegative; furthermore, it vanishes if and only if the data is identically the vacuum. The positive energy theorem assumes either asymptotic

[1] Here we can either take the perspective where postselection is allowed, i.e. non-unitary gates are allowed, or we can take this to compute the size of the circuit before including any postselection, and correct this by including the python's lunch [7].

[2] That is, a set of allowed unitary gates, or in the case of Nielsen geometry [27], a cost function.

flatness [31–33], or asymptotic hyperbolicity [34–37], and in the latter case it is sufficient to assume the weak energy condition.

Is there an analogous equally general theorem that applies to $\mathcal{C}_F$ under the assumptions of regularity of the initial data and some energy condition? Here we prove rigorously for a broad class of spacetimes that the answer is yes, and that the relevant energy condition is the AdS weak curvature condition (to which we refer as the WCC for short):

$$\mathcal{E} \equiv t^a t^b \left(R_{ab} - \frac{1}{2} g_{ab} R - \frac{d(d-1)}{2L^2} g_{ab} \right) \geq 0, \qquad \forall \text{ timelike } t^a, \tag{8.4}$$

which simplifies to the usual weak energy condition $T_{ab}t^a t^b \geq 0$ in Einstein gravity (under an appropriate shift of a constant from the matter Lagrangian to the gravitational Lagrangian). Our positive (holographic) complexity theorem then states:

Let Σ be a set of regular and complete asymptotically AdS_4 initial data on a maximal volume slice anchored at a static slice of the conformal boundary, and with n asymptotic boundaries. If Σ satisfies the WCC, then it has nonnegative vacuum-subtracted volume:

$$\mathrm{vol}[\Sigma] - n\,\mathrm{vol}[\Sigma_{\mathrm{AdS}}] \geq 0,$$

provided each outermost minimal surface on Σ is connected. Equality holds if and only if Σ is a static slice of pure AdS.

Thus, among all asymptotically AdS_4 spacetimes satisfying the WCC and our assumption on minimal surfaces, the vacuum is strictly the simplest: like the positive energy theorem, our positive complexity theorem provides a new vacuum rigidity result.

The theorem is proved in Sec. 8.2.4 via the application of a powerful mathematical result by Brendle and Chodosh [38, 39] that grants control over volumes of asymptotically hyperbolic Riemannian manifolds. For technical reasons, the result applies for maximal volume slices where each outermost minimal surface (see Sec. 8.2) is connected, effectively restricting to spacetimes where for any connected component $\mathscr{I}$ of the conformal boundary, $\partial J^-[\mathscr{I}] \cap \Sigma$ is connected on the maximal volume slice Σ.

Is the WCC actually necessary? Let us address this by taking inspiration from the positive energy theorem. Positivity of the mass and rigidity of the vacuum is indeed violated by false vacuum decay, which violates the dominant energy condition. We may expect a similar phenomenon here, as the WCC serves the same purpose here as the dominant energy condition: it excludes false vacuum decay and negative local energy densities. Indeed, existing violations of positivity of $\mathcal{C}_F$ [30, 40] do indeed violate our assumptions; we explore this in more detail in a companion article [41], where we will demonstrate that matter fields (e.g. scalar tachyons) violating the WCC – but satisfying the Null Energy Condition $T_{ab}k^a k^b \geq 0$ – can indeed violate the positive complexity theorem, just as they can violate the positive energy theorem [42]. Perhaps more surprisingly, it also turns out that the inclusion of compact dimensions can violate positivity of $\mathcal{C}_F$, even when the WCC holds. Common to all of the violations in [41] is the presence of VEVs for relevant CFT scalar primaries, and so the results proven in this paper could in more general theories likely be viewed as constraining how certain subsets of operators affect $\mathcal{C}_V$.

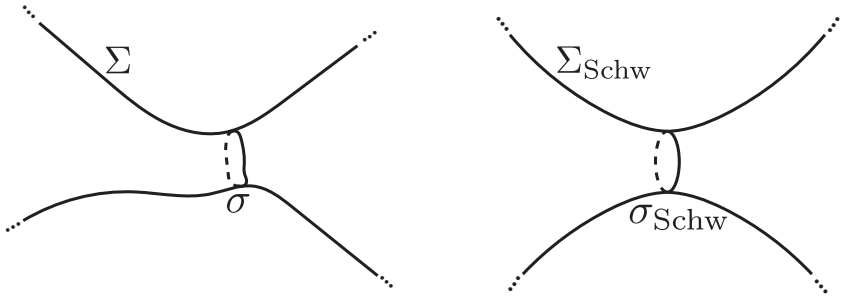


Figure 8.1: Illustration of the volume comparison theorem at fixed throat area. Σ_{Schw} is a static slice of Schwarzschild, while Σ is a maximal volume slice anchored at a static boundary time.

What about our restriction to four dimensions? Unlike the WCC or absence of nontrivial compact factors, we expect that this restriction can be lifted. The generalization to arbitrary dimensions can be proven from a generalization of a well-known mathematical conjecture by Schoen [43]; without relying on conjectures, we give a proof under the assumption of spherical or planar symmetry that $\mathcal{C}_F$ is positive (again assuming the WCC). We also highlight an existing mathematical theorem [44] proving that sufficiently small (but finite) WCC-preserving deformations of pure AdS result in positive $\mathcal{C}_F$, thus establishing in all dimensions that pure AdS is a local minimum of $\mathcal{C}_F$ within the space of WCC-preserving asymptotically AdS spacetimes. We consider the combination of our results, Schoen's conjecture and the small-deformation results [44] to be strong evidence in favor of the higher-dimensional generalization of the positive complexity theorem.

Rigidity of the vacuum implies that the vacuum has the lowest complexity of formation – it is simplest, under assumption of $\mathcal{C} \sim \text{vol}[\Sigma]$; there is an analogous expectation that the thermofield double is a fairly simple state that minimizes holographic complexity within some class. The corresponding AdS dual would be a proof that the Schwarzschild-AdS geometry minimizes the volume within some class of wormholes (see related work by [45]). Such a result about $\mathcal{C}_F$ would in fact constitute a parallel with a less famous companion to the positive energy theorem, the Riemannian Penrose Inequality, a lower bound on the mass given the existence of a minimal area surface [46–49]. More precisely, the area of a minimal surface on a maximal Cauchy slice in a spacetime of a given mass is conjectured to be identical to that of Schwarzschild with the same mass if and only if the spacetime is identical to Schwarzschild.[3]

Can our techniques provide a rigorous backing to this expectation that Scwarzschild-AdS is simplest in some precise sense? The analogy with the positive energy theorem would suggest that the answer is yes, and indeed it is: the same methodology used to prove $\mathcal{C}_F \geq 0$ yields a holographic complexity parallel of the Penrose Inequality (not restricted to a moment

[3]Note that while some results are established for the Riemannian Penrose Inequality in AdS [49–54], the full statement remains conjectural. We bring it up here to draw a parallel.

of time symmetry).

Let Σ be a complete maximal volume slice in an asymptotically AdS_4 spacetime with two disconnected asymptotic boundaries, equipped with initial data satisfying the WCC, and anchored at a static boundary slice. If A is the area of an outermost minimal-area surface on Σ, then

$$\mathrm{vol}[\Sigma] - \mathrm{vol}[\Sigma_{\mathrm{Schw}}] \geq 0,$$

provided each outermost minimal surface on Σ is connected. Equality holds if and only if Σ is identical to Σ_{Schw}, the static slice of AdS-Schwarzschild with throat area A.

That is, the area of the throat fixes the lowest possible complexity (as computed by the volume), which is that of Schwarzschild with the same throat area. See Fig. 8.1 for an illustration. In general dimensions we prove this under the restriction of spherical or planar symmetry. Once again, as we show in our companion paper [41], the WCC is necessary: it is possible to construct classical wormhole solutions satisfying the NEC with negative $\mathcal{C}_F$.

Finally, we make use of the tools developed in our proof of the positive complexity theorem to investigate a closely-related conjecture regarding the growth of complexity: Lloyd's bound [55]. This bound proposes that the rate of computation of a quantum system is bounded by an expression proportional to energy, and in the form discussed in [14, 15], it reads

$$\frac{\mathrm{d}\mathcal{C}}{\mathrm{d}\tau} \leq cE,$$

for some constant c. A large portion of the holographic complexity literature is dedicated to the investigation of this bound in specific spacetimes. In this work, for a class of solutions, we establish a similar result rigorously. To be precise, in minimally coupled Einstein-Maxwell-Scalar theory with spherical symmetry and more than three bulk dimensions (assuming the WCC), the growth of holographic complexity is bounded by the spacetime mass M:

$$\frac{1}{G_N L}\left|\frac{\mathrm{d}}{\mathrm{d}\tau}\,\mathrm{vol}[\Sigma_\tau]\right| \leq \frac{8\pi M}{d-1} f(M), \qquad f(M) = \begin{cases} 1 & M \leq \hat{M} \\ 1 + 2\left(\frac{M}{\hat{M}}\right)^{\frac{1}{d-2}} & M > \hat{M} \end{cases}, \tag{8.5}$$

where $\hat{M}$ is a mass scale near the Hawking-Page transition (see (8.42) for the specific value), and Σ_τ a maximal volume slice anchored at round sphere on the boundary – i.e. at constant τ in the boundary metric $-\mathrm{d}\tau^2 + L^2\mathrm{d}\Omega^2$. The takeaway is that late time holographic complexity growth is at most linear,[4] with the slope restricted by the mass. Since [56] showed that circuit complexity has a linearly growing upper bound in time (with an appropriate definition of complexity and time), this provides significant evidence for consistency of the CV proposal.

The paper is structured as follows. In Sec. 8.2 we state our proven lower bounds on complexity and proven Penrose inequality. The proofs are then given in a separate subsection, which can be skipped by readers only interested in the primary results, with the caveats that (1) the technology developed there reappears in the proof of Lloyd's bound and (2) some of the results of the proof section are interesting on their own right. In Sec. 8.3 we discuss

[4] More precisely, there is a linearly growing upper bound on holographic complexity.

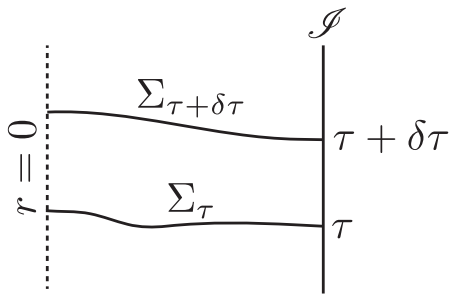


Figure 8.2: Illustration of two static anchored maximal volume slices separated by boundary static time $\delta\tau$. (8.5) provides a speed limit on the growth of $\frac{\mathrm{d}\,\mathrm{vol}[\Sigma_\tau]}{\mathrm{d}\tau}$.

upper bounds on complexity growth, with a proof of (8.5). We also give a novel formula for complexity growth in spacetimes with spherical or planar symmetry and with matter with compact spatial support (in the radial direction). This formula is valid independent of the WCC and the theory considered, although when assuming the former it gives a novel upper bound on complexity growth. Finally, in Sec. 8.4 we discuss implications of our result. A number of explicit computations are relegated to an appendix.

8.2 The Positive Complexity Theorem

In this section, we prove the positive volume (complexity) theorem for AdS_4, state the necessary conjecture for the proof of the statement in $\mathrm{AdS}_{d+1\neq 4}$, and prove the general d case for certain classes of bulk spacetimes. We also state our Penrose inequality for spatially symmetric spacetimes. The reader only interested in the results may skip section 8.2.4, which consists of the proofs.

8.2.1 The Vacuum is Simplest

We work with maximal volume slices Σ anchored on static boundary time slices: that is, in the conformal frame where the boundary manifold has coordinates

$$\mathrm{d}s^2 = -\mathrm{d}\tau^2 + L^2\mathrm{d}Q^2, \tag{8.6}$$

with $\mathrm{d}Q^2$ the metric of a maximally symmetric spacelike manifold (i.e. a sphere, plane, or hyperbolic space), we take our maximal volume slices to be anchored at constant τ, which we shall refer to as *static-anchored*; we further take Σ to be a manifold without boundaries or singularities – a *complete* Cauchy slice.[5] In particular, this means that our results do not apply to spacetimes terminating at end-of-the-world branes, in which Cauchy slices are manifolds with boundary.

[5] Σ will of course have a boundary in the conformal completion. The absence of singularities means that the intrinsic geometry of Σ is geodesically complete, which rules out intrinsic curvature singularities on Σ (unless they are distributional). Geodesic completeness on Σ is equivalent to the statement that all closed and bounded sets on Σ are compact [57].

Next, in order to discuss volume-regularization and assumptions on falloffs, it is useful to recall that for any choice of representative of the boundary conformal structure [58–60], the Fefferman-Graham gauge gives a unique coordinate system in a neighborhood of $\mathscr{I}$:

$$\mathrm{d}s^2 = \frac{L^2}{z^2}\left[\mathrm{d}z^2 + \gamma_{\mu\nu}(z,x)\mathrm{d}x^\mu \mathrm{d}x^\nu\right], \tag{8.7}$$

where the conformal boundary is at $z=0$, and where $\gamma_{\mu\nu}(z=0,x)$ is the chosen representative of the boundary conformal structure. The resulting Fefferman-Graham expansion reads

$$\gamma_{\mu\nu}(z,x) = \gamma^{(0)}_{\mu\nu}(x) + z^2\gamma^{(2)}_{\mu\nu}(x) + \ldots + \gamma^{(d)}_{\mu\nu}z^d + z^d \log z \bar{\gamma}^{(d)}_{\mu\nu} + \ldots. \tag{8.8}$$

We will assume that matter falloffs are sufficiently slow so that the $\gamma^{(n)}_{\mu\nu}$ for $n<d$, under the equations of motion, only depend on the boundary geometry $\gamma^{(0)}_{\mu\nu}$ and its derivatives. This is the case for pure Einstein-gravity with a negative cosmological constant, and also upon the inclusion various standard forms of matter appearing in holography [61] (provided we only turn on normalizable modes). We will call a spacetime asymptotically AdS (AAdS) whenever it has the above falloffs. This ensures that (1) a finite gravitational mass can be defined from the asymptotic value of the metric (as in [62]), and (2) the divergence structure of maximal volumes slices is universal and coincident with that of a static slice of pure AdS. Once the boundary conformal frame is fixed, this provides a canonical way to regularize the volume: cut off the spacetime at $z=\epsilon$ for some small $\epsilon>0$. After taking the volume difference with the regularized volume of a static slice of pure AdS, the result remains finite as $\epsilon\to 0$.

Finally, before turning to our results, we introduce one additional definition. We will say that a compact surface σ on Σ is *outermost minimal* if (1) σ divides Σ into two connected components (with one of them possibly trivial): Σ_{in} and Σ_{out}, where Σ_{out} intersects the conformal boundary $\mathscr{I}$ in the conformal completion on a complete Cauchy slice of a single connected component of $\mathscr{I}$, (2) σ is a local minimum of the area functional on Σ, and (3) there is no other surface satisfying (1) and (2) that is contained in Σ_{out}. We note that (1) outermost minimal surfaces have the property that every surface in Σ_{out} has area greater than σ [47] and (2) such surfaces always exist in wormhole geometries (see proof of Theorem (12)).

Our first theorem is a lower bound on maximal volume in four bulk dimensions.

Theorem 12. *Let Σ be a static-anchored complete maximal volume slice of an asymptotically globally*[6] *AdS_4 spacetime (M,g) satisfying the WCC (8.4). If the conformal boundary of (M,g) has n connected components $\mathscr{I}_i$, and the outermost minimal surface in Σ with respect to each $\mathscr{I}_i$ is connected (if non-empty), then the volume of Σ is bounded from below by the volume of n static slices Σ_{AdS_4} of pure AdS_4:*

$$\text{vol}[\Sigma] - n\,\text{vol}[\Sigma_{\text{AdS}_4}] \geq 0, \tag{8.9}$$

with equality if and only if (M,g) is identical to pure AdS_4 on Σ.

As mentioned above, while the volumes are formally infinite, comparison with the vacuum is well defined and finite; the vacuum-subtracted volume is finite and positive as the regulator is removed.

[6] By this we mean that the bulk approaches global AdS_4 at the asymptotic boundary, so that Q is a sphere.

Without the holographic interpretation of the vacuum-subtracted volume as the complexity of formation, this is a positive (renormalized) volume theorem. With $\mathcal{C} \propto \text{vol}[\Sigma]$ with some fixed proportionality constant, however, we have the desired positive complexity result: given a state $|\psi(\tau)\rangle$ of a holographic CFT_3 on n copies of the static cylinder, the complexity of $|\psi(\tau)\rangle$ is bounded from below by the complexity of the vacuum $|0\rangle$ on n disjoint spheres:

$$\mathcal{C}(|\psi\rangle) \geq \mathcal{C}(|0\rangle^{\otimes n}), \tag{8.10}$$

with equality if and only if $|\psi(\tau)\rangle = |0\rangle^{\otimes n}$. The complexity of formation is guaranteed to be positive for any bulk spacetime satisfying the WCC, and with a connected outermost minimal surface. The latter will be true for a single n-sided black hole (i.e. a wormhole when $n \geq 1$).

As noted in Sec. 8.1, the rigidity statement of the above theorem bears a strong parallel to vacuum rigidity following from the positive energy theorem: under assumptions of the latter, Minkowkski space [31, 32] is the unique asymptotically flat spacetime with vanishing mass. Analogous positive energy results have been established for pure AdS [34–37] under the assumption of the WCC. From the statement of Theorem 12, it is clear that a (WCC-satisfying) excitation of volume above the pure AdS value is possible if and only there is a corresponding non-zero mass: so the uniqueness of the minimal mass spacetime ensures the uniqueness of the minimal volume spacetime.

For bulk dimensions other than four, the technical results are restricted to either (1) have spherical or planar symmetry,[7] or (2) consist of small but finite deformations of the vacuum. Our general expectation is nevertheless that the results of four bulk dimensions continue to hold more generally. However, in the absence of a general proof to this effect, we state the lower bounds on complexity of formation in general dimensions as rigorously as possible, starting with spacetimes with spatial symmetry:[8]

Theorem 13. *Let Σ be a static-anchored complete maximal volume slice of an asymptotically $AdS_{d+1\geq 3}$ spacetime (M, g) with spherical or planar symmetry, and assume (M, g) satisfies the WCC. If the conformal boundary of (M, g) has n connected components, then the vacuum-subtracted volume of Σ is nonnegative:*

$$\text{vol}[\Sigma] - n\,\text{vol}[\Sigma_{\text{AdS}_{d+1}}] \geq 0, \tag{8.11}$$

with equality if and only if (M, g) is identical to n copies of pure AdS_{d+1} on Σ.

The upshot of this theorem is that any AAdS spacetime with spherical or planar symmetry (but arbitrary time dependence) that satisfies the WCC has a positive complexity of formation unless it is exactly pure AdS on the maximal volume slice, and thus also in its domain of dependence.[9] Note that spherically symmetric spacetimes should be compared with global AdS, and planar with Poincaré-AdS.

[7] Meaning that (M, g) can be foliated by a family of codimension-2 surfaces left invariant by an algebra of isometries equal to that of the $d-1$ dimensional sphere or plane. The codimension-2 submanifolds σ invariant under these symmetries are isometric to the sphere or plane, or quotients thereof.

[8] In fact, this theorem would also apply to hyperbolic symmetry if it could be shown analytically that static hyperbolic black holes have volumes greater than the vacuum. This is easily checked to be true numerically, but given the reliance on a numerical computation this result is not rigorously established.

[9] Note that this result holds even for AAdS spacetimes that do not have the falloffs assumed throughout this paper, although in this case $\text{vol}[\Sigma] - n\,\text{vol}[\Sigma_{\text{AdS}_{d+1}}]$ is positive and divergent as the regulator is removed. In this case the spacetime in some sense has infinite mass, since the nonstandard falloffs result in a divergent Balasubramanian-Kraus [62] stress tensor.

A known mathematics theorem immediately yields the desired result for perturbations around pure AdS in general dimensions:

Theorem 14 ([44])**.** *Let δg be a metric perturbation to global pure AdS_{d+1}, where the perturbation approaches zero sufficiently quickly to not affect the asymptotics [44]. A sufficiently small such δg satisfying the WCC cannot increase the volume of static-anchored maximal-volume slices.*

In this result δg need not be infinitesimal, as long as it is sufficiently small in an appropriate norm [44].

8.2.2 A Lower Bound on Wormhole Volume

In four bulk spacetime dimensions, we also prove the analogous result about wormhole complexity:

Theorem 15. *Let (M, g) be a WCC-satisfying, connected, globally asymptotically AdS_4 spacetime with two conformal boundaries. Let Σ be a static-anchored complete maximal volume slice in M, and let σ be the globally minimal surface on Σ.*[10] *Then, if the outermost minimal surfaces in Σ with respect to each component is connected,*

$$\mathrm{vol}[\Sigma] \geq \mathrm{vol}[\Sigma_{\mathrm{Schw}}], \tag{8.12}$$

where Σ_{Schw} is the totally geodesic slice of Schwarzschild-AdS whose bifurcation surface has area given by $\mathrm{Area}[\sigma]$.

This theorem resembles a Riemannian Penrose inequality [63] for volumes, or alternatively, for complexity: the area of a minimal surface on the maximal volume slice Σ puts a lower bound on the complexity. The observation that Penrose-like inequalites exist for volume was first made in [39], in the context of a result similar to Theorem 18, which is the crucial in our proof of Theorems 14 and 15.

Similar techniques facilitate a proof of the above result in other dimensions under assumption of spatial symmetry. We will use the abbreviation Stat_k to refer to the static AdS black hole with metric

$$\mathrm{d}s^2 = -\left(k + \frac{r^2}{L^2} - \frac{m}{r^{d-2}}\right)\mathrm{d}t^2 + \frac{\mathrm{d}r^2}{k + \frac{r^2}{L^2} - \frac{m}{r^{d-2}}} + r^2 \mathrm{d}Q_k^2, \tag{8.13}$$

where $\mathrm{d}Q_k^2$ is the metric of the $d-1$ dimensional unit sphere, plane, or unit hyperbolic space for $k = 1, 0, -1$, respectively. The constant m is proportional to the mass. We include the hyperbolic case, since interesting intermediate results in the next section apply for this case as well. We will use the term *maximal spatial symmetry* to refer to spacetimes with either spherical, planar or hyperbolic symmetry.

Under the assumption of maximal spatial symmetry, we prove the following:

[10] Such a surface must exist. See the proof of Theorem 12.

Theorem 16. *Let (M, g) be a connected asymptotically $AdS_{d+1\geq 3}$ WCC-satisfying spacetime with spherical or planar symmetry, and with a conformal boundary $\mathscr{I}$ with two connected components. Let Σ be a static-anchored complete maximal volume slice in M, and take $\sigma \in \Sigma$ to be the globally minimal symmetric surface on Σ.*[11] *Then*

$$\mathrm{vol}[\Sigma] \geq \mathrm{vol}[\Sigma_{\mathrm{Stat}}], \tag{8.14}$$

where Σ_{Stat} is the totally geodesic slice of Stat_k (depending on which symmetry (M, g) has) whose bifurcation surface has area $\mathrm{Area}[\sigma]$.

8.2.3 A Riemannian Penrose Inequality with Spatial Symmetry

Here we present a novel application of our results, which is somewhat tangential to the main subject of the paper. Uninterested readers may choose to skip ahead to the proofs of the theorems in Sec. 8.2.4. The main upshot is that our results directly imply the Riemannian Penrose inequality for maximal spatial symmetry. Using Ω_k to denote the volume of the unit sphere ($k = 1$), plane ($k = 0$) or unit hyperbolic plane ($k = -1$), we have

Theorem 17. *Let (Σ, h_{ab}, K_{ab}) be a complete asymptotically $AdS_{d+1\geq 3}$ maximal volume initial data set respecting the WCC and with maximal spatial symmetry. Let σ be the outermost stationary symmetric surface (possibly empty, in which case we define $\mathrm{Area}[\sigma] = 0$). Then*

$$\frac{16\pi G_N}{(d-1)\Omega_k} M \geq k \left(\frac{\mathrm{Area}[\sigma]}{\Omega_k}\right)^{\frac{d-2}{d-1}} + \frac{1}{L^2}\left(\frac{\mathrm{Area}[\sigma]}{\Omega_k}\right)^{\frac{d}{d-1}}, \tag{8.15}$$

where $k = 1, 0, -1$ for spherical, planar, or hyperbolic symmetry, respectively, and with equality if and only if (Σ, h_{ab}, K_{ab}) is a static slice of Stat_k or pure AdS.

This result does not require a moment of time symmetry, but σ will be (anti)trapped rather than marginally (anti)trapped unless $K_{ab} = 0$ at σ. Also note that while Ω_k may be formally infinite (unless we use isometries to compactify the symmetric surfaces), $\mathrm{Area}[\Sigma]/\Omega_k$ and M/Ω_k are well defined.[12]

8.2.4 Proofs of Theorems

Proof of Theorems 12 and 15

To prove Theorem 12, our main tool is an extant result about volumes of regions of conformally compact Riemannian 3-manifolds with certain asymptotics [38, 39]. The proof of Theorem 12 is constructed by establishing that (1) the requisite technical assumptions used in the mathematical literature hold for static-anchored maximal volume slices Σ of AAdS_4 spacetimes when the WCC holds, and (2) if Σ has n conformal boundaries it also has n disjoint asymptotic regions of the kind appearing in the result of [38, 39].

[11] For a connected two-sided complete slice Σ with the given symmetries, such a surface clearly exists.

[12] A formulation of the non-compact cases ($k = 0, -1$) of this inequality to spacetimes without symmetry would presumably have to involve some background subtraction. For example, this could likely be done for spacetimes that, on a spatial slice, approach a static planar or hyperbolic black hole outside a compact set in the conformal completion.

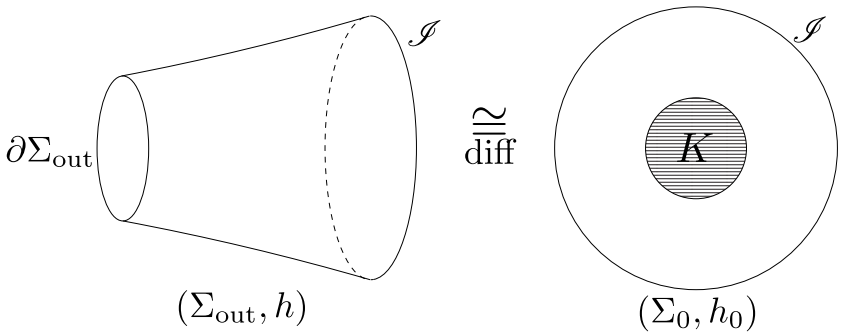


Figure 8.3: Example of geometry appearing in Theorem 18. On the left we see an asymptotic "arm" (Σ_{out}, h) of an asymptotically hyperbolic manifold, where $\partial\Sigma_{\text{out}}$ is outermost minimal. The region Σ_{out} is diffeomorphic to a subset $\Sigma_0 \setminus K$ of the hyperbolic plane (Σ_0, h_0), and h_0 and h are both metrics on Σ_{out} (via a judicious choice of diffeomorphism).

We can now state the result of [38, 39], which after a rephrasing reads [13]

Theorem 18 ([38], [39])**.** *Let (Σ_0, h_0) be hyperbolic 3-space and (Σ_m, h_m) the totally geodesic hypersurface of Schwarzschild-AdS of mass m, restricted to one side of the bifurcation surface. Consider a Riemannian metric h on $\Sigma_{\text{out}} = \Sigma_0 \setminus K$ where K is a compact set with smooth connected boundary. Assume that we can pick K so that*

- *(Σ_{out}, h) is asymptotically hyperbolic, meaning that*

$$\begin{aligned} |h - h_0|_{h_0} &= \mathcal{O}(r^{-2-\delta}), \quad \delta > 0, \\ \lim_{r\to\infty} |D_0(h - h_0)|_{h_0} &= 0, \end{aligned} \tag{8.16}$$

  where D_0 is the metric-compatible connection of h_0, and where r is defined through the coordinates

$$h_0 = \frac{1}{1+r^2}\mathrm{d}r^2 + r^2\mathrm{d}\Omega^2.$$

- *The Ricci scalar of (Σ, h) satisfies $R[h] \geq -6/L^2$.*
- *$\partial\Sigma_{\text{out}}$ is an outermost minimal surface with respect to h, and*

$$\text{Area}[\partial\Sigma_{\text{out}}, h] \geq \text{Area}[\partial\Sigma_m, h_m]$$

  for some $m \geq 0$.

Then the renormalized volumes are finite and satisfy

$$\text{vol}_{\text{ren}}(\Sigma_{\text{out}}, h) \geq \text{vol}_{\text{ren}}(\Sigma_m, h_m),$$

where

$$\text{vol}_{\text{ren}}(\Sigma_{\text{out}}, h) = \lim_{i\to\infty} \left[\text{vol}(\Sigma_{\text{out}} \cap \Omega_i, h) - \text{vol}(\Omega_i, h_0)\right], \tag{8.17}$$

[13] Note that while the theorem is stated with a strict inequality $m > 0$ in Ref. [38], Proposition 5.3 in Ref. [39] includes the case $m = 0$, where we should take K empty so that $\partial\Sigma_0 = \varnothing$.

and $\{\Omega_i\}$ *is an exhaustion of* Σ_0 *by compact sets, where asymptotic hyperbolicity ensures that the renormalized volume is well defined and finite. Moreover, if and only if equality holds, then* h *is isometric to* h_m*.*

See Fig. 8.3 for an illustration for quantities appearing in this theorem. Viewing (Σ_{out}, h) as a subset of a complete Riemannian manifold (Σ, h), the theorem says that the volume of (Σ, h) contained outside the outermost minimal surface of each asymptotic region is greater than the volume of a totally geodesic slice of Schwarzschild-AdS on one side of the bifurcation surface – provided that the first two conditions hold. Note that if we take (Σ_{out}, h) to be one half of a totally geodesic slice of Schwarzschild-AdS, then we find that as expected the volume of the static slice of Schwarzschild-AdS$_4$ monotonically increases with mass. As a special case, of course, for any positive mass the volume is greater than twice the volume of a static slice of pure AdS.

Before applying the result to volumes of maximal volume slices in AAdS$_4$ spacetimes, we must understand how restrictive asymptotic hyperbolicity is. We relegate the details to Appendix 8.5.1, where we prove the following:

Lemma 7. *Let* (M, g) *be an asymptotically globally AdS*$_4$ *spacetime. Then maximal volume static-anchored slices* Σ *in* (M, g) *are asymptotically hyperbolic.*

We now prove Theorem 12.

Proof. By Lemma 7, a maximal volume slice Σ is asymptotically hyperbolic. Next, by the Gauss-Codazzi equation, the WCC and maximality ($K = 0$), the Ricci scalar of the induced metric h on Σ is bounded from below:

$$R[h] = K^{\alpha\beta}K_{\alpha\beta} + 2n^a n^b \left(R_{ab}[g] - \frac{1}{2} g_{ab} R[g] \right) = K^{\alpha\beta}K_{\alpha\beta} - \frac{6}{L^2} + 2\mathcal{E} \geq -\frac{6}{L^2}, \tag{8.18}$$

where n^a is a unit normal to Σ. Thus, subsets Σ_{out} of Σ that lies outside a connected outermost minimal surface will satisfy the criteria for Theorem 18.

Let σ_i for $i = 1, \ldots, n$ be the outermost minimal surface for a given connected component $\mathscr{I}_i$ of $\mathscr{I}$. Minimal surfaces homologous to $\mathscr{I}_i$ are guaranteed to exist by the results of [64, 65] when $n \geq 2$.[14] Assume each σ_i is connected. For $n = 1$ we have that σ_1 can be empty. If σ_1 is not empty, then by two applications of Theorem 18 we have (since for any value of Area$[\sigma_i]$, there is an m such that Area$[\sigma_i] \geq$ Area$[\partial\Sigma_m]$)

$$\text{vol}[\Sigma_{\text{out},i}] \geq \frac{1}{2} \text{vol}[\Sigma_{\text{Schw}[\sigma_i]}] \geq \text{vol}[\Sigma_{\text{AdS}_4}], \tag{8.19}$$

where $\Sigma_{\text{Schw}[\sigma_i]}$ is the totally geodesic hypersurface of Schwarzschild-AdS whose bifurcation surface has the same area as σ_i. If σ_1 is empty, then a single application of Theorem 18 directly

[14] The result of [64, 65] guarantees the existence of a marginally outer trapped surface (MOTS) in an initial dataset with an outer trapped inner boundary and an outer untrapped outer boundary. If we take our initial dataset (Σ, h, K_{ab}) and turn it into the new dataset $(\Sigma, h, K_{ab} = 0)$, then MOTS in this second dataset are exactly the minimal surfaces of (Σ, h), and so the theorems of [64, 65] guarantee the existence of minimal surfaces when we have at least two asymptotic regions. Note that the results of [64, 65] do not rely on constraint equations or energy conditions, so setting $K_{ab} = 0$ without altering the intrinsic geometry causes no problems.

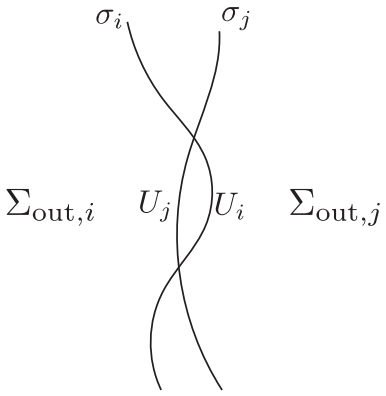


Figure 8.4: If two outermost minimal surface were to "weave" through each other, we would have a contradiction since either $(\sigma_i - U_i) \cup U_j$ would have area less than σ_i, or $(\sigma_j - U_j) \cup U_i$ would have area less than σ_j.

gives $\text{vol}[\Sigma_{\text{out},1}] \geq \text{vol}[\Sigma_{\text{AdS}_4}]$. For $n = 1$ this completes our proof, since $V[\Sigma] \geq V[\Sigma_{\text{out},1}]$, so assume now $n > 1$.

Take now $\Sigma_{\text{out},i}$ so that they are closed as subsets of Σ, that is, σ_i is included in $\Sigma_{\text{out},i}$. If the interiors of $\Sigma_{\text{out},i}$ and $\Sigma_{\text{out},j}$ intersect for some $i \neq j$, then either σ_i is fully contained in the interior of $\Sigma_{\text{out},j}$, or the two surfaces σ_i and σ_j "weave" through each other. The former is not possible, by outermost minimality, so consider the second case. Then σ_i has a subset U_i lying in $\Sigma_{\text{out},j}$, and σ_j a subset U_j in $\Sigma_{\text{out},i}$ such that $\hat{\sigma}_i = (\sigma_i - U_i) \cup U_j$ and $\hat{\sigma}_j = (\sigma_j - U_j) \cup U_i$ lie in Σ_i and Σ_j respectively – see Fig. 8.4. Outermost minimality of σ_i implies that

$$A[\sigma_i] \leq A[\hat{\sigma}_i] \qquad \Rightarrow \qquad A[U_i] \leq A[U_j]. \tag{8.20}$$

But this implies $A[\hat{\sigma}_j] \leq A[\sigma_j]$ contradicting the outermost minimality of σ_j, and so we conclude that Σ_i and Σ_j only can intersect on a measure 0 set, giving finally that

$$V[\Sigma] \geq \sum_{i=1}^{n} V[\Sigma_{\text{out},i}] = nV[\Sigma_{\text{AdS}_4}]. \tag{8.21}$$

□

Could the above theorem be true also for higher dimensions? A hint that this likely to be the case a comes from the following conjecture of Schoen [43][15]

Conjecture 1 ([43]). *Let (Σ, h_0) be a closed hyperbolic Riemannian manifold with constant negative scalar curvature $R[h_0]$. Let h be another metric on Σ with scalar curvature $R[h] \geq R[h_0]$. Then* $\text{vol}(\Sigma, h) \geq \text{vol}(\Sigma, h_0)$.

As stated the conjecture applies to compact rather than conformally compact manifolds. Nevertheless, it adds plausibility to the possibility that Theorem 12 holds in general dimensions. It also lends support to the idea that Theorem 12 should hold without the assumption that the outermost minimal surface is connected, since the conformally compact generalization of

[15]The conjecture is written in a different form in [43]. The version stated here can be found in [66].

Schoen's conjecture immediately implies Theorem 12 with this assumption removed. In a moment we add further evidence for this by proving it for $d \geq 3$ spacetimes when we have maximal spatial symmetry and when h_0 is hyperbolic space.

The proof of Theorem 15 follows that of Theorem 12 *mutatis mutandis* by replacing (8.19) with

$$\mathrm{vol}[\Sigma_{\mathrm{out},i}] \geq \frac{1}{2}\,\mathrm{vol}[\Sigma_{\mathrm{Schw}[\sigma_i]}] \geq \frac{1}{2}\,\mathrm{vol}[\Sigma_{\mathrm{Schw}[\sigma^*]}], \tag{8.22}$$

where now $\Sigma_{\mathrm{Schw}[\sigma_*]}$ is the totally geodesic hypersurface of Schwarzschild-AdS with a bifurcation surface with area equal to the globally minimal surface σ_* on Σ. This second inequality holds due to monotonicity of the Schwarzschild-volume (and bifurcation surface area) with mass.

Proof of Theorems 13, 16 and 17

The proof of Theorem 13 is threefold: first, we establish that the intrinsic metric of an extremal hypersurface in an asymptotically AdS spacetime with maximal spatial symmetry can be written as

$$\mathrm{d}s^2 = \frac{1}{k + \frac{r^2}{L^2} - \frac{\omega(r)}{r^{d-2}}}\mathrm{d}r^2 + r^2\mathrm{d}Q_k^2, \tag{8.23}$$

where $\omega(r)$ is a monotonously non-decreasing function whenever the WCC holds. Second, we show that at a stationary surface $r = r_*$, ω satisfies the equation

$$k + \frac{r_*^2}{L^2} - \frac{\omega(r_*)}{r_*^{d-2}} = 0, \tag{8.24}$$

and by the monotonicity of ω the volume contained outside r_* is greater than the one we obtain by replacing $\omega(r)$ with the constant $\omega(r_*)$. Third, we note that constant ω corresponds to a constant$-t$ slice of Stat_k, and taken together with (8.24) we see that the geometry outside r_* after the replacement $\omega(r) \to \omega(r_*)$ is exactly that of half a constant-t slice of Stat_k. Thus, the volume outside a stationary surface at r_* is greater than that of half a constant-t slice of Stat_k with mass parameter $\omega(r_*)$, and this is in turn greater than the volume of a static AdS slice.

We show in Appendix 8.5.2 that there is a geometric functional $\omega[\sigma, \Sigma]$ that, in an AAdS spacetime with maximal spatial symmetry, reduces to $\omega(r)$ when σ is a surface invariant under the spatial symmetry – which we shall call a 'symmetric' surface – having area radius r – henceforth denoted σ_r. The quantity

$$M_\Sigma[\sigma_r] \equiv \omega[\sigma_r, \Sigma]\frac{(d-1)\Omega_k}{16\pi G_N} \tag{8.25}$$

asymptotes to the spacetime mass as σ_r approaches the asymptotic boundary [67], where we remind that Ω_k is the volume of the unit sphere, the plane or the unit hyperbolic plane.[16] In Sec. 8.3 we will see that $\omega(r) \geq 0$ whenever the WCC holds, $d \geq 3$ and the spatial symmetry is spherical or planar. As an aside, in the asymptotically flat context, this quantity was used early partial proofs of the positive energy theorem [69, 70].

[16] For the falloffs and symmetries assumed in this paper this mass agrees with the CFT energy up to a fixed offset corresponding to the Casimir energy. For weaker falloffs we have that $M_\Sigma[\sigma]$ diverges towards the boundary. For $d = 3$, $M_\Sigma[\sigma]$ is the so-called Geroch-Hawking mass [34, 68–72].

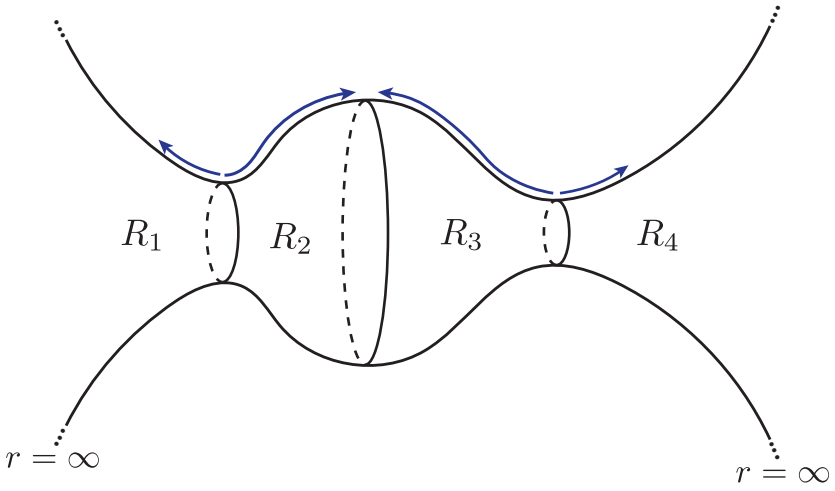


Figure 8.5: Example of a two-dimensional spherically symmetric hypersurface Σ with three stationary surfaces. This manifold can be covered by four coordinate systems of the form (8.26), and $\omega(r)$ increases along the blue arrows.

We now prove the properties of $\omega(r)$ claimed in the beginning of this subsection (see Fig. 8.5 for a illustration of the quantities appearing in this lemma).

Lemma 8. *If Σ is a complete maximal volume slice in an $AAdS_{d+1}$ spacetime with maximal spatial symmetry satisfying the WCC, then*

- *Any region R that lies on Σ between two stationary symmetric surfaces can be covered by the coordinates*

$$\mathrm{d}s^2|_\Sigma = \frac{1}{k + \frac{r^2}{L^2} - \frac{\omega(r)}{r^{d-2}}}\mathrm{d}r^2 + r^2\mathrm{d}Q_k^2. \tag{8.26}$$

  In particular, this is true when one of the two stationary surfaces is $\varnothing$ ($r = 0$) or asymptotic infinity ($r = \infty$).

- $$\omega'(r) \geq 0. \tag{8.27}$$

- *There is always an outermost symmetric stationary surface $\sigma_{\tilde{r}}$ such that (8.26) covers all $r > \tilde{r}$, where $\sigma_{\tilde{r}} = \varnothing$ and $\tilde{r} = 0$ is allowed.*

Proof. Since by assumption Σ is foliated by symmetric surfaces, locally on Σ we can pick an ADM coordinate system with respect to this foliation with vanishing shift and unit lapse:

$$\mathrm{d}s^2 = \mathrm{d}\rho^2 + \gamma_{ij}(\rho, x)\mathrm{d}x^i\mathrm{d}x^j,$$

where x^i are coordinates on the symmetric surfaces. Maximal spatial symmetry implies that $\gamma_{ij}(\rho, x)\mathrm{d}x^i\mathrm{d}x^j = f(\rho)\mathrm{d}Q_k^2$ for some function $f(\rho)$ that is nonnegative by the spacelike signature of Σ. Defining the new coordinate r through

$$r = \sqrt{f(\rho)}, \tag{8.28}$$

gives a metric of the form

$$\mathrm{d}s^2 = B(r)\mathrm{d}r^2 + r^2\mathrm{d}Q_k^2, \qquad B(r) = \frac{4f}{(\partial_\rho f)^2}. \tag{8.29}$$

These coordinates break down at $r = 0$ and at radii where $\partial_\rho f(\rho) = 0, f \neq 0 \Leftrightarrow B = \infty$, i.e. at symmetric surfaces that are stationary. The case of $B = 0$ corresponds to a singularity (the Kretschmann scalar is divergent for $d \geq 3$, and there is a conical singularity for $d = 2$), which is incompatible with completeness of Σ.[17] To bring the metric to the required form, we simple define $\omega(r)$ through

$$B(r) = \frac{1}{k + \frac{r^2}{L^2} - \frac{\omega(r)}{r^{d-2}}}. \tag{8.30}$$

For the second point, we use the Gauss-Codazzi equation applied to Σ, which yields

$$(d-1)\frac{\omega'(r)}{r^{d-1}} = K^{\alpha\beta}K_{\alpha\beta} + 2\mathcal{E} \geq 0, \tag{8.31}$$

where we have used the WCC, extremality of Σ and positivity of $K_{\alpha\beta}K^{\alpha\beta}$.

Finally, since the spacetime has finite mass, $\omega(r)$ converges to some finite number m at large r. Thus, the quantity

$$B(r)^{-1} = k + \frac{r^2}{L^2} - \frac{\omega(r)}{r^{d-2}} \tag{8.32}$$

must have a last zero $\tilde{r}$, unless it has none, in which case we set $\tilde{r} = 0$. Consequently, for $r > \tilde{r}$ there are no stationary surfaces, and so a single coordinate patch (8.29) covers this region. □

Next we prove our main volume comparison result:

Theorem 19. *Let (M, g) be a maximally spatially symmetric asymptotically $AdS_{d+1\geq 3}$ spacetime satisfying the WCC, and let Σ be a static-anchored complete maximal volume slice of M. Let σ_{r_i} be the outermost symmetric stationary surface with respect to the connected component $\mathscr{I}_i$ of $\mathscr{I}$. Then*

$$V(\Sigma) \geq \sum_i V(\Sigma_{m_i}), \tag{8.33}$$

with

$$V(\Sigma_{m_i}) = \Omega_k \int_{r_i}^{\infty} \mathrm{d}r \frac{r^{d-1}}{\sqrt{k + \frac{r^2}{L^2} - \frac{m_i}{r^{d-2}}}}, \tag{8.34}$$

and with m_i defined by

$$k + \frac{r_i^2}{L^2} - \frac{m_i}{r_i^{d-2}} = 0. \tag{8.35}$$

If $\mathscr{I}$ is connected, (8.33) is an equality if and only if Σ has the same extrinsic geometry as that of a static slice of pure AdS. If $\mathscr{I}$ has multiple connected components, (8.33) is an equality if and only if Σ has the same extrinsic geometry as that of a totally geodesic slice of a static AdS black hole with spherical, planar, or hyperbolic horizon topology.

[17] A discontinuous but nonzero $B(r)$ corresponds to a distributional energy shock on Σ, which does not lead to breakdown of the coordinates.

Proof. It is illustrative to begin with the one-sided case. Let σ_{r_1} be the outermost symmetric stationary surface on Σ. If there is none then $r_1 = 0$. Clearly the full volume of Σ is greater or equal to the volume outside σ_{r_1}, and since by Lemma 8, $\omega'(r) \geq 0$, and the region outside σ_{r_1} is covered by one patch of the coordinates (8.26), we get

$$\begin{aligned} V(\Sigma) &\geq \Omega_k \int_{r_1}^{\infty} \frac{\mathrm{d}r r^{d-1}}{\sqrt{k + \frac{r^2}{L^2} - \frac{\omega(r)}{r^{d-2}}}} \\ &\geq \Omega_k \int_{r_1}^{\infty} \frac{\mathrm{d}r r^{d-1}}{\sqrt{k + \frac{r^2}{L^2} - \frac{\omega(r_1)}{r^{d-2}}}} \\ &= V(\Sigma_{\omega(r_1)}), \end{aligned} \tag{8.36}$$

where $\omega(r_1)$ satisfies $0 = k + \frac{r_1^2}{L^2} - \frac{\omega(r_1)}{r_1^{d-2}}$ by stationarity of σ_{r_1}. The two-sided cases follows from the same argument applied twice.

Finally, the above inequalities can all be replaced by equalities if and only if $\omega(r)$ is constant, and with $n \in \{1, 2\}$ coordinate patches fully covering Σ when there are n conformal boundaries. This implies the intrinsic geometry corresponds to a constant$-t$ slice of either pure AdS ($n = 1$) or Stat_k ($n = 2$). From (8.31), we see that ω is constant only if $K_{ab} = 0$, and so the full extrinsic geometry is that of a constant$-t$ slice embedded in pure AdS or Stat_k. □

The final ingredient for the proof of Theorems 13 and 16 is the behavior of the volume of totally geodesic slices of Stat_k as a function of the mass. In the case $d = 2$, i.e. for the BTZ black hole, we readily find that the volume of a totally geodesic hypersurface relative to that of static slices of pure AdS is a positive constant independent of mass. Consider thus now $d \geq 3$, and define the vacuum-subtracted volumes:

$$\begin{aligned} I_k(m) &= \Omega_k \lim_{r_c \to \infty} \left[\int_{r_h(m)}^{r_c} \mathrm{d}r \frac{r^{d-1}}{\sqrt{k + \frac{r^2}{L^2} - \frac{m}{r^{d-2}}}} - \int_0^{r_c} \mathrm{d}r \frac{r^{d-1}}{\sqrt{k + \frac{r^2}{L^2}}} \right], \qquad k \in \{0, 1\} \\ I_{-1}(m) &= \Omega_{-1} \lim_{r_c \to \infty} \left[\int_{r_h(m)}^{r_c} \mathrm{d}r \frac{r^{d-1}}{\sqrt{-1 + \frac{r^2}{L^2} - \frac{m}{r^{d-2}}}} - \int_L^{r_c} \mathrm{d}r \frac{r^{d-1}}{\sqrt{-1 + \frac{r^2}{L^2}}} \right]. \end{aligned} \tag{8.37}$$

Here $r_h(m)$ satisfies $k + \frac{r_h^2}{L^2} - \frac{m}{r_h^{d-2}} = 0$. For $k \in \{0, 1\}$, smoothness and completeness of Σ is only compatible with $m \geq 0$. For $k = -1$, we can also have black holes with $\hat{m} \leq m < 0$, where [73, 74]

$$\hat{m} = -\frac{2L^{d-2}}{d-2} \left(\frac{d-2}{d} \right)^{\frac{d}{2}}. \tag{8.38}$$

The function $I_0(m)$ was computed explicitly in [18] was be seen to be positive and monotonically increasing with m [18]. For $k = 1$ and $d = 3$ the same properties follow from Theorem 18, as discussed previously. In the Appendix 8.5.3 we prove that this extends to all $d \geq 3$ by generalizing a proof from [38]:

Lemma 9. *The volume of a totally geodesic hypersurface of Schwarzschild-AdS$_{d+1}$ is monotonically increasing with mass:*

$$\partial_m I_1(m) > 0.$$

For $k = -1$ numerical studies [18] indicate that $I_{-1}(m)$ is non-negative, has a global minimum equal to zero at $m = 0$, and increases monotonically with m for $m > 0$; at $m < 0$ it appears to monotonically decrease, and it diverges at $m = \hat{m}$. However, at the moment there is no rigorous proof these results, so we only state the $k \in \{0, 1\}$ cases as theorems.

It follows from the above that that totally geodesic hypersurfaces of Stat$_k$ black holes for $k \in \{0, 1\}$ have volumes greater than $2\,\mathrm{vol}[\Sigma_{\mathrm{AdS}_{d+1}}]$, which implies that the right-hand side of (8.33) is greater than one or two times $\mathrm{vol}[\Sigma_{\mathrm{AdS}_{d+1}}]$. This proves Theorem 13.

To prove Theorem 16, note that $\partial_m I_k(m) > 0$ for $k \in \{0, 1\}$ implies $\partial_A I_k(m(A)) > 0$, with A the area of the bifurcation surface, since the area of the bifurcation surface increases with mass. In the case of two conformal boundaries, we can now bound the right-hand side of (8.33) from below by $\mathrm{vol}(\Sigma_*)$, where Σ_* is a totally geodesic slice of Stat$_k$ with bifurcation surface area equal to the globally minimal surface area.

Finally, we note that the stationarity condition $k + r_*^2 - \omega(r_*)/r_*^{d-2} = 0$ together with the monotonicity of ω (giving $\omega(\infty) \geq \omega(r_*)$) immediately proves Theorem 17.

8.3 Upper Bounds on Complexity Growth

8.3.1 Lloyd's bound

Lloyd's bound [55] is a conjectured constraint on general quantum systems given by

$$\frac{1}{\Delta\tau} \leq \frac{2E}{\pi}, \tag{8.39}$$

where E is the energy of the system and $\Delta\tau$ the minimal time that a quantum state takes to transition from an initial state to a new orthogonal state. This relation constitutes a bound on the maximal speed of computation. The existence of a bound on the growth of either bulk volume or action in terms of CFT energy would provide evidence for the holographic complexity proposals, and there is a large literature devoted to investigations of holographic complexity growth [10–12, 14, 15, 20, 75–116]. In particular, the points of interest is whether this growth is linear at late times,[18] as hypothesized in [117] , and whether there is an upper bound on the growth of holographic complexity of the form

$$\frac{\mathrm{d}\mathcal{C}}{\mathrm{d}\tau} \leq cE, \tag{8.40}$$

for the CFT energy E and some constant c. In the holography literature (8.40) is typically referred to as Lloyd's bound. For the CV-proposal with the AdS radius as L, it is natural to set $c = \frac{8\pi}{d-1}$, so that (8.40) is an upper bound in terms of the late time volume growth in

[18] But not so late as to reach the quantum recurrence time [117], where complexity saturates and fluctuates. That circuit complexity indeed grows linearly until the recurrence time was recently argued for random quantum circuits in [118].

Schwarzschild-AdS as $M \to \infty$ (with the identification of E with the Schwarzschild mass M [20]). For charged and rotating spacetimes it is expected that (8.40) can be strengthed via the inclusion of terms depending on the charge and angular momentum on the right hand side of (8.40) [14, 15, 75].

While specific case studies are invaluable as tests of the complexity proposals, a robust test of whether (8.40) holds for holographic complexity via CV requires a broad investigation of the growth of $\mathcal{C}_V$, which has heretofore been elusive.[19] Here we make strides towards closing that lacuna:

Theorem 20. *A WCC-satisfying spherically symmetric $AAdS_{d+1\geq 4}$ spacetime with any number of minimally coupled $U(1)$ gauge fields and scalars, both charged and neutral satisfies*

$$\left|\frac{\mathrm{d}\mathcal{C}_V}{\mathrm{d}\tau}\right| \leq \frac{8\pi M}{d-1} f(M), \tag{8.41}$$

where M is the spacetime mass and

$$f = \begin{cases} 1 & M \leq \hat{M} \equiv \frac{(d-1)\,\mathrm{vol}(S^{d-1})L^{d-2}}{16\pi G_N}\left[\frac{d}{4(d-2)}\right]^{\frac{d-2}{2}} \\ 1+\left(\frac{M}{\hat{M}}\right)^{\frac{1}{d-2}} & \text{otherwise.} \end{cases} . \tag{8.42}$$

Here τ is the time parameter in the boundary conformal frame $-\mathrm{d}\tau^2 + L^2\mathrm{d}\Omega^2$.

Theorem 20 is an immediate proof that $\frac{\mathcal{C}_V(\tau)}{\tau}$ is bounded at late times (in spacetimes subject to our assumptions). Thus if $\mathcal{C}_V$ is not oscillating at late times, then the late time growth of $\mathcal{C}_V$ is at most linear. Finally, if in some multiboundary spacetime the different boundaries have different masses, then M in (8.41) is the one associated to the conformal boundary at which we perturb the time anchoring.

Provided that we can identify the spacetime mass M with the CFT energy E, our bound is identical to Lloyd's bound (8.40) for small masses, while for large masses ($M \geq \hat{M}$) it is qualitatively similar, although the energy dependence is now non-linear. Thus, for spacetimes with ordinary falloffs, Theorem 20 constitutes a partial proof of a complexity growth bound, albeit with non-linear energy dependence, and where it is the vacuum-subtracted CFT energy that appears in the bound.[20]

Before turning to proving our result, let us finally note that the proof of (8.41) puts the CV and CA proposals on quite different footings when it comes to complexity growth. It is known that Lloyd's bound cannot hold for the CA proposal [20, 115]; spacetimes with ordinary AdS asymptotics and finite gravitational mass can have negative and divergent action growth [20].

[19] Some partial results on Lloyd's bound at late times in static spacetimes exist for the Complexity=Action proposal [119].

[20] Our assumptions about the falloffs are naturally violated in spacetimes where the ordinary gravitational mass M, as defined either as in [62] (with subtraction of Casimir energy) or as the Geroch-Hawking mass at infinity, diverges. In such case the growth of extremal surface volumes can be infinite, as in e.g. [84]. In such cases the field theory energy after subtraction of the Casimir energy is no longer equal to M, and additional counterterms depending on scalar fields contributes to give a finite CFT energy E [67, 120–124]. In this scenario (8.41) is true but vacuous since both sides of the inequality diverge.

An easy way to compute complexity change: Under the assumption of maximal spatial symmetry, the extrinsic geometry of the initial maximal volume slice at some time is sufficient to yield a simple expression for $\dot{\mathcal{C}}_V$: given a one-parameter family of extremal slices Σ_τ ($\Sigma_0 \equiv \Sigma$) with boundary $\partial\Sigma_\tau$, extremality ensures that any volume change from a perturbation must be a boundary term (see for example the appendices of [125, 126]):

$$\frac{\mathrm{d}V[\Sigma_\tau]}{\mathrm{d}\tau} = \int_{\partial\Sigma_\tau} \eta^a N_a, \tag{8.43}$$

where $\eta^a = (\partial_\tau)^a$ is the displacement field for $\partial\Sigma_\tau$ and N_a the outwards unit normal to $\partial\Sigma_\tau$ in Σ_τ. We want to work with Σ_τ which technically do not have boundaries. However, the above formula can be used by cutting off Σ_τ at a finite radius, applying the formula, and then afterward taking the limit where $\partial\Sigma_\tau$ goes to the conformal boundary.

Assume now that the gravitational mass is finite and remains so on evolution of the initial data on Σ. Expressing $\eta^a N_a$ in terms of quantities over which we have control, carried out in Appendix 8.5.4, we find for spherical or planar symmetry that

$$\frac{\mathrm{d}\mathcal{C}_V}{\mathrm{d}\tau} = \frac{\Omega_k}{G_N L(d-1)} \lim_{r\to\infty} r^d K_{ab} r^a r^b, \tag{8.44}$$

where r^a is the outwards unit normal to σ_r tangent to Σ. Thus, knowledge of K_{ab} near the boundary is sufficient to compute $\dot{\mathcal{C}}_V$.[21] In Appendix 8.5.4 we show that we can rewrite this expression in Einstein+matter gravity as

$$\frac{\mathrm{d}\mathcal{C}_V}{\mathrm{d}\tau} = \frac{1}{G_N L(d-1)} \left[r_{\text{throat}} \int_{\partial\Sigma_{\text{out}}} \sqrt{2\theta_k\theta_\ell} + 8\pi G_N \int_{\Sigma_{\text{out}}} r T_{ab} n^a r^b \right], \tag{8.45}$$

where now $\partial\Sigma_{\text{out}}$ is an outermost stationary surface on Σ at radius $r = r_{\text{throat}}$, and Σ_{out} is the part of Σ outside this surface. Here n^a is the future normal to Σ, k and ℓ the future null normals to $\partial\Sigma_{\text{out}}$ normalized so $k\cdot\ell = -1$, and θ_k and θ_ℓ are the null geodesic expansions. It follows from (8.139) in the appendix that $\theta_k\theta_\ell \geq 0$ by stationarity of $\partial\Sigma_{\text{out}}$, so the square root is always real This expression (8.45) is a special case of the more general momentum-complexity correspondence proven in [23], with the difference that we write our integral on only part of Σ at the cost of a boundary term. In spacetimes where $n^a r^b T_{ab} = 0$, which includes AdS-Schwarzschild and AdS-Reissner-Nordström, we see that the complexity growth is proportional $\sqrt{\theta_k\theta_\ell}$ at the throat of Σ. This suggests a connection between trapped and anti-trapped regions and complexity growth, at least in these spacetimes. The more trapped the throat of Σ is, the faster the complexity growth.

With our convenient formulae for $\dot{\mathcal{C}}_V$ in hand, we can turn to proving that the mass indeed bounds $\dot{\mathcal{C}}_V$ in Einstein-Maxwell-Scalar theory.

[21] Note however that a near boundary analysis is not enough to obtain K_{ab} in the first place. If we treat the determination of an extremal slice as a shooting problem from the conformal boundary, a general initial condition at the boundary is incompatible with smoothness in the bulk, so determining $\dot{\mathcal{C}}_V$ still requires deep-bulk information.

Proof of a Lloyd's bound in Einstein-Maxwell-Scalar theory (with spherical symmetry)

Explicit constraint solutions

Consider Einstein-gravity minimally coupled to a $U(1)$ gauge field and a charged scalar:

$$S = \frac{1}{8\pi G_N}\int_M \mathrm{d}^{d+1}x\sqrt{-g}\left[\frac{1}{2}R + \frac{d(d-1)}{2L^2} - |\mathcal{D}\phi|^2 - V(\phi,\phi^\dagger) - \frac{1}{4}F_{ab}F^{ab}\right], \tag{8.46}$$

where $\mathcal{D}$ is covariant derivative associated with the $U(1)$ gauge field. The stress tensor reads

$$8\pi G_N T_{ab} = \mathcal{D}_a\phi(\mathcal{D}_b\phi)^\dagger + \mathcal{D}_b\phi(\mathcal{D}_a\phi)^\dagger - g_{ab}|\mathcal{D}\phi|^2 - g_{ab}V(\phi,\phi^\dagger) + F_a{}^cF_{bc} - \frac{1}{4}g_{ab}F_{cd}F^{cd}. \tag{8.47}$$

We now restrict to spherical, planar or hyperbolic symmetry and work in the coordinates (8.29):

$$\mathrm{d}s^2|_\Sigma = B(r)\mathrm{d}r^2 + r^2\mathrm{d}\Omega_k^2 = \frac{1}{k + \frac{r^2}{L^2} - \frac{\omega(r)}{r^{d-2}}}\mathrm{d}r^2 + r^2\mathrm{d}\Omega_k^2. \tag{8.48}$$

We then get that the energy density and radial energy current on Σ are

$$\begin{aligned}\mathcal{E} &= |\mathcal{D}_t\phi|^2 + \frac{1}{B(r)}|\mathcal{D}_r\phi|^2 + V(\phi,\phi^\dagger) + \frac{1}{2B(r)}F_{tr}^2 + \frac{1}{4}F^{ij}F_{ij}\\ J_r &= \mathcal{D}_t\phi(\mathcal{D}_r\phi)^\dagger + \mathcal{D}_r\phi(\mathcal{D}_t\phi)^\dagger,\end{aligned} \tag{8.49}$$

where $\mathcal{D}_t \equiv n^a\mathcal{D}_a$ and $F_{tb} \equiv n^aF_{ab}$, with n^a is the future unit normal to Σ. The i,j-indices run over coordinates of the transverse space. Symmetry sets $F_{ri} = F_{ti} = 0$. Certain nonzero F_{ij} are possible, but the constraints of Maxwell theory might set these to zero, depending on the topology of Σ. The WCC, here equivalent to the WEC, holds when $V(\phi,\phi^\dagger) \geq 0$.

Symmetry and $K = 0$ dictates that

$$K_{\alpha\beta}\mathrm{d}x^\alpha\mathrm{d}x^\beta = K_{rr}(r)\left[\mathrm{d}r^2 - \frac{r^2}{B(r)(d-1)}\mathrm{d}\Omega_k^2\right]. \tag{8.50}$$

Defining the function $\mathcal{K}$ through $K_{rr} = B(r)\mathcal{K}(r)$, the constraint equations simply read[22]

$$(d-1)\frac{\omega'(r)}{r^{d-1}} = 2\mathcal{E}(r) + \frac{d}{d-1}\mathcal{K}(r)^2, \tag{8.51}$$

$$\frac{1}{r^d}\frac{\mathrm{d}}{\mathrm{d}r}\left[r^d\mathcal{K}(r)\right] = J_r(r). \tag{8.52}$$

[22]Note that (8.51) and (8.52) are true independent of the particular matter we are studying.

These equations are linear ODEs and are readily solved to give

$$\begin{aligned}
\mathcal{K}(r) &= \frac{1}{r^d}\left\{\mathcal{K}(r_0)r_0^d + \int_{r_0}^{r} \mathrm{d}\rho\rho^d \left[\mathcal{D}_t\phi(\mathcal{D}_r\phi)^\dagger, +\mathcal{D}_r\phi(\mathcal{D}_t\phi)^\dagger\right]\right\} \\
\omega(r) &= e^{-h(r)}\left[\omega(r_0) + \frac{1}{d-1}\int_{r_0}^{r}\mathrm{d}\rho e^{h(\rho)}\rho^{d-1}\chi(\rho)\right], \\
h(r) &= \frac{1}{d-1}\int_{r_0}^{r}\mathrm{d}\rho\rho\left(2|\mathcal{D}_r\phi|^2 + F_{tr}^2\right), \\
\chi(r) &= \frac{d}{d-1}\mathcal{K}(\rho)^2 + 2\left(k+\frac{\rho^2}{L^2}\right)|\mathcal{D}_r\phi|^2 + 2|\mathcal{D}_t\phi|^2 + 2V(\phi) \\
&\quad + \left(k+\frac{\rho^2}{L^2}\right)F_{tr}^2 + \frac{1}{2}F^{ij}F_{ij}.
\end{aligned} \tag{8.53}$$

The solution for a canonically normalized real scalar is obtained by sending $\phi \to \frac{1}{\sqrt{2}}\phi$ and replacing $\mathcal{D}_a$ with ∂_a.

Noting now that $K_{ab}r^a r^b = \mathcal{K}$ in an outermost coordinate patch where r^α points to increasing r, we find that the complexity change for spherical or planar symmetry is

$$\frac{\mathrm{d}\mathcal{C}_V}{\mathrm{d}\tau} = \frac{\Omega_k}{G_N L(d-1)}\lim_{r\to\infty} r^d\mathcal{K}(r), \tag{8.54}$$

while the mass reads

$$M = \lim_{r\to\infty} M_\Sigma[\sigma_r] = \frac{(d-1)\Omega_k}{16\pi G_N}\omega(\infty). \tag{8.55}$$

Thus, proving Lloyd's bound now amounts to constraining the asymptotic values of the solution (8.53).

Some technical lemmas

We begin by establishing positivity of ω:

Lemma 10. *$\omega(r)$ is positive on any complete maximal volume slice in an $AAdS_{d+1\geq 4}$ spacetime satisfying the WCC with spherical or planar symmetry.*

Proof. Consider first a one-sided spacetime. A finite Ricci and Kretschmann scalar as $r \to 0$ on Σ requires $\lim_{r\to 0}\omega(r) = 0$, which together with monotonicity from Lemma 8 implies

$$\omega(r) \geq 0. \tag{8.56}$$

In the two-sided case, the existence of a globally minimal surface means there is some minimal radius r_0 where

$$\omega(r_0) = kr_0^{d-2} + \frac{r_0^d}{L^2}, \tag{8.57}$$

which for $k \in \{0, 1\}$ is positive, and so by monotonicity we have $\omega(r) \geq 0$. □

Next, it will be convenient to introduce

$$\begin{aligned}\delta_\pm \equiv 2\left|\mathcal{D}_t\phi \mp \frac{r}{L}\mathcal{D}_r\phi\right|^2 + 2\left(k - \frac{\omega(r)}{r^{d-2}}\right)|\mathcal{D}_r\phi|^2 + 2V(\phi,\phi^\dagger) \\ + \frac{d}{d-1}\mathcal{K}(r)^2 + \frac{1}{B(r)}F_{tr}^2 + \frac{1}{2}F^{ij}F_{ij},\end{aligned} \tag{8.58}$$

so that the constraint (8.51) can be written

$$(d-1)\frac{\omega'}{r^{d-1}} = \pm\frac{2r}{L}J_r + \delta_\pm(r). \tag{8.59}$$

In terms of $\delta_\pm$ we can then state the following lemma:

Lemma 11. *Consider the solution (8.53) with spherical symmetry, $d \geq 3$ and $V(\phi,\phi^\dagger) \geq 0$, so that the WCC holds. Assume that*

$$1 - \frac{\omega(r)}{r^{d-2}} \tag{8.60}$$

vanishes somewhere. Then there exists a radius $\hat{r}$ such that

$$1 - \frac{\omega(\hat{r})}{\hat{r}^{d-2}} = 0, \tag{8.61}$$

$$2\mathcal{E}(\hat{r}) + \frac{d}{d-1}\mathcal{K}(\hat{r})^2 \leq \frac{(d-1)(d-2)}{\hat{r}^2}, \tag{8.62}$$

$$\delta_\pm(r) \geq 0 \qquad \forall r \geq \hat{r}, \tag{8.63}$$

and

$$\pm\int_{\hat{r}}^\infty \mathrm{d}r r^d J_r(r) = \frac{(d-1)L}{2}\left[\omega(\infty) - \hat{r}^{d-2} - \frac{1}{d-1}\int_{\hat{r}}^\infty \mathrm{d}r r^{d-1}\delta_\pm(r)\right]. \tag{8.64}$$

Proof. Denote for convenience $P(r) = 1 - \frac{\omega(r)}{r^{d-2}}$. Since $\omega(r)$ converges to a finite positive number at $r = \infty$ and $d > 2$, P is everywhere positive above some large radius. Since P is negative somewhere, we have by continuity that there must be a last zero of P that is approached from negative P, so that $P' > 0$ there. Let us denote this zero by $r = \hat{r}$.

Writing (8.51) in terms of P, we find

$$P'(r) = \frac{(d-2)}{r}(1-P) - \frac{r}{d-1}\left[2\mathcal{E} + \frac{d}{d-1}\mathcal{K}(r)^2\right], \tag{8.65}$$

which through $P'(\hat{r}) \geq 0, P(\hat{r}) = 0$ implies

$$\frac{d-2}{\hat{r}} - \frac{\hat{r}}{d-1}\left[2\mathcal{E}(\hat{r}) + \frac{d}{d-1}\mathcal{K}(\hat{r})^2\right] \geq 0, \tag{8.66}$$

giving (8.62).

Next, note that a single coordinate system is valid for all $r \geq \hat{r}$ since $\frac{1}{B(r)} = r^2 + P$ must be strictly positive there. Then, integrating (8.59) from $r = \hat{r}$ up to $r = \infty$ and remembering that $\omega(\hat{r}) = \hat{r}^{d-2}$, we find

$$(d-1)\left[\omega(\infty) - \hat{r}^{d-2}\right] = \int_{\hat{r}}^{\infty} \mathrm{d}r r^{d-1}\left[\pm\frac{2r}{L}J_r + \delta_{\pm}(r)\right], \tag{8.67}$$

yielding (8.64).

Finally, note that the only potentially negative term in $\delta_{\pm}$ is $2P|\mathcal{D}_r\phi|^2$. However, since P is positive above $\hat{r}$, so is $\delta_{\pm}$, giving (8.63). □

Proving a Lloyd's bound in Einstein-Maxwell-Scalar theory

We now finally prove our bound on complexity growth (Theorem 20):

Proof. Assume first that $P = 1 - \frac{\omega(r)}{r^{d-2}}$ is not everywhere positive. The two-sided case will always be this category, since at a minimal surface, which must always be present on a complete slice, we have $P = -r^2$. Then by Lemma 11 there exists an $r = \hat{r} > 0$ such that

$$\pm\int_{\hat{r}}^{\infty} \mathrm{d}r r^d J_r(r) \leq \frac{(d-1)L}{2}[\omega(\infty) - \hat{r}^{d-2}]. \tag{8.68}$$

Adding $\pm\mathcal{K}(\hat{r})\hat{r}^d$ to both sides of the inequality, we see from the solution of the constraint (8.53) that we get

$$\pm\lim_{r\to\infty} r^d\mathcal{K}(r) \leq \frac{(d-1)L}{2}\left[\omega(\infty) - \hat{r}^{d-2} \pm \frac{2}{(d-1)L}\mathcal{K}(\hat{r})\hat{r}^d\right] \tag{8.69}$$

Let us now without loss of generality assume that $\mathcal{K}(\hat{r}) \geq 0$ – the proof of the opposite sign is entirely analogous. Taking the lower sign inequality, multiplying by $-\frac{\Omega_{+1}}{G_N L(d-1)}$, neglecting the two positive terms proportional to $\mathcal{K}$ and $\hat{r}^{d-2}$, and using (8.54) and (8.55), we get

$$\dot{\mathcal{C}}_V \geq -\frac{8\pi M}{d-1}. \tag{8.70}$$

Next, (8.62) together with the positivity of $\mathcal{E}$ from the WEC gives $\mathcal{K}(\hat{r})^2 \leq \frac{(d-1)^2(d-2)}{d\hat{r}^2}$. Using this, the upper sign inequality reads

$$\lim_{r\to\infty} r^d\mathcal{K}(r) \leq \frac{(d-1)L}{2}\left[\omega(\infty) + \hat{r}^{d-2}\left(\frac{2}{L}\sqrt{\frac{d-2}{d}}\hat{r} - 1\right)\right]. \tag{8.71}$$

By monotonicity of ω we know that

$$\hat{r}^{d-2} = \omega(\hat{r}) \leq \omega(\infty). \tag{8.72}$$

Thus, if

$$\frac{2}{L}\sqrt{\frac{d-2}{d}}\omega(\infty)^{\frac{1}{d-2}} - 1 \leq 0, \tag{8.73}$$

then the second bracket of (8.71) is negative, and so we find $\dot{\mathcal{C}}_V \leq \frac{8\pi M}{d-1}$ after multiplying by $\Omega_{+1}/G_N L(d-1)$. In terms of a mass, the bound (8.73) reads

$$M \leq \frac{(d-1)\,\mathrm{vol}[S^{d-1}]L^{d-2}}{16\pi G_N}\left[\frac{d}{4(d-2)}\right]^{\frac{d-2}{2}} \equiv \hat{M}, \tag{8.74}$$

where we use the notation $\Omega_{+1} = \mathrm{vol}[S^{d-1}]$. If M does not satisfy this bound, we instead neglect the $-\hat{r}^{d-2}$ term in (8.71) and use $\hat{r}^{d-1} \leq \omega(\infty)^{\frac{d-1}{d-2}}$. Multiplying by the usual factor we get

$$\dot{\mathcal{C}}_V \leq \frac{8\pi M}{d-1}\left[1 + \left(\frac{M}{\hat{M}}\right)^{\frac{1}{d-2}}\right]. \tag{8.75}$$

Since we proved $\dot{\mathcal{C}}_V \geq -\frac{8\pi M}{d-1}$ for any mass, the above bound also holds for the absolute value, and so our bound is proven in the case where P vanishes somewhere.

Assume now $P \geq 0$ everywhere, which is only possible in the one-sided case. Then we can cover the whole of Σ with one coordinate system, and (8.68) holds with $\hat{r} = 0$, which immediately gives $|\dot{\mathcal{C}}_V| \leq \frac{8\pi}{d-1}M$ after multiplying an overall factor.

Finally, we note that the proof for real scalars is entirely analogous, and the only change with multiple fields is that $\delta_\pm$ contains a linear sum over the various fields. Lemma 11 remains true in this case also. Thus the above proof applies equally well to any number of gauge fields and scalars. □

8.3.2 A simple formula for $\dot{\mathcal{C}}_V$ for matter of compact support

Theorem 21. *Consider an AAdS spacetime with spherical or planar symmetry, and let r be the area radius. Let Σ be a maximal volume slice and assume that the matter has support only for $r \leq \rho$ on Σ. Let σ_ρ be the $r = \rho$ surface. Then*

$$\dot{\mathcal{C}}_V^2 = \frac{\mathrm{Area}[\sigma_\rho]}{4G_N}\frac{64\pi\rho}{(d-1)L^2}(M - M_\Sigma[\sigma_\rho]), \tag{8.76}$$

If in addition the spacetime satisfies the WCC, then

$$\dot{\mathcal{C}}_V^2 \leq \frac{16\pi\rho\,\mathrm{Area}[\sigma_\rho]}{(d-1)G_N L^2}M. \tag{8.77}$$

Proof. Since matter has compact support, we can explicitly solve (8.51) and (8.52) outside the support of the matter:

$$\begin{aligned}\mathcal{K}(r) &= \frac{\rho^d}{r^d}\mathcal{K}(\rho)\\ \omega(r) &= \omega(\rho) + \frac{1}{(d-1)^2}\mathcal{K}(\rho)^2\left(\frac{\rho}{r}\right)^d(r^d - \rho^d),\end{aligned} \tag{8.78}$$

Using that

$$\begin{aligned}\dot{\mathcal{C}}_V &= \frac{\Omega_k}{G_N L(d-1)}\rho^d\mathcal{K}(\rho),\\ M_\Sigma(\rho) &= \frac{(d-1)\Omega_k}{16\pi G_N}\omega(\rho),\end{aligned} \tag{8.79}$$

we find

$$M = \frac{(d-1)G_N L^2}{16\pi\Omega_k}\frac{1}{\rho^d}\dot{\mathcal{C}}^2 + M_\Sigma(\rho), \tag{8.80}$$

giving (8.76). If the WCC holds, then we know that $M_\Sigma(\rho) \geq 0$ from Lemma 10, giving (8.77). □

These formulas make reference to bulk quantities, and so they are useful only when working on the gravitational side. Furthermore, the compact support restriction makes the result less relevant when gauge fields are present. But the bound can be useful for other kinds of matter, or when compact support is a good approximation. Note also that we do not need compact support in spacetime – only on a spatial slice.

8.4 Discussion

The volumes of maximal slices are among the most natural diffeomorphism invariant gravitational observables in AAdS spacetimes; these are sensitive to the black hole interior and more generally constitute a more fine-grained gravitational observable than e.g. the areas of extremal surfaces. The Complexity=Volume relation stands to shed light on significant aspects of the holographic correspondence if the details of the proposal can be made precise and the proposal can be rigorously established (see [127] for some steps in this direction).

Here we have investigated the consistency of CV from a bulk perspective: if the proposal is a fundamental entry in the holographic dictionary, it dictates constraints on the behavior of maximal volume slices that should be provable independently using just geometry, analogous to the geometric proof of strong subadditivity of holographic entanglement entropy [3, 128], entanglement wedge nesting [3], causal wedge inclusion [3], etc. Under an interpretation of $\mathcal{C}_F$ as complexity with the vacuum as reference state, vacuum-subtracted volumes must be strictly positive in all spacetimes not identical to pure AdS. We established this result rigorously in broad generality in four bulk dimensions, assuming the weak energy condition. We have also established a weaker statement in other dimensions, that the vacuum-subtracted volume is positive in spacetimes with sufficient symmetry or in perturbations of the AdS vacuum [44] (again assuming the weak energy condition). The more general statement for arbitrary spacetimes satisfying the weak energy condition would follow from a modification of a well-known mathematical conjecture [43, 66] from compact to conformally compact manifolds.

Until now, broadly applicable results on maximal volume slices in holography have been sparse in comparison with those on (quantum) extremal surfaces.[23] This gap is at least partly a consequence of relatively few available techniques for maximal volumes; the holographic entanglement entropy proposal benefited from a readily-available arsenal of geometric tools controlling the behavior of codimension-two surfaces long predating holography [129–131]. Here we initiated the construction of a similar toolbox for maximal volumes, adapting results from mathematics [38, 39, 64, 65] in four dimensions and developing new techniques in general D. The utility of our technology is immediate: beyond the positive complexity theorem in four dimensions, the new tools have given a derivation of a version of a Lloyd's bound for

[23] Although not entirely absent: see [93] for example.

spatially symmetric maximal volume slices in a large class of matter, which is thus far the broadest proof of the bound on holographic complexity growth. We have also shown that wormhole complexity is indeed bounded from below by the thermofield double (with given energy) in general spacetimes satisfying the weak energy condition.

It would be interesting to see if our methods could be refined to derive stronger bounds on complexity growth when charge is present, given that our proof amounted to writing a formula $(d-1)|\dot{\mathcal{C}}| = 8\pi M - \Delta$ where there pure gauge field terms always contribute positively to Δ.[24] A strictly positive lower bound on their contribution to Δ in terms of charge would give a strengthening of our bound. It is also a possibility that our bound on complexity growth is not strict, and that the bound is true with $f(M) = 1$. Yet another possible avenue for further exploration is in spacetimes of different asymptotics, such as those of [132–136].

The ubiquity of the weak energy condition as an assumption in our theorems raises a potential question: why must we exchange the null energy condition, typically used to prove consistency in the holographic entropy context, for the more restrictive weak energy condition? As we will show in a companion article, the weak energy condition is in fact necessary, and violations of it such as vacuum decay can indeed result in negative vacuum-subtracted complexity, even in wormhole geometries. A potential conclusion is that the gap between the weak energy condition and the null energy condition is sourced by classical matter whose fine-grained properties in holography are qualitatively different from weak energy condition respecting fields in ways that are less obvious in coarser observables such as entropy. In particular, $\mathcal{C}_F$ as currently defined cannot be reinterpreted as the complexity itself for spacetimes with such matter (consisting of e.g. tachyonic scalars satisfying the Breitenlohner-Freedman bound.)

The assumption of hyperbolicity in our theorems is a more innocuous one in AdS/CFT, although it is nevertheless possible to violate it via the inclusion of compact dimensions. That is, an asymptotically $\mathrm{AdS}_{d+1} \times K$ spacetime, K can be picked to be a compact space that spoils the assumption of hyperbolicity even for static slices of pure AdS (when the compact dimensions are included). In an upcoming paper [41] we show that when this happens (which can be the case in many supergravity theories of interest), then the comparison theorems can be violated even when the weak energy condition holds, resulting in negative $\mathcal{C}_F$. This suggests that the effects of compact dimensions on the CV proposal are nontrivial and deserve further study.

Let us now briefly discuss a potential application of our results to the open question of holographic complexity as applied to subregions.

Mixed state holographic complexity: Given a reduced density matrix ρ_R, what is the holographic dual of the least complex purification of ρ_R (under the constraint of no unentangled qubits in the purifier)? This question was initially asked by [137], who dubbed the corresponding quantity the purification complexity $C_P(\rho_R)$.[25] The natural bulk dual to this quantity would require a minimax procedure: consider the maximal volume slices of all

[24] More precisely, we have $\pm(d-1)\dot{\mathcal{C}}_V = 8\pi M - \Delta_\pm$ where $\Delta_\pm$ is proportional to the intergral of the quantity $r^{d-1}\delta_\pm$ given by (8.58).

[25] It was proposed by [13] that the relevant geometric quantity is the volume of Σ_{E_R}, the maximal volume slice of the entanglement wedge of R. However, this proposal falls short of satisfying the requisite qualitative properties predicted by tensor network models [137].

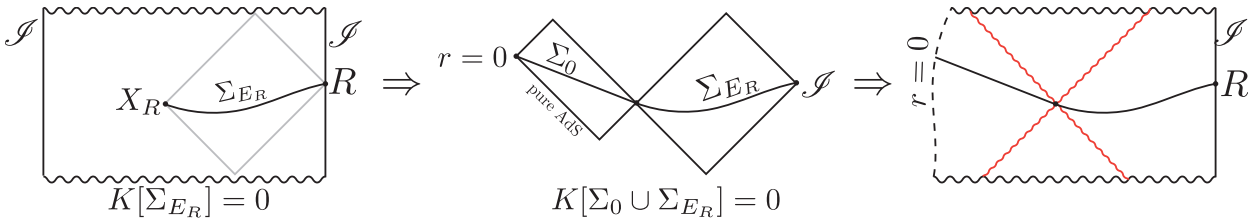


Figure 8.6: Example of the proceedure used to build the simplest WCC respecting spherically symmetric spacetime (at time R) containing the entanglement wedge E_R of R. The extremal Cauchy slice Σ_{E_R} of an entanglement wedge E_R (left) is glued to a ball Σ_0 of the hyperbolic plane with extrinsic curvature $K_{ab} = 0$ (center). Then the maximal Cauchy evolution of $\Sigma_{E_R} \cup \Sigma_0$ gives the wanted spacetime (right). The red wiggly lines describe positive energy shocks.

possible classical[26] bulks that complete the entanglement wedge of R into an inextendible spacetime, and then minimize its volume over the spacetime geometries. The relevant question then appears to be: given ρ_R, what is the spacetime of least complexity containing the entanglement wedge of R?

The volume of the maximal volume slice of this spacetime is the obvious candidate to $\mathcal{C}_P(\rho_R)$.[27] The results from Sec. 8.2 allow for a concrete computation. Consider the case of a two-sided spherically symmetric connected spacetime, and take R to be a complete timeslice of one of the two conformal boundaries. Assuming the weak energy condition, a computation carried out in Appendix 8.5.5 proves that the spherically symmetric spacetime of least $\mathcal{C}_V$ complexity that contains E_R is a one-sided spacetime with $\mathcal{C}_V$ complexity

$$\frac{V(\Sigma_{E_R})}{G_N L} + S_{\text{vN}}(\rho_R) \frac{4r}{dL} {}_2F_1\left(\frac{1}{2}, \frac{d}{2}, \frac{2+d}{2}, -\frac{r^2}{L^2}\right), \tag{8.81}$$

where r is the area radius of the HRT surface.[28] The above is just the volume obtained after gluing a ball of the hyperbolic plane to the maximal volume slice of the entanglement wedge, so that the resulting surface is a complete initial data set (with $K = 0$). See Fig. 8.6 for an illustration.

Ref. [137] argued for several qualitative features of $\mathcal{C}_P$ that are satisfied by (8.81). First, $\mathcal{C}_P(\rho_R)$ is expected to go as $c_1 n + |c_2| S(\rho_R)$ for unknown c_i, where n is the number of qubits

[26] It is of course possible that the optimal purification does not have a semiclassical dual. However, at least in some cases – e.g. for optimized (n-point) correlation measures – it was established in [138] that a minimization over semiclassical bulk spacetimes will in fact accomplish a global minimum over a holographic CFT's Hilbert space.

[27] This way of computing $\mathcal{C}_P(\rho_R)$ was proposed in [139], although they also minimized the complexity over bulk cutoffs, allowing the bulk cutoff to move deep into the bulk. Then then solution is that Σ_{E_R} itself is the Cauchy slice of the purified bulk spacetime, with the conformal boundary moved in sufficiently far that the HRT surface itself now corresponds to a piece of the conformal boundary. This leads to a peculiar situation where the state proposed to purify ρ_R is a state on a QFT where the UV cutoff drastically differs in different regions of space. This appears in tension with purity, since the theory on the HRT surface must purify the state on the real conformal boundary, even as the UV cutoff is taken arbitrarily small and the number of qubits there grows without bound. Avoiding the minimization over the cutoff altogether appears to be the more natural option.

[28] We cannot rule out that there is a spacetime that breaks spherical symmetry with even lower complexity.

in the state. Since the leading UV-divergence of the volume of Σ_{E_R} is of the form $\text{vol}[R]/\epsilon^{d-1}$, which can be thought of as counting the number of lattice sites in the discretized theory, the two terms in Eq. (8.81) have this form. Second, $\mathcal{C}_P$ was argued to satisfy subadditivity for the thermofield double state: $2\mathcal{C}_P(\rho_\beta) > \mathcal{C}(|\text{TFD}, \beta\rangle)$. Again, this is realized by our result when applied to one side of a Schwarzschild black hole, unlike in the case where the second term of (8.81) is not included.

We can make further progress on this front by making the assumption that the optimal purification is always given by completing Σ_{E_R} with a compact subset H of the hyperbolic plane.[29] This assumption is motivated by the conformally compact generalization of Schoen's conjecture and the following theorem (together with the fact that hyperbolic space has constant sectional curvature sectional curvature $-\frac{1}{L^2}$):

Theorem 22 ([140])**.** *Let Σ be a compact domain in a d-dimensional complete simply connected Riemannian manifold with sectional curvature bounded from above by $-\kappa^2$. Then*

$$\text{vol}[\Sigma] \leq \frac{\text{Area}[\partial\Sigma]}{(d-1)|\kappa|}. \tag{8.82}$$

This immediately yields the following lower and upper bounds on purification complexity:

$$\frac{\text{vol}[\Sigma_{E_R}]}{G_N L} \leq \mathcal{C}_P(\rho_R) \leq \frac{\text{vol}[\Sigma_{E_R}]}{G_N L} + \frac{4}{d-1} S_{\text{vN}}(\rho_R). \tag{8.83}$$

Note that if the WCC is violated but there is still a lower bound on local energy density, we could get a similar bound, but with an altered prefactor in the entropy term.

Acknowledgments

It is a pleasure to Chris Akers, Simon Brendle, Shira Chapman, Otis Chodosh, Sebastian Fischetti, Daniel Harlow, Daniel Roberts, Shreya Vardhan and Ying Zhao for discussions. This work is supported in part by NSF grant no. PHY-2011905 and the MIT department of physics. The work of NE was also supported in part by the U.S. Department of Energy, Office of Science, Office of High Energy Physics of U.S. Department of Energy under grant Contract Number DE-SC0012567 (High Energy Theory research) and by the U.S. Department of Energy Early Career Award DE-SC0021886. The work of ÅF is also supported in part by an Aker Scholarship.

[29] Of course we should prescribe K_{ab} to have initial data. One can show that gluing a subset of the hyperbolic plane to an extremal surface respects the WCC and the NEC if we assign $K_{ab} = 0$ to H. If the AdS WCC does not hold but there is still a lower bound on energy density in our theory, we should then choose hyperbolic space with the smallest cosmological constant consistent with our minimal energy density, so as to minimize its volume.

8.5 Appendix

8.5.1 Asymptotic hyperbolicity

Let now $(\tilde{\Sigma}, \tilde{h})$ be hyperbolic 3-space, with metric

$$\mathrm{d}s^2 = \frac{1}{1+r^2}\mathrm{d}r^2 + r^2\mathrm{d}\Omega^2. \tag{8.84}$$

Consider a second Riemannian 3-manifold (Σ, h), potentially with boundary, where $\Sigma = \tilde{\Sigma}\setminus K$ for a compact (possibly empy) set K. Then (Σ, h) is in Ref. [38] defined to be asymptotically hyperbolic if

$$\begin{aligned} |h-\tilde{h}|_{\tilde{h}} &= \mathcal{O}(r^{-2-\delta}), \quad \delta > 0, \\ \lim_{r\to\infty} |\tilde{D}(h-\tilde{h})|_{\tilde{h}} &= 0, \end{aligned} \tag{8.85}$$

where $|\cdot|_{\tilde{h}}$ is the pointwise tensor norm taken with respect to $\tilde{h}$.

Consider now an AAdS_4 spacetime (M, g) and pick a boundary conformal frame so that

$$\mathrm{d}s^2|_{\partial M} = -\mathrm{d}t^2 + \mathrm{d}\Omega^2. \tag{8.86}$$

We proceed now to show that an extremal hypersurface Σ in an AAdS_4 spacetime anchored at $t = \text{const}$ slice on the boundary is asymptotically hyperbolic.

First, note that the Fefferman-Graham coordinates of pure AdS_4 adapted to the Einstein static universe of unit spatial radius on the boundary reads

$$\mathrm{d}s^2 = \frac{1}{z^2}\left[\mathrm{d}z^2 - \frac{(4+z^2)^2}{16}\mathrm{d}t^2 + \frac{(4-z^2)^2}{16}\mathrm{d}\Omega^2\right]. \tag{8.87}$$

Going to a general AAdS_4 spacetime with the falloffs we are considering, we have

$$\mathrm{d}s^2 = \frac{1}{z^2}\left[\mathrm{d}z^2 - \left(1+\frac{z^2}{2}\right)\mathrm{d}t^2 + \left(1-\frac{z^2}{2}\right)\mathrm{d}\Omega^2 + z^3\mathcal{T}_{ij}(x)\mathrm{d}x^i\mathrm{d}x^j + \mathcal{O}(z^4)\right], \tag{8.88}$$

where i, j are boundary coordinate indices running over t, θ, φ.

Consider now a spatial hypersurface Σ in (M, g) that is extremal, and with intrinsic coordinates $x^\alpha = (z, \theta, \varphi)$ and embedding coordinates $X^\mu = (z, t(z,\theta,\varphi), \theta, \varphi)$. Generically, once the boundary anchoring time is fixed, there is only one slice that will be smooth and complete in the bulk, and so the integration constant which determines how Σ leaves the boundary is fixed once the boundary anchoring location of Σ is fixed. However, from a near boundary analysis it is impossible to know this integration constant, so we will have to allow it to be general.

Let us take the ansatz for the expansion of $t(z, \theta, \varphi)$ near the boundary to be given by

$$t(z,\theta,\varphi) = \sum_{n=0}^{\infty} t_n(\theta,\varphi) z^n. \tag{8.89}$$

Since we consider a slice of constant t on the boundary, we have that t_0 is a constant. With this expansion we can compute the mean curvature K in a small$-z$ expansion and demand

that it vanishes order by order in z. The exact form of the equations are ugly and not needed. We only need the basic structure. We find that

$$t_1(\theta,\varphi) = t_2(\theta,\varphi) = t_3(\theta,\varphi) = 0. \tag{8.90}$$

If the anchoring time t_0 was not constant, then t_2 would become nonzero and given by an algebraic expression of the derivatives of t_0. The function t_4 is the integration constant referred to above. It encodes deep bulk information and is fixed by requiring that Σ is a smooth slice. Higher t_n are fixed by t_0, t_4, $\mathcal{T}_{ij}$ and higher order terms in the metric (which in turn depends only on the boundary conformal structure and $\mathcal{T}_{ij}$).

The induced metric on Σ now reads

$$\begin{aligned} h_{\alpha\beta} &= g_{\alpha\beta} + g_{tt}\partial_\alpha t \partial_\beta t \\ &= \frac{1}{z^2}\begin{pmatrix} 1 & 0 & 0 \\ 0 & (1-\frac{z^2}{2}) + z^3\mathcal{T}_{\theta\theta} & \mathcal{T}_{\theta\phi}z^3 \\ 0 & \mathcal{T}_{\phi\theta}z^3 & (1-\frac{1}{2}z^2)\sin\theta^2 + \mathcal{T}_{\phi\phi}z^3 \end{pmatrix} + \mathcal{O}(z^2). \end{aligned} \tag{8.91}$$

In the above, higher order terms depend both on the spacetime geometry and t_4. Since the t_4 dependence is in the higher order term, setting $\mathcal{T}_{ij} = 0$ gives the metric of the hyperbolic plane, $\tilde{h}_{\alpha\beta}$, up to $\mathcal{O}(z^2)$ corrections. Defining for notational convenience $\mathcal{T}_{\alpha\beta} = \mathcal{T}_{ij}$ when $\alpha,\beta,i,j \in \{\theta,\varphi\}$ and $\mathcal{T}_{z\alpha} = 0$, we have

$$\begin{aligned} |h-\tilde{h}|^2_{\tilde{h}} &= \tilde{h}^{\alpha\gamma}\tilde{h}^{\beta\delta}(h-\tilde{h})_{\alpha\beta}(h-\tilde{h})_{\gamma\delta} \\ &= \frac{1}{z^4}\tilde{h}^{\alpha\gamma}\tilde{h}^{\beta\delta}\left(z^3\mathcal{T}_{\alpha\beta} + \mathcal{O}(z^4)\right)\left(z^3\mathcal{T}_{\gamma\delta} + \mathcal{O}(z^4)\right) \\ &= \mathcal{O}(z^6). \end{aligned} \tag{8.92}$$

But to leading order, we have $z = \frac{1}{r} + \mathcal{O}(r^{-2})$, where r is the coordinate used to define asymptotic hyperbolicity. Thus we find

$$|h-\tilde{h}|_{\tilde{h}} = \mathcal{O}(r^{-3}), \tag{8.93}$$

meaning that Σ satisfies the first condition for asymptotic hyperbolicity. Next, note that

$$\tilde{D}_\gamma(h-\tilde{h})_{\alpha\beta} = \tilde{D}_\gamma\left[z\mathcal{T}_{\alpha\beta}(x) + \mathcal{O}(z^2)\right] = \mathcal{O}(1). \tag{8.94}$$

Since the three inverse metrics involved in calculating $|\tilde{D}(h-\tilde{h})|^2$ brings a total power of z^6, we find that

$$|\tilde{D}(h-\tilde{h})| = \mathcal{O}(z^3) = \mathcal{O}(r^{-3}) \tag{8.95}$$

and so the second condition,

$$\lim_{r\to\infty}|\tilde{D}(h-\tilde{h})| = 0, \tag{8.96}$$

holds. Hence Σ is asymptotically hyperbolic.

Finally we note that if t_0 was not constant, then t_2 would not vanish, and we would have $\mathcal{O}(1)$ corrections in h_{zz} depending on t_2. This factor would not be present for the metric of hyperbolic space, and so $(h-\tilde{h})_{\alpha\beta}$ would now be $\mathcal{O}(z^2)$ rather than $\mathcal{O}(z^3)$, and so the falloff in Eq. (8.92) would end up being $\mathcal{O}(z^4)$ instead, which would mean Σ was not asymptotically hyperbolic in general.

8.5.2 *d*-dimensional Geroch-Hawking-mass with a cosmological constant

Consider a maximal volume slice Σ of a spacetime with maximal spatial symmetry. The mean curvature of a constant$-r$ surface σ in Σ in the coordinate system (8.29) reads

$$H = D_\alpha r^\alpha = \frac{d-1}{r\sqrt{B(r)}}, \tag{8.97}$$

where $r^\alpha = \frac{1}{\sqrt{A}}(\partial_r)^\alpha$ is the unit normal pointing to increasing r. We find

$$H^2 = \frac{(d-1)^2}{r^2}\left(k + \frac{r^2}{L^2} - \frac{\omega(r)}{r^{d-2}}\right). \tag{8.98}$$

Noting that a constant$-r$ surface has intrinsic Ricci scalar

$$\mathcal{R} = k\frac{(d-1)(d-2)}{r^2},$$

we can rewrite ω as follows:

8.5.3 Monotonicity of the volume of Stat_k with respect to mass

Assume $d > 2$ and consider the regularized volume of (half a) totally geodesic slice of Stat_k:

$$V(r_h) = \int_{r_h}^{r_c} \mathrm{d}r r^{d-1}\left[k + r^2 - \left(\frac{r_h}{r}\right)^{d-2}(k + r_h^2)\right]^{-1/2}, \tag{8.99}$$

where we pick units where $L = 1$ and divide out the overall factor of Ω_k. Changing now variables

$$r = e^t r_h \qquad \mathrm{d}r = r\mathrm{d}t, \quad \tau \equiv t(r_c) = \log\left(\frac{r_c}{r_h}\right), \tag{8.100}$$

we get

$$\begin{aligned} V(r_h) &= r_h^d \int_0^\tau \mathrm{d}t e^{dt}\left[k + e^{2t}r_h^2 - e^{-(d-2)t}(k + r_h^2)\right]^{-1/2} \\ &= r_h^d \int_0^\tau \mathrm{d}t e^{(d-1)t}\left[k\left(e^{-2t} - e^{-dt}\right) + r_h^2\left(1 - e^{-dt}\right)\right]^{-1/2} \end{aligned} \tag{8.101}$$

and

$$\begin{aligned} V(e^\alpha r_h) &= r_h^d \int_0^{\tau-\alpha} \mathrm{d}t e^{d(t+\alpha)}\left[k\left(1 - e^{-(d-2)t}\right) + r_h^2 e^{2\alpha}\left(e^{2t} - e^{-(d-2)t}\right)\right]^{-1/2} \\ &= r_h^d \int_\alpha^\tau \mathrm{d}t e^{dt}\left[k\left(1 - e^{-(d-2)(t-\alpha)}\right) + r_h^2 e^{2\alpha}\left(e^{2(t-\alpha)} - e^{-(d-2)(t-\alpha)}\right)\right]^{-1/2} \\ &= r_h^d \int_\alpha^\tau \mathrm{d}t e^{(d-1)t}\left[k\left(e^{-2t} - e^{(d-2)\alpha}e^{-dt}\right) + r_h^2\left(1 - e^{d\alpha}e^{-dt}\right)\right]^{-1/2}. \end{aligned} \tag{8.102}$$

Consider now the difference in volume of a slice with horizon $e^{\alpha}r_h$ and one with r_h. Furthermore, introduce a new regulator $\epsilon \ll 1$ by modifying the denominators above with the replacement $[\ldots] \to [\epsilon + \ldots]$. The (rescaled and ϵ-regulated) volume difference reads

$$\Delta_\epsilon(\alpha, r_h) \equiv \lim_{\tau\to\infty} \frac{V_\epsilon(e^{\alpha}r_h) - V_\epsilon(r_h)}{r_h^d}. \tag{8.103}$$

We have that $\partial_{r_h} V(r_h)$, $\partial_m I_k(m)$ and $\partial_\alpha \Delta_\epsilon(\alpha, r_h)|_{\alpha=0,\epsilon=0}$ all have the same sign, so let us focus on the latter.

We have

$$\begin{aligned}\Delta_\epsilon(\alpha, r_h) = \int_\alpha^\infty \mathrm{d}t e^{(d-1)t}\Big\{ &\left[\epsilon + k\left(e^{-2t} - e^{(d-2)\alpha}e^{-dt}\right) + r_h^2\left(1 - e^{d\alpha}e^{-dt}\right)\right]^{-1/2} \\ &- \left[\epsilon + k\left(e^{-2t} - e^{-dt}\right) + r_h^2\left(1 - e^{-dt}\right)\right]^{-1/2}\Big\} \\ - \int_0^\alpha \mathrm{d}t e^{(d-1)t} &\left[\epsilon + k\left(e^{-2t} - e^{-dt}\right) + r_h^2\left(1 - e^{-dt}\right)\right]^{-1/2},\end{aligned} \tag{8.104}$$

and so taking the α-derivative we obtain

$$\partial_\alpha \Delta_\epsilon(\alpha) = \int_\alpha^\infty \mathrm{d}t e^{(d-1)t}\partial_\alpha\{\ldots\} - e^{(d-1)\alpha}\epsilon^{-1/2}. \tag{8.105}$$

Next, note that

$$\partial_\alpha[\ldots]^{-1/2} = \frac{1}{2}[\ldots]^{-3/2} d\left(k\frac{d-2}{d}e^{-2\alpha} + r_h^2\right)e^{\alpha d}e^{-dt}. \tag{8.106}$$

Defining now $\mu = k\frac{d-2}{d} + r_h^2$, we find that

$$\begin{aligned}\partial_\alpha\Delta_\epsilon|_{\alpha=0} &= -\epsilon^{-1/2} \\ &\quad + \frac{d\mu}{2}\int_0^\infty \mathrm{d}t e^{-t}\left[\epsilon + \mu\left(1 - e^{-dt}\right) + k\left(e^{-2t} - e^{-dt}\right) - k\frac{d-2}{d}(1 - e^{-dt})\right]^{-3/2} \\ &= -\epsilon^{-1/2} + \frac{d\mu}{2}\int_0^\infty \mathrm{d}t e^{-t}\left[\epsilon + \mu\left(1 - e^{-dt}\right) + kf(t)\right]^{-3/2}\end{aligned} \tag{8.107}$$

where

$$f(t) \equiv e^{-2t} - e^{-dt} - \frac{d-2}{d}\left(1 - e^{-dt}\right). \tag{8.108}$$

Now we must consider various cases.

The case of $k \geq 0$:

For all $t \geq 0$ we have that

$$f(t) \leq 0. \tag{8.109}$$

Furthermore, for $k \geq 0$ we have $\mu \geq 0$ for all $r_h > 0$, and so replacing $f(t) \to 0$ we get the inequality

$$\partial_\alpha\Delta_\epsilon|_{\alpha=0} \geq -\epsilon^{-1/2} + \frac{d\mu}{2}\int_0^\infty \mathrm{d}t e^{-t}\left[\epsilon + \mu\left(1 - e^{-dt}\right)\right]^{-3/2}. \tag{8.110}$$

This intergral is a hypergeometric function whose expansion about $\epsilon = 0$ reads

$$\partial_\alpha \Delta_\epsilon|_{\alpha=0} \geq -\epsilon^{-1/2} + \frac{d\mu}{2}\left[\frac{2}{d\mu\sqrt{\epsilon}} - \frac{2\sqrt{\pi}\Gamma\left(1+\frac{1}{d}\right)}{\mu^{3/2}\Gamma\left(-\frac{1}{2}+\frac{1}{d}\right)} + \mathcal{O}(\sqrt{\epsilon})\right], \tag{8.111}$$

giving finally that

$$\partial_\alpha \Delta_\epsilon|_{\alpha=0,\epsilon=0} \geq -\frac{d\sqrt{\pi}\Gamma\left(1+\frac{1}{d}\right)}{\mu^{1/2}\Gamma\left(-\frac{1}{2}+\frac{1}{d}\right)} > 0, \tag{8.112}$$

where we use that $d > 2$ so that the Gamma-function in the denominator is negative. Thus, for a spherical or planar static black hole we have that the complexity of formation is positive and monotonically increasing with horizon radius, and thus also mass.

The case of $k = -1$:

Below we only show monotonicity for sufficiently large masses.

Consider $\mu > 0$. Then the integral in (8.107) can be expanded

$$\begin{aligned} J &\equiv \frac{d}{2\sqrt{\mu}}\int_0^\infty \mathrm{d}t e^{-t}\left[\epsilon\mu^{-1} + \left(1-e^{-dt}\right) - \mu^{-1}f(t)\right]^{-3/2} \\ &= \frac{d}{2\sqrt{\mu}}\sum_{n=0}^\infty c_n \mu^{-n}\int_0^\infty \mathrm{d}t e^{-t}\frac{f(t)^n}{\left[\epsilon\mu^{-1}+(1-e^{-dt})\right]^{3/2+n}}, \end{aligned}$$

where c_n are the positive coefficients appearing in

$$\frac{1}{(1-x)^{3/2}} = \sum_{n=0}^\infty c_n x^n.$$

Note that $f(t) \sim \mathcal{O}(t^2)$ at small t, so in this limit the numerator behaves as $\sim t^{2n}$ while the denominator behaves as $\sim (\epsilon+t)^{3/2+n}$. After carrying out the integral and sending $\epsilon \to 0$ only the $n = 0$ term diverges, so after carrying out the integral each $n \geq 1$ term scales as $\mathcal{O}(\epsilon^0)$. Thus up to $\mathcal{O}(\epsilon)$ corrections we have

$$J = \frac{d}{2\sqrt{\mu}}\int_0^\infty \mathrm{d}t e^{-t}\frac{1}{\left[\epsilon\mu^{-1}+1-e^{-dt}\right]^{3/2}} + \sum_{n=1}^\infty \frac{c_n d\mu^{-n-\frac{1}{2}}}{2}\int_0^\infty \mathrm{d}t \frac{e^{-t}f(t)^n}{\left[1-e^{-dt}\right]^{3/2+n}}.$$

The first term is exactly the integral computed above, so we find

$$\partial_\alpha \Delta_\epsilon|_{\alpha=0,\epsilon=0} = -\frac{d\sqrt{\pi}\Gamma\left(1+\frac{1}{d}\right)}{\mu^{1/2}\Gamma\left(-\frac{1}{2}+\frac{1}{d}\right)} + \sum_{n=1}^\infty \frac{c_n d\mu^{-n-\frac{1}{2}}}{2}\int_0^\infty \mathrm{d}t \frac{e^{-t}f(t)^n}{\left[1-e^{-dt}\right]^{3/2+n}}.$$

We can readily check that, for $t \geq 0$, we have the lower bound

$$f(t) \geq -\frac{d-2}{d}(1-e^{-dt})^2.$$

giving

$$f(t)^n \geq \begin{cases} 0 & n \text{ even} \\ -(\frac{d-2}{d})^n (1-e^{-dt})^{2n} & n \text{ odd.} \end{cases}$$

Thus we find

$$\begin{aligned} \partial_\alpha \Delta_\epsilon|_{\alpha=0,\epsilon=0} &\geq -\frac{d\sqrt{\pi}\Gamma\left(1+\frac{1}{d}\right)}{\mu^{1/2}\Gamma\left(-\frac{1}{2}+\frac{1}{d}\right)} - \sum_{n=1,\text{odd}}^{\infty} \frac{c_n d \mu^{-n-\frac{1}{2}}}{2} \left(\frac{d-2}{d}\right)^n \int_0^\infty \mathrm{d}t e^{-t} \left[1-e^{-dt}\right]^{n-\frac{3}{2}} \\ &= -\frac{d\sqrt{\pi}\Gamma\left(1+\frac{1}{d}\right)}{\mu^{1/2}\Gamma\left(-\frac{1}{2}+\frac{1}{d}\right)} - \sum_{n=1,\text{odd}}^{\infty} \frac{c_n d \mu^{-n-\frac{1}{2}}}{2} \left(\frac{d-2}{d}\right)^n \frac{\Gamma\left(1+\frac{1}{d}\right)\Gamma\left(-\frac{1}{2}+n\right)}{\Gamma\left(-\frac{1}{2}+\frac{1}{d}+n\right)}. \end{aligned}$$

This last sum can be carried out exactly and evalutes to a hypergeometric function, but the exact expression is not particularly useful. However, it does scale like $\mathcal{O}(\mu^{-3/2})$ at large μ, so since the first term which goes as $\mathcal{O}(\mu^{-1/2})$ has a positive coefficient, this means that there exists a $\hat{\mu} > 0$ such that $\partial_\alpha \Delta_\epsilon|_{\alpha=0,\epsilon=0} \geq 0$ for all $\mu \geq \hat{\mu}$.

8.5.4 Computing $\dot{\mathcal{C}}$

We study only spacetimes that have finite gravitational mass, so that the mass (8.55) coincides the CFT energy up to the Casimir energy. In the coordinates (8.29) this means that the following integrals must be assumed finite:

$$\int^\infty \mathrm{d}r r^{d-1} \mathcal{E}, \qquad \int^\infty \mathrm{d}r r^d J_r(\rho), \tag{8.113}$$

where J_r is given by

$$J_r = 8\pi G_N n^a (\partial_r)^b T_{ab}, \tag{8.114}$$

and where $8\pi G_N T_{ab} = R_{ab} - \frac{1}{2} g_{ab} R - \frac{d(d-1)}{2L^2} g_{ab}$.

We now want to express $\eta^a N_a$, as defined in Sec. 8.3.1, in terms of quantities over which we have control. To do this we temporarily introduce an ADM coordinate system x^μ on spacetime with vanishing shift, and lapse equal to r:

$$\mathrm{d}s^2|_M = g_{\mu\nu}\mathrm{d}x^\mu \mathrm{d}x^\nu = -r^2 \mathrm{d}t^2 + h_{\alpha\beta}(t, x^\alpha)\mathrm{d}x^\alpha \mathrm{d}x^\beta, \tag{8.115}$$

where

$$\begin{aligned} \mathrm{d}s^2|_\Sigma &= h_{\alpha\beta}(t=0, x^\alpha)\mathrm{d}x^\alpha \mathrm{d}x^\beta = B(r)\mathrm{d}r^2 + r^2 \mathrm{d}\Omega_k^2, \\ \partial_t h_{\alpha\beta}(x)|_{t=0} &= 2r K_{\alpha\beta}(x). \end{aligned} \tag{8.116}$$

We proceed to locate the conformal boundary in a small neighborhood around Σ. This will let us determine η^a, which is tangent to ∂M.

Let us for notational convenience now take the spherically symmetric case, and let us install coordinates $y^i = (\tau, \Omega)$ on ∂M and take $\partial M \cap \Sigma = \partial\Sigma$ to be located $(r = r_c, t = \tau = 0)$, where we temporarily work with a finite cutoff at $r = r_c$. We want to find embedding coordinates $(r(\tau), t(\tau))$ for ∂M so that the induced metric reads

$$\mathrm{d}s^2|_{\partial M} = \gamma_{ij}\mathrm{d}y^i \mathrm{d}y^j = \frac{r_c^2}{L^2}(-\mathrm{d}\tau^2 + L^2 \mathrm{d}\Omega^2). \tag{8.117}$$

Thus, we must solve the equations

$$\begin{aligned}\gamma_{\tau\tau} &= g_{rr}\dot{r}^2 - r^2\dot{t}^2 = -r_c^2/L^2,\\ \gamma_{\theta\theta} &= g_{\theta\theta} = r_c^2,\end{aligned} \tag{8.118}$$

where dots are derivatives with respect to τ. In fact we will only need $\dot{r}(0)$ and $\dot{\tau}(0)$. Taking a derivative of the second equation and then setting $\tau = 0$ gives the pair of equations

$$\begin{aligned}\dot{r}(0)^2 g_{rr}(0, r_c) - r_c^2\dot{t}(0)^2 &= -r_c^2/L^2,\\ \partial_t g_{\theta\theta}(0, r_c)\dot{t}(0) + \partial_r g_{\theta\theta}(0, r_c)\dot{r}(0) &= 0,\end{aligned} \tag{8.119}$$

or

$$\begin{aligned}\dot{r}(0)^2 B(r_c) - r_c^2\dot{t}(0)^2 &= -r_c^2/L^2,\\ 2r_c K_{\theta\theta}(r_c)\dot{t}(0) + 2r_c\dot{r}(0) &= 0,\end{aligned} \tag{8.120}$$

which are easily solved to give

$$\begin{aligned}\dot{t}(0) &= \frac{1}{\sqrt{1 - BK_{\theta\theta}^2 r^{-2}L^2}}\Big|_{r=r_c},\\ \dot{r}(0) &= -\frac{K_{\theta\theta}}{\sqrt{1 - BK_{\theta\theta}^2 r^{-2}L^2}}\Big|_{r=r_c},\end{aligned} \tag{8.121}$$

where we choose the branch where $\dot{t} > 0$. Remembering from Eq. (8.50) that

$$K_{\theta\theta} = -\frac{K_{rr}r^2}{(d-1)B(r)}, \tag{8.122}$$

we find that the term $BK_{\theta\theta}^2 r^{-2}$ at large r behaves like

$$BK_{\theta\theta}^2 r^{-2} \sim \frac{K_{rr}^2 r^2}{B} \sim K_{rr}^2 r^4 L^{-2} \tag{8.123}$$

The solution of (8.52) gives that $\mathcal{K}(r) \sim \mathcal{O}(r^{-d})$, so we have

$$K_{rr} = B(r)\mathcal{K}(r) \sim \frac{1}{r^{d+2}}, \tag{8.124}$$

implying that $BK_{\theta\theta}^2 r^{-2} \sim \mathcal{O}\left(r^{-2d}\right)$. We thus find

$$\begin{aligned}\dot{t}(0) &= 1 + \mathcal{O}\left(r_c^{-2d}\right),\\ \dot{r}(0) &= \frac{K_{rr}r^2}{(d-1)B(r)}\left(1 + \mathcal{O}\left(r^{-2d}\right)\right)\Big|_{r=r_c}.\end{aligned}$$

In our spacetime coordinates, η^a and N^a reads

$$\begin{aligned}\eta^a &= (\partial_\tau)^a = \dot{t}(0)(\partial_t)^a + \dot{r}(0)(\partial_r)^a,\\ N^a &= \frac{1}{\sqrt{B(r_c)}}(\partial_r)^a,\end{aligned}$$

giving finally that

$$\eta^a N_a = \sqrt{B(r_c)}\dot{r}(0) = \frac{K_{rr} r^2}{(d-1)\sqrt{B(r)}} \left[1 + \mathcal{O}\left(r^{-2d}\right)\right] \Big|_{r=r_c}. \tag{8.125}$$

This implies

$$\begin{aligned} \frac{\mathrm{d}V[\Sigma]}{\mathrm{d}\tau} &= \frac{\Omega_k r^{d+1} K_{rr}}{(d-1)\sqrt{B(r)}} \left[1 + \mathcal{O}(r^{-2d})\right] |_{r=r_c} \\ &= \frac{\Omega_k r^{d+2} K_{rr}}{d-1} \left[1 + \mathcal{O}(r^{-2}, \omega(r) r^{-d})\right] |_{r=r_c} \end{aligned} \tag{8.126}$$

This is clearly finite when $\omega(\infty)$ is finite. Furthermore $r^{d+2} K_{rr} \sim \mathcal{O}(1)$ and so we have

$$\frac{\mathrm{d}\mathcal{C}_V}{\mathrm{d}\tau} = \frac{\Omega_k}{G_N L(d-1)} \lim_{r\to\infty} r^{d+2} K_{rr} = \frac{\Omega_k}{G_N L(d-1)} \lim_{r\to\infty} r^d \mathcal{K}(r). \tag{8.127}$$

Note that the above derivations holds upon the replacement of $\mathrm{d}\Omega^2$ with $\mathrm{d}\boldsymbol{x}^2$, and θ with one of the Cartesian boundary directions. The hyperbolic case requires minor modifications, but is not needed for us.

If we plug in the solution of Eq. (8.52) into the expression for $\dot{\mathcal{C}}$, we find

$$\begin{aligned} \frac{\mathrm{d}\mathcal{C}_V}{\mathrm{d}\tau} &= \frac{\Omega_k}{G_N L(d-1)} \left[\mathcal{K}(r_0) r_0^d + \int_{r_0}^{\infty} \mathrm{d}r r^d J_r(r)\right] \\ &= \frac{1}{G_N L(d-1)} \left[r_0 \int_{\partial\Sigma_{\mathrm{out}}(r_0)} r^\alpha r^\beta K_{\alpha\beta} + 8\pi G_N \int_{\Sigma_{\mathrm{out}}(r_0)} r T_{\alpha\beta} n^\alpha r^\beta\right], \end{aligned} \tag{8.128}$$

If we take r_0 to be at an outermost stationary surface, $r_0 = r_{\mathrm{throat}}$, where $\mathcal{K}(r_0) = \sqrt{2\theta_k \theta_\ell}$, we find the expression

$$\frac{\mathrm{d}\mathcal{C}_V}{\mathrm{d}\tau} = \frac{1}{G_N L(d-1)} \left[r_{\mathrm{throat}} \int_{\partial\Sigma_{\mathrm{out}}(r_{\mathrm{throat}})} \sqrt{2\theta_k \theta_\ell} + 8\pi G_N \int_{\Sigma_{\mathrm{out}}(r_{\mathrm{throat}})} r T_{\alpha\beta} n^\alpha r^\beta\right]. \tag{8.129}$$

8.5.5 A simple purification

A spherically symmetric simplification

Consider a spherically symmetric bulk causal diamond D anchored at a bulk subregion R that is just a boundary sphere. D could be an entanglement wedge (intersected with the Wheeler-de-Witt patch of R – i.e. the domain of dependence is taken in the strict bulk sense), but need not be. Let Σ be the maximal volume slice of D. Then the edge of D, which equals $\partial\Sigma$, has the intrinsic metric of a sphere. We now want to find a surface Σ_0 and initial data (Σ_0, h, K_{ab}) such that $\Sigma \cup \Sigma_0$ (1) is complete, (2) is an extremal hypersurface, (3) satisfies our energy conditions and (4) has minimal volume consistent with (1)–(3). Let us only consider Σ_0 to be spherically symmetric – in principle there might be a Σ_0 with even less volume satisfying our above conditions which breaks spherical symmetry.

If it is compatible with energy conditions to have Σ_0 be compact, i.e. without a conformal boundary, then this is always the correct choice, since adding a second conformal boundary

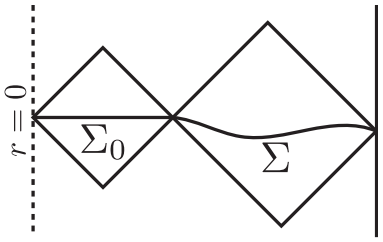


Figure 8.7: Construction of the simplified spacetime. The slice Σ_0 has intrinsic and extrinsic geometry of a static slice of AdS.

introduces new UV divergences. Let us thus assume Σ_0 is one-sided and later return to when this is consistent. Our computations in Sec. 8.2.4 showing that pure AdS was the least complex space did not rely on $r = \infty$ being the upper bound of the volume integral. Let Γ be some compact spherically symmetric extremal spacelike manifold with a spherical boundary of radius r. Then

$$\begin{aligned} V[\Gamma] &= \text{vol}[S^{d-1}] \int_0^r \frac{\mathrm{d}\rho \rho^{d-1}}{\sqrt{1+\rho^2/L^2 - \frac{\omega(\rho)}{\rho^{d-2}}}} \\ &\geq \text{vol}[S^{d-1}] \int_0^r \frac{\mathrm{d}\rho \rho^{d-1}}{\sqrt{1+\rho^2/L^2}} \\ &= V[\Sigma_0] = \frac{\text{vol}[S^{d-1}] r^d}{d} {}_2F_1\left(\frac{1}{2}, \frac{d}{2}, \frac{2+d}{2}, -\frac{r^2}{L^2}\right), \end{aligned} \tag{8.130}$$

where Σ_0 is the ball in the hyperbolic plane with boundary of area radius r. Here we used that $\omega(r) \geq 0$, which is shown in Lemma 10. Thus, of all spherically symmetric manifolds with boundary of area radius r, a ball in hyperbolic space has the least volume – assuming the WCC. Thus, we should just choose the completion Σ_0 of Σ to be a ball in hyperbolic space, provided that the energy shell induced at the junction between Σ and Σ_0 is compatible with the WCC and the NEC, which we return to in a moment. See Fig. 8.7 for an illustration of the new spacetime.

Now let us prove that any other spacetime with spherical symmetry has higher complexity: Let $(\tilde{M}, \tilde{g})$ be any spherically symmetric spacetime containing D, and let $\tilde{R}$ be a boundary Cauchy slice containing R. If R is a strict subset of $\tilde{R}$, then the maximal volume slice anchored at $\tilde{R}$ trivially has larger volume due to additional UV-divergences. Thus, assume $R = \tilde{R}$, meaning that $(\tilde{M}, \tilde{g})$ is one-sided. Let Γ be an extremal Cauchy slice of the inner wedge $I_W[\partial\Sigma]$ of $\partial\Sigma$ in $(\tilde{M}, \tilde{g})$ – see Fig. 8.8. Let $\tilde{\Sigma}$ be the maximal volume slice of $(\tilde{M}, \tilde{g})$ anchored at R. Since $\Sigma \cup \Gamma$ is a complete but non-extremal slice due to the kink at the joining, we have that its volume is less than the volume of $\tilde{\Sigma}$:

$$V[\tilde{\Sigma}] \geq V[\Sigma] + V[\Gamma]. \tag{8.131}$$

But Γ is a spherically symmetric compact manifold with boundary being a sphere, so it has greater or equal volume than the ball Σ_0 of the hyperbolic plane with the same area of its boundary. Thus

$$V[\tilde{\Sigma}] \geq V[\Sigma] + V[\Gamma] \geq V[\Sigma] + V[\Sigma_0], \tag{8.132}$$

completing our proof.

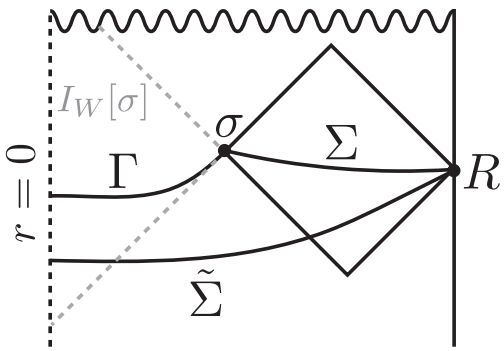


Figure 8.8: A spherically symmetric spacetime $(\tilde{M}, \tilde{g})$ containing $D[\Sigma]$. The maximal volume slice $\tilde{\Sigma}$ must have volume greater or equal to (8.81).

Energy conditions at the junction

By trying to glue Σ to a subset of the hyperbolic plane Σ_0, the thing that can go wrong is that we might be forced to violate the WCC or the NEC, since a junction supports a distributional Ricci tensor that might or might not respect our energy conditions. Below we show that the WCC (and thus the NEC) is preserved by the energy shell at the junction if $K_{\alpha\beta} = 0$ on Σ_0 and if

$$\sqrt{2}\theta_k[\partial\Sigma] \leq H_{\Sigma_0}[\partial\Sigma_0], \qquad \sqrt{2}\theta_\ell[\partial\Sigma] \geq -H_{\Sigma_0}[\partial\Sigma_0], \tag{8.133}$$

where k, ℓ are the future directed normals to $\partial\Sigma$ with k pointing towards R, and with

$$k \cdot n = \ell \cdot n = -\frac{1}{\sqrt{2}}, \tag{8.134}$$

where n^a is the future unit normal to Σ. Here $H_{\Sigma_0}[\partial\Sigma_0]$ is the mean curvature in Σ_0 with respect to the normal of $\partial\Sigma_0$ pointing toward Σ. For the case of spherical symmetry the conditions read

$$\theta_k[\partial\Sigma] \leq \frac{d-1}{\sqrt{2}}\sqrt{\frac{1}{r^2}+\frac{1}{L^2}}, \qquad \theta_\ell[\partial\Sigma] \geq -\frac{d-1}{\sqrt{2}}\sqrt{\frac{1}{r^2}+\frac{1}{L^2}}, \tag{8.135}$$

These are clearly satisfied for an HRT surface, since $\theta_k[\partial\Sigma] = \theta_\ell[\partial\Sigma] = 0$.

Let us now derive these conditions. Consider a surface σ contained in an extremal hypersurface Σ. Consider the null vectors

$$k^a = \frac{1}{\sqrt{2}}(n^a + r^a), \quad \ell^a = \frac{1}{\sqrt{2}}(n^a - r^a), \tag{8.136}$$

where we n^a is the future timelike unit normal to Σ and r^a the spacelike outwards normal to σ contained in Σ. Letting P^{ab} be the projector Σ and h^{ab} the projector on σ, we find

$$\begin{aligned}\sqrt{2}\theta_k &= h^{ab}\nabla_a(n_b + r_b) = (P^{ab} - r^a r^b)(\nabla_a n_b + \nabla_a r_b) \\ &= K + P^{ab}\nabla_a r_b - r^a r^b \nabla_a n_b \\ &= K + H_\Sigma[\sigma] - r^\alpha r^\beta K_{\alpha\beta}.\end{aligned} \tag{8.137}$$

A similar calculation shows that

$$\sqrt{2}\theta_\ell = K - H_\Sigma[\sigma] - r^\alpha r^\beta K_{\alpha\beta}. \tag{8.138}$$

Thus,

$$H_\Sigma[\sigma] = \frac{1}{\sqrt{2}}\left[\theta_{k_\Sigma} - \theta_{\ell_\Sigma}\right]. \tag{8.139}$$

If $K_{\alpha\beta} = 0$, which is the case for a constant-t slice of AdS, then

$$\theta_k = -\theta_\ell, \tag{8.140}$$

and

$$H_\Sigma[\sigma] = \sqrt{2}\theta_k = -\sqrt{2}\theta_\ell. \tag{8.141}$$

Let us now consider the junction conditions. In [141], it is shown that for a null junction between two spacetime regions M^+ and M^- generated by k^a, we have

$$R_{\ell\ell}|_{\text{singular}} = \frac{\delta(\tau)}{-k\cdot n}\left(\theta_\ell^- - \theta_\ell^+\right), \tag{8.142}$$

where n^a is tangent to a time-like congruence crossing the junction and τ the proper time from the junction along the congruence generated by n^a. For this formula to be valid we must have

$$[n\cdot k] \equiv n\cdot k|_+ - n\cdot k|_- = 0. \tag{8.143}$$

The region M^+ is the region to the future of the junction and M^- to the past. Taking n^a to be future directed, and taking k^a and to be future-directed and to point towards R, we have that the M^--region contains Σ. Thus the NEC demands

$$\theta_\ell[\partial\Sigma] - \theta_\ell[\partial\Sigma_0] \geq 0. \tag{8.144}$$

Similarly, for a null-junction generated by ℓ we find

$$R_{kk}|_{\text{sing}} = \frac{\delta(\tau)}{-\ell\cdot n}\left(\theta_k^- - \theta_k^+\right), \tag{8.145}$$

where $[\ell\cdot n] = 0$. This time however, the M^--region contains Σ_0, and so the NEC demands

$$\theta_k[\partial\Sigma_0] - \theta_k[\partial\Sigma] \geq 0. \tag{8.146}$$

Finally let us now consider what (non-null) energy shock is required. We do this by directly applying the spacelike junction conditions in the Riemannian setting, forgetting about the ambient spacetime. Define

$$[A] = A|_{\text{out}} - A|_{\text{in}}. \tag{8.147}$$

The Riemannian Einstein tensor is given by [141]

$$G_{\alpha\beta} = \text{non-singular} + \delta(s)S_{\alpha\beta}, \tag{8.148}$$

where s is the proper distance from σ and

$$S_{\alpha\beta} = -[H_{\alpha\beta}] + [H]h_{\alpha\beta}, \tag{8.149}$$

where we remind that $h_{\alpha\beta}$ is the intrinsic metric on Σ and $H_{\alpha\beta}$ the extrinsic curvature of σ in Σ with respect the outwards normal r^α (pointing towards R). This computation is purely geometric and does not assume Einstein's equations. Taking the trace, we thus find

$$R(h)\left(1 - \frac{d}{2}\right) = \text{non-singular} + \delta(s)(d-2)[H]. \tag{8.150}$$

Intergrating R on a small curve tangent r^a accross σ, we find

$$\int_{\delta}^{-\delta} \mathrm{d}s R(h) = -2[H] + \mathcal{O}(\delta). \tag{8.151}$$

Applying the Gauss-Codazzi equation, this gives

$$\int_{-\delta}^{\delta} \mathrm{d}\ell(2\mathcal{E} + K^{ab}K_{ab}) = -2[H] + \mathcal{O}(\delta). \tag{8.152}$$

Thus, for an arbitrarily small $\delta > 0$ the WCC dictates that

$$\int_{-\delta}^{\delta} \mathrm{d}\ell\mathcal{E} = H|_{\text{in}} - H|_{\text{out}} - \int \frac{1}{2}K^{ab}K_{ab} \geq 0. \tag{8.153}$$

It appears likely that $K_{ab}K^{ab}$ should not make any contribution to this integra in general, since $D_\alpha K^{\alpha\beta} = J^\beta$ should be solvable with a Green's function that smears the potential delta function in J, causing $K_{\alpha\beta}$ to be merely discontinous rather than distributional, and so

$$H|_{\text{in}} \geq H_{\text{out}}. \tag{8.154}$$

should be a sufficient condition for a positive energy shell. For spherical, planar and hyperbolic symmetry we can explicitly check that this indeed is exactly what happens. By formula (8.139) we see that the constraint on H reduces to

$$\theta_k[\partial\Sigma_0] - \theta_\ell[\partial\Sigma_0] \geq \theta_k[\partial\Sigma] - \theta_\ell[\partial\Sigma], \tag{8.155}$$

which is true when conditions on the NEC holds.

References

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

[2] V. E. Hubeny, M. Rangamani, and T. Takayanagi, *A Covariant holographic entanglement entropy proposal*, *JHEP* **0707** (2007) 062, [arXiv:0705.0016].

[3] A. C. Wall, *Maximin Surfaces, and the Strong Subadditivity of the Covariant Holographic Entanglement Entropy*, *Class.Quant.Grav.* **31** (2014), no. 22 225007, [arXiv:1211.3494].

[4] T. Faulkner, A. Lewkowycz, and J. Maldacena, *Quantum corrections to holographic entanglement entropy*, *JHEP* **1311** (2013) 074, [arXiv:1307.2892].

[5] X. Dong, *Holographic Entanglement Entropy for General Higher Derivative Gravity*, *JHEP* **01** (2014) 044, [arXiv:1310.5713].

[6] N. Engelhardt and A. C. Wall, *Quantum Extremal Surfaces: Holographic Entanglement Entropy beyond the Classical Regime*, *JHEP* **01** (2015) 073, [arXiv:1408.3203].

[7] A. R. Brown, H. Gharibyan, G. Penington, and L. Susskind, *The Python's Lunch: geometric obstructions to decoding Hawking radiation*, *JHEP* **08** (2020) 121, [arXiv:1912.00228].

[8] N. Engelhardt, G. Penington, and A. Shahbazi-Moghaddam, *A World without Pythons would be so Simple*, arXiv:2102.07774.

[9] N. Engelhardt, G. Penington, and A. Shahbazi-Moghaddam, *Finding Pythons in Unexpected Places*, arXiv:2105.09316.

[10] D. Stanford and L. Susskind, *Complexity and Shock Wave Geometries*, *Phys. Rev. D* **90** (2014), no. 12 126007, [arXiv:1406.2678].

[11] D. A. Roberts, D. Stanford, and L. Susskind, *Localized shocks*, *JHEP* **03** (2015) 051, [arXiv:1409.8180].

[12] L. Susskind and Y. Zhao, *Switchbacks and the Bridge to Nowhere*, arXiv:1408.2823.

[13] M. Alishahiha, *Holographic Complexity*, *Phys. Rev. D* **92** (2015), no. 12 126009, [arXiv:1509.06614].

[14] A. R. Brown, D. A. Roberts, L. Susskind, B. Swingle, and Y. Zhao, *Holographic Complexity Equals Bulk Action?*, *Phys. Rev. Lett.* **116** (2016), no. 19 191301, [arXiv:1509.07876].

[15] A. R. Brown, D. A. Roberts, L. Susskind, B. Swingle, and Y. Zhao, *Complexity, action, and black holes*, *Phys. Rev. D* **93** (2016), no. 8 086006, [arXiv:1512.04993].

[16] L. Lehner, R. C. Myers, E. Poisson, and R. D. Sorkin, *Gravitational action with null boundaries*, *Phys. Rev. D* **94** (2016), no. 8 084046, [arXiv:1609.00207].

[17] J. Couch, W. Fischler, and P. H. Nguyen, *Noether charge, black hole volume, and complexity*, *JHEP* **03** (2017) 119, [arXiv:1610.02038].

[18] S. Chapman, H. Marrochio, and R. C. Myers, *Complexity of Formation in Holography*, *JHEP* **01** (2017) 062, [arXiv:1610.08063].

[19] D. Carmi, R. C. Myers, and P. Rath, *Comments on Holographic Complexity*, *JHEP* **03** (2017) 118, [arXiv:1612.00433].

[20] D. Carmi, S. Chapman, H. Marrochio, R. C. Myers, and S. Sugishita, *On the Time Dependence of Holographic Complexity*, *JHEP* **11** (2017) 188, [arXiv:1709.10184].

[21] L. Susskind and Y. Zhao, *Complexity and Momentum*, arXiv:2006.03019.

[22] J. L. F. Barbón, J. Martín-García, and M. Sasieta, *Momentum/Complexity Duality and the Black Hole Interior*, *JHEP* **07** (2020) 169, [arXiv:1912.05996].

[23] J. L. F. Barbon, J. Martin-Garcia, and M. Sasieta, *Proof of a Momentum/Complexity Correspondence*, *Phys. Rev. D* **102** (2020), no. 10 101901, [arXiv:2006.06607].

[24] J. L. F. Barbon, J. Martin-Garcia, and M. Sasieta, *A Generalized Momentum/Complexity Correspondence*, *JHEP* **04** (2021) 250, [arXiv:2012.02603].

[25] L. Susskind, *Black Holes at Exp-time*, arXiv:2006.01280.

[26] A. Bouland, B. Fefferman, and U. Vazirani, *Computational pseudorandomness, the wormhole growth paradox, and constraints on the AdS/CFT duality*, arXiv:1910.14646.

[27] M. A. Nielsen, M. R. Dowling, M. Gu, and A. C. Doherty, *Quantum Computation as Geometry*, *Science* **311** (Feb., 2006) 1133–1135, [quant-ph/0603161].

[28] Z. Fu, A. Maloney, D. Marolf, H. Maxfield, and Z. Wang, *Holographic complexity is nonlocal*, *JHEP* **02** (2018) 072, [arXiv:1801.01137].

[29] M. Flory and N. Miekley, *Complexity change under conformal transformations in AdS_3/CFT_2*, *JHEP* **05** (2019) 003, [arXiv:1806.08376].

[30] A. Bernamonti, F. Galli, J. Hernandez, R. C. Myers, S.-M. Ruan, and J. Simón, *Aspects of The First Law of Complexity*, arXiv:2002.05779.

[31] R. Schoen and S. T. Yau, *On the proof of the positive mass conjecture in general relativity*, *Comm. Math. Phys.* **65** (1979), no. 1 45–76.

[32] R. Schoen and S. T. Yau, *Proof of the positive mass theorem. ii*, *Comm. Math. Phys.* **79** (1981), no. 2 231–260.

[33] E. Witten, *A Simple Proof of the Positive Energy Theorem*, *Commun. Math. Phys.* **80** (1981) 381.

[34] X. Wang, *The Mass of Asymptotically Hyperbolic Manifolds*, *Journal of Differential Geometry* **57** (2001), no. 2 273 – 299.

[35] P. T. Chrusciel and M. Herzlich, *The mass of asymptotically hyperbolic Riemannian manifolds*, *arXiv Mathematics e-prints* (Oct., 2001) math/0110035, [math/0110035].

[36] L. Andersson, M. Cai, and G. J. Galloway, *Rigidity and Positivity of Mass for Asymptotically Hyperbolic Manifolds*, *Annales Henri Poincare* **9** (2008) 1–33, [math/0703259].

[37] P. T. Chruściel and G. J. Galloway, *Positive mass theorems for asymptotically hyperbolic Riemannian manifolds with boundary*, arXiv:2107.05603.

[38] S. Brendle and O. Chodosh, *A volume comparison theorem for asymptotically hyperbolic manifolds*, *Communications in Mathematical Physics* **332** (May, 2014) [arXiv:1305.6628].

[39] O. Chodosh, *Large isoperimetric regions in asymptotically hyperbolic manifolds*, *Commun. Math. Phys.* **343** (2016), no. 2 393–443, [arXiv:1403.6108].

[40] S. Chapman, D. Ge, and G. Policastro, *Holographic Complexity for Defects Distinguishes Action from Volume*, *JHEP* **05** (2019) 049, [arXiv:1811.12549].

[41] N. Engelhardt and Å. Folkestad, "To appear.".

[42] T. Hertog, G. T. Horowitz, and K. Maeda, *Negative energy in string theory and cosmic censorship violation*, *Phys. Rev. D* **69** (2004) 105001, [hep-th/0310054].

[43] R. M. Schoen, *Variational theory for the total scalar curvature functional for riemannian metrics and related topics*, in *Topics in Calculus of Variations* (M. Giaquinta, ed.), (Berlin, Heidelberg), pp. 120–154, Springer Berlin Heidelberg, 1989.

[44] X. Hu, D. Ji, and Y. Shi, *Volume comparison of conformally compact manifolds with scalar curvature $r \geq -n\,(n-1)$*, *Annales Henri Poincaré* **17** (2014), no. 4 953–977, [arXiv:1309.5430].

[45] S. H. Shenker and D. Stanford, *Black holes and the butterfly effect*, *JHEP* **03** (2014) 067, [arXiv:1306.0622].

[46] G. Huisken and T. Ilmanen, *The Inverse Mean Curvature Flow and the Riemannian Penrose Inequality*, *Journal of Differential Geometry* **59** (2001), no. 3 353 – 437.

[47] H. L. Bray, *Proof of the Riemannian Penrose Conjecture Using the Positive Mass Theorem*, *arXiv Mathematics e-prints* (Nov., 1999) math/9911173, [math/9911173].

[48] H. L. Bray and D. A. Lee, *On the Riemannian Penrose inequality in dimensions less than 8*, *Duke Math. J.* **148** (2009) 81–106, [arXiv:0705.1128].

[49] V. Husain and S. Singh, *Penrose inequality in anti–de Sitter space*, *Phys. Rev. D* **96** (2017), no. 10 104055, [arXiv:1709.02395].

[50] M. Dahl, R. Gicquaud, and A. Sakovich, *Penrose type inequalities for asymptotically hyperbolic graphs*, *Annales Henri Poincare* **14** (2013) 1135–1168, [arXiv:1201.3321].

[51] L. Lopes de Lima and F. Girão, *Positive mass and Penrose type inequalities for asymptotically hyperbolic hypersurfaces*, *arXiv e-prints* (Jan., 2012) arXiv:1201.4991, [arXiv:1201.4991].

[52] Y. Ge, G. Wang, J. Wu, and C. Xia, *A penrose inequality for graphs over Kottler space*, *arXiv e-prints* (Sept., 2013) arXiv:1309.6248, [arXiv:1309.6248].

[53] D. A. Lee and A. Neves, *The Penrose inequality for asymptotically locally hyperbolic spaces with nonpositive mass*, *arXiv e-prints* (Oct., 2013) arXiv:1310.3002, [arXiv:1310.3002].

[54] M. Mars, *Present status of the Penrose inequality*, *Class. Quant. Grav.* **26** (2009) 193001, [arXiv:0906.5566].

[55] S. Lloyd, *Ultimate physical limits to computation*, *Nature* **406** (Aug, 2000) 1047–1054.

[56] S. Aaronson, *The Complexity of Quantum States and Transformations: From Quantum Money to Black Holes*, 7, 2016. arXiv:1607.05256.

[57] H. Hopf and W. Rinow, *Ueber den begriff der vollständigen differentialgeometrischen fläche*, *Commentarii Mathematici Helvetici* **3** (Dec., 1931) 209–225.

[58] C. Fefferman and C. R. Graham, *Conformal invariants*, *Elie Cartan et les Mathématiques d'aujourdhui* (Astérisque) p. 95. 1985.

[59] C. Graham and J. M. Lee, *Einstein metrics with prescribed conformal infinity on the ball*, *Advances in Mathematics* **87** (1991), no. 2 186–225.

[60] C. R. Graham and E. Witten, *Conformal anomaly of submanifold observables in AdS / CFT correspondence*, *Nucl. Phys. B* **546** (1999) 52–64, [hep-th/9901021].

[61] S. de Haro, S. N. Solodukhin, and K. Skenderis, *Holographic reconstruction of space-time and renormalization in the AdS / CFT correspondence*, *Commun. Math. Phys.* **217** (2001) 595–622, [hep-th/0002230].

[62] V. Balasubramanian and P. Kraus, *A Stress tensor for Anti-de Sitter gravity*, *Commun. Math. Phys.* **208** (1999) 413–428, [hep-th/9902121].

[63] R. Penrose, *Naked Singularities*, in *Sixth Texas Symposium on Relativistic Astrophysics* (D. J. Hegyi, ed.), vol. 224 of *Annals of the New York Academy of Sciences*, p. 125, 1973.

[64] L. Andersson and J. Metzger, *The Area of horizons and the trapped region*, *Commun. Math. Phys.* **290** (2009) 941–972, [arXiv:0708.4252].

[65] L. Andersson, M. Eichmair, and J. Metzger, *Jang's equation and its applications to marginally trapped surfaces*, in *4th International Conference on Complex Analysis and Dynamical Systems*, 6, 2010. arXiv:1006.4601.

[66] H. L. Bray, *The Penrose inequality in general relativity and volume comparison theorems involving scalar curvature*. PhD thesis, STANFORD UNIVERSITY, Nov., 1997.

[67] T. Hertog and G. T. Horowitz, *Towards a big crunch dual*, *JHEP* **07** (2004) 073, [hep-th/0406134].

[68] S. W. Hawking, *Gravitational radiation in an expanding universe*, *Journal of Mathematical Physics* **9** (1968), no. 4 598–604, [https://doi.org/10.1063/1.1664615].

[69] R. Geroch, *Energy extraction**, *Annals of the New York Academy of Sciences* **224** (1973), no. 1 108–117.

[70] P. Jang and R. Wald *J. Math. Phys.* **18** (1977) 41.

[71] P. T. Chrusciel and W. Simon, *Towards the classification of static vacuum space-times with negative cosmological constant*, *J. Math. Phys.* **42** (2001) 1779–1817, [gr-qc/0004032].

[72] H. Bray, S. Hayward, M. Mars, and W. Simon, *Generalized inverse mean curvature flows in spacetime*, *Commun. Math. Phys.* **272** (2007) 119–138, [gr-qc/0603014].

[73] D. Birmingham, *Topological black holes in Anti-de Sitter space*, *Class. Quant. Grav.* **16** (1999) 1197–1205, [hep-th/9808032].

[74] R. Emparan, *AdS / CFT duals of topological black holes and the entropy of zero energy states*, *JHEP* **06** (1999) 036, [hep-th/9906040].

[75] R.-G. Cai, S.-M. Ruan, S.-J. Wang, R.-Q. Yang, and R.-H. Peng, *Action growth for AdS black holes*, *JHEP* **09** (2016) 161, [arXiv:1606.08307].

[76] D. Momeni, M. Faizal, S. Bahamonde, and R. Myrzakulov, *Holographic complexity for time-dependent backgrounds*, *Phys. Lett. B* **762** (2016) 276–282, [arXiv:1610.01542].

[77] W.-J. Pan and Y.-C. Huang, *Holographic complexity and action growth in massive gravities*, *Phys. Rev. D* **95** (2017), no. 12 126013, [arXiv:1612.03627].

[78] R.-G. Cai, M. Sasaki, and S.-J. Wang, *Action growth of charged black holes with a single horizon*, *Phys. Rev. D* **95** (2017), no. 12 124002, [arXiv:1702.06766].

[79] P. Wang, H. Yang, and S. Ying, *Action growth in $f(R)$ gravity*, *Phys. Rev. D* **96** (2017), no. 4 046007, [arXiv:1703.10006].

[80] W.-D. Guo, S.-W. Wei, Y.-Y. Li, and Y.-X. Liu, *Complexity growth rates for AdS black holes in massive gravity and $f(R)$ gravity*, *Eur. Phys. J. C* **77** (2017), no. 12 904, [arXiv:1703.10468].

[81] W. Cottrell and M. Montero, *Complexity is simple!*, *JHEP* **02** (2018) 039, [arXiv:1710.01175].

[82] S. A. Hosseini Mansoori and M. M. Qaemmaqami, *Complexity growth, butterfly velocity and black hole thermodynamics*, *Annals Phys.* **419** (2020) 168244, [arXiv:1711.09749].

[83] B. Swingle and Y. Wang, *Holographic Complexity of Einstein-Maxwell-Dilaton Gravity*, *JHEP* **09** (2018) 106, [arXiv:1712.09826].

[84] M. Moosa, *Divergences in the rate of complexification*, *Phys. Rev. D* **97** (2018), no. 10 106016, [arXiv:1712.07137].

[85] Y.-S. An and R.-H. Peng, *Effect of the dilaton on holographic complexity growth*, *Phys. Rev. D* **97** (2018), no. 6 066022, [arXiv:1801.03638].

[86] M. Alishahiha, A. Faraji Astaneh, M. R. Mohammadi Mozaffar, and A. Mollabashi, *Complexity Growth with Lifshitz Scaling and Hyperscaling Violation*, *JHEP* **07** (2018) 042, [arXiv:1802.06740].

[87] P. A. Cano, R. A. Hennigar, and H. Marrochio, *Complexity Growth Rate in Lovelock Gravity*, *Phys. Rev. Lett.* **121** (2018), no. 12 121602, [arXiv:1803.02795].

[88] S. Chapman, H. Marrochio, and R. C. Myers, *Holographic complexity in Vaidya spacetimes. Part I*, *JHEP* **06** (2018) 046, [arXiv:1804.07410].

[89] S. Chapman, H. Marrochio, and R. C. Myers, *Holographic complexity in Vaidya spacetimes. Part II*, *JHEP* **06** (2018) 114, [arXiv:1805.07262].

[90] Y.-S. An, R.-G. Cai, and Y. Peng, *Time Dependence of Holographic Complexity in Gauss-Bonnet Gravity*, *Phys. Rev. D* **98** (2018), no. 10 106013, [arXiv:1805.07775].

[91] R. Auzzi, S. Baiguera, M. Grassi, G. Nardelli, and N. Zenoni, *Complexity and action for warped AdS black holes*, *JHEP* **09** (2018) 013, [arXiv:1806.06216].

[92] K. Nagasaki, *Complexity growth of rotating black holes with a probe string*, *Phys. Rev. D* **98** (2018), no. 12 126014, [arXiv:1807.01088].

[93] J. Couch, S. Eccles, T. Jacobson, and P. Nguyen, *Holographic Complexity and Volume*, *JHEP* **11** (2018) 044, [arXiv:1807.02186].

[94] M. Ghodrati, *Complexity growth rate during phase transitions*, *Phys. Rev. D* **98** (2018), no. 10 106011, [arXiv:1808.08164].

[95] S. Mahapatra and P. Roy, *On the time dependence of holographic complexity in a dynamical Einstein-dilaton model*, *JHEP* **11** (2018) 138, [arXiv:1808.09917].

[96] M. Reza Tanhayi, R. Vazirian, and S. Khoeini-Moghaddam, *Complexity Growth Following Multiple Shocks*, *Phys. Lett. B* **790** (2019) 49–57, [arXiv:1809.05044].

[97] J. Jiang, *Action growth rate for a higher curvature gravitational theory*, *Phys. Rev. D* **98** (2018), no. 8 086018, [arXiv:1810.00758].

[98] K. Meng, *Holographic complexity of Born–Infeld black holes*, *Eur. Phys. J. C* **79** (2019), no. 12 984, [arXiv:1810.02208].

[99] Z.-Y. Fan and M. Guo, *Holographic complexity under a global quantum quench*, *Nucl. Phys. B* **950** (2020) 114818, [arXiv:1811.01473].

[100] X.-H. Feng and H.-S. Liu, *Holographic Complexity Growth Rate in Horndeski Theory*, *Eur. Phys. J. C* **79** (2019), no. 1 40, [arXiv:1811.03303].

[101] J. Jiang, *Holographic complexity in charged Vaidya black hole*, *Eur. Phys. J. C* **79** (2019), no. 2 130, [arXiv:1811.07347].

[102] M. Alishahiha, K. Babaei Velni, and M. Reza Tanhayi, *Complexity and near extremal charged black branes*, *Annals Phys.* **425** (2021) 168398, [arXiv:1901.00689].

[103] D. Ageev, *Holographic complexity of local quench at finite temperature*, *Phys. Rev. D* **100** (2019), no. 12 126005, [arXiv:1902.03632].

[104] J. Jiang and M. Zhang, *Holographic complexity of the electromagnetic black hole*, *Eur. Phys. J. C* **80** (2020), no. 2 85, [arXiv:1905.07576].

[105] R. J. Caginalp, *Holographic Complexity in FRW Spacetimes*, *Phys. Rev. D* **101** (2020), no. 6 066027, [arXiv:1906.02227].

[106] Z.-Y. Fan and H.-Z. Liang, *Time dependence of complexity for Lovelock black holes*, *Phys. Rev. D* **100** (2019), no. 8 086016, [arXiv:1908.09310].

[107] Y.-S. An, R.-G. Cai, L. Li, and Y. Peng, *Holographic complexity growth in an FLRW universe*, *Phys. Rev. D* **101** (2020), no. 4 046006, [arXiv:1909.12172].

[108] S. Chapman and H. Z. Chen, *Charged Complexity and the Thermofield Double State*, *JHEP* **02** (2021) 187, [arXiv:1910.07508].

[109] K. Nagasaki, *Complexity growth for topological black holes by holographic method*, *Int. J. Mod. Phys. A* **35** (2020), no. 25 2050152, [arXiv:1912.03567].

[110] Y.-T. Zhou, X.-M. Kuang, Y.-Z. Li, and J.-P. Wu, *Holographic subregion complexity under a thermal quench in an Einstein-Maxwell-axion theory with momentum relaxation*, *Phys. Rev. D* **101** (2020), no. 10 106024, [arXiv:1912.03479].

[111] A. Bhattacharyya, S. Das, S. Shajidul Haque, and B. Underwood, *Cosmological Complexity*, *Phys. Rev. D* **101** (2020), no. 10 106020, [arXiv:2001.08664].

[112] W.-J. Pan, Y.-l. Li, M. Song, W.-b. Xie, and S. Zhang, *Holographic Complexity Growth Rate in a dual FLRW Universe*, arXiv:2003.11415.

[113] A.-C. Li, *Holographic complexity growth for a charged AdS-dilaton black holes with fixed and dynamical boundary respectively*, arXiv:2007.09520.

[114] H. Razaghian, *Complexity Growth of Dyonic Black holes with Quartic Field Strength Corrections*, arXiv:2009.03948.

[115] A. Al Balushi, R. A. Hennigar, H. K. Kunduri, and R. B. Mann, *Holographic complexity of rotating black holes*, arXiv:2010.11203.

[116] Y.-T. Zhou, X.-M. Kuang, and J.-P. Wu, *Complexity growth of massive black hole with a probe string*, arXiv:2104.12998.

[117] L. Susskind, *Computational Complexity and Black Hole Horizons*, *Fortsch. Phys.* **64** (2016) 24–43, [arXiv:1402.5674]. [Addendum: Fortsch.Phys. 64, 44–48 (2016)].

[118] J. Haferkamp, P. Faist, N. B. T. Kothakonda, J. Eisert, and N. Y. Halpern, *Linear growth of quantum circuit complexity*, arXiv:2106.05305.

[119] R.-Q. Yang, *Strong energy condition and complexity growth bound in holography*, *Phys. Rev. D* **95** (2017), no. 8 086017, [arXiv:1610.05090].

[120] M. Henneaux, C. Martinez, R. Troncoso, and J. Zanelli, *Black holes and asymptotics of 2+1 gravity coupled to a scalar field*, *Phys. Rev. D* **65** (2002) 104007, [hep-th/0201170].

[121] G. T. Horowitz, *Creating naked singularities and negative energy*, *Phys. Scripta T* **117** (2005) 86–91, [hep-th/0312123].

[122] M. Henneaux, C. Martinez, R. Troncoso, and J. Zanelli, *Asymptotically anti-de Sitter spacetimes and scalar fields with a logarithmic branch*, *Phys. Rev. D* **70** (2004) 044034, [hep-th/0404236].

[123] T. Hertog and K. Maeda, *Black holes with scalar hair and asymptotics in $N = 8$ supergravity*, *JHEP* **07** (2004) 051, [hep-th/0404261].

[124] M. Henneaux, C. Martinez, R. Troncoso, and J. Zanelli, *Asymptotic behavior and Hamiltonian analysis of anti-de Sitter gravity coupled to scalar fields*, *Annals Phys.* **322** (2007) 824–848, [hep-th/0603185].

[125] S. Fischetti and T. Wiseman, *A Bound on Holographic Entanglement Entropy from Inverse Mean Curvature Flow*, *Class. Quant. Grav.* **34** (2017), no. 12 125005, [arXiv:1612.04373].

[126] N. Bao, C. Cao, S. Fischetti, and C. Keeler, *Towards Bulk Metric Reconstruction from Extremal Area Variations*, *Class. Quant. Grav.* **36** (2019), no. 18 185002, [arXiv:1904.04834].

[127] L. V. Iliesiu, M. Mezei, and G. Sárosi, *The volume of the black hole interior at late times*, arXiv:2107.06286.

[128] M. Headrick and T. Takayanagi, *A Holographic proof of the strong subadditivity of entanglement entropy*, *Phys.Rev.* **D76** (2007) 106013, [arXiv:0704.3719].

[129] R. Penrose, *Gravitational collapse and space-time singularities*, *Phys. Rev. Lett.* **14** (1965) 57–59.

[130] S. W. Hawking and R. Penrose, *The Singularities of gravitational collapse and cosmology*, *Proc. Roy. Soc. Lond. A* **314** (1970) 529–548.

[131] S. W. Hawking and G. F. R. Ellis, *The large scale stucture of space-time*. Cambridge University Press, Cambridge, England, 1973.

[132] K. Balasubramanian and J. McGreevy, *Gravity duals for non-relativistic CFTs*, *Phys. Rev. Lett.* **101** (2008) 061601, [arXiv:0804.4053].

[133] S. Kachru, X. Liu, and M. Mulligan, *Gravity duals of Lifshitz-like fixed points*, *Phys. Rev. D* **78** (2008) 106005, [arXiv:0808.1725].

[134] M. Taylor, *Non-relativistic holography*, arXiv:0812.0530.

[135] D. T. Son, *Toward an AdS/cold atoms correspondence: A Geometric realization of the Schrodinger symmetry*, *Phys. Rev. D* **78** (2008) 046003, [arXiv:0804.3972].

[136] X. Dong, S. Harrison, S. Kachru, G. Torroba, and H. Wang, *Aspects of holography for theories with hyperscaling violation*, *JHEP* **06** (2012) 041, [arXiv:1201.1905].

[137] C. A. Agón, M. Headrick, and B. Swingle, *Subsystem Complexity and Holography*, *JHEP* **02** (2019) 145, [arXiv:1804.01561].

[138] N. Cheng, *Optimized Correlation Measures in Holography*, *Phys. Rev. D* **101** (2020), no. 6 066009, [arXiv:1909.09334].

[139] E. Cáceres, J. Couch, S. Eccles, and W. Fischler, *Holographic Purification Complexity*, *Phys. Rev. D* **99** (2019), no. 8 086016, [arXiv:1811.10650].

[140] S.-T. Yau, *Isoperimetric constants and the first eigenvalue of a compact riemannian manifold*, *Scientific annals of the 'Ecole Normale Sup érieure* **Ser. 4, 8** (1975), no. 4 487–507.

[141] E. Poisson, *A Relativist's Toolkit: The Mathematics of Black-Hole Mechanics*. Cambridge University Press, 12, 2009.

Chapter 9

Negative Complexity of Formation: the Compact Dimensions Strike Back

Abstract: We show that the vacuum-subtracted maximal volume, the proposed holographic dual to complexity of formation, can be negative when contributions from compact directions are included. We construct explicit solutions with arbitrarily negative complexity of formation in asymptotically $\mathrm{AdS}_4 \times S^7$ SUGRA. These examples rely critically on the compact directions, specifically the fact that the full eleven-dimensional spacetime is not asymptotically AdS_{11}. While there is some ambiguity in the extension of the holographic complexity proposal to the compact directions, we show that the two natural candidates can both have arbitrarily negative complexity of formation in SUGRA solutions. We further find examples in which complexity can even *decrease* at late times, including cases of both single-sided geometries and two-sided wormholes. In particular, we construct a cosmological wormhole with simultaneously negative and decreasing complexity of formation (as computed by volume) at late times. We find a distinguished role for relevant primaries in these constructions and comment on possible interpretations.

9.1 Introduction

Recent progress on the emergence of spacetime has crucially relied on the geometrization of quantum information theoretic quantities [1–16]. A relative newcomer to this set of connections has been the geometrization of computational complexity [11–14], either through the proposed Complexity=Action [13, 14] or Complexity=Volume duality [12]. The latter, which is our focus in this article, relates the circuit complexity $\mathcal{C}$ of a given holographic CFT$_d$ state $|\psi(\tau)\rangle$ relative to some reference state $|R\rangle$ to regulated bulk spatial volumes:

$$\mathcal{C}\big(\,|\psi(\tau)\rangle\,,|R\rangle\,\big) = \max_{\Sigma}\frac{\text{vol}[\Sigma]}{G_N L} \equiv \mathcal{C}_V\left(|\psi(\tau)\rangle\right), \tag{9.1}$$

where Σ a bulk hypersurface that intersects the conformal boundary on the timeslice τ; L is a length scale which we will take to be the AdS radius, and $\mathcal{C}_V$ a convenient shorthand for the gravitational quantity.

In a recent paper [17], we proved that the *complexity of formation* $\mathcal{C}_F$ [14, 18] satisfies

$$\mathcal{C}_F(|\psi\rangle) \equiv \mathcal{C}_V\left(|\psi\rangle\right) - \mathcal{C}_V\left(|0\rangle\right) \geq 0, \tag{9.2}$$

with equality if and only if $|\psi\rangle = |0\rangle$, where $|0\rangle$ is the vacuum dual to pure AdS$_{d+1}$. We established this for asymptotically AdS$_{d+1}$ spacetimes under the assumption of the weak curvature condition (WCC):

$$t^a t^b \left(R_{ab} - \frac{1}{2}g_{ab}R - \frac{d(d-1)}{L^2}g_{ab}\right) \geq 0, \qquad \forall \text{ timelike } t^a, \tag{9.3}$$

which in Einstein gravity reduces to the weak energy condition (WEC): $T_{ab}t^a t^b \geq 0$ for all timelike t^a. The result (9.2) implies that among the states with a classical asymptotically AdS$_{d+1}$ dual respecting the WCC, the vacuum is the least $\mathcal{C}_V$–complex. That is, WCC-respecting excitations of the vacuum move away from the reference state as measured by the complexity. The necessity of the WCC is clear from [19, 20], who found examples in which the vacuum-subtracted volume is negative; in those examples the WCC is violated.

The assumption of the WCC is somewhat unnatural from a holographic perspective: consistency conditions in the large-N, large-λ limit of the AdS/CFT correspondence are typically proven using the Null Curvature Condition (NCC) $R_{ab}k^a k^b \geq 0$ for null vectors k^a. The latter (strictly weaker) condition is expected to be true for any valid classical matter; the same, however, is not true for the WCC [21]. Nevertheless, it turns out that the WCC holds in type II and eleven-dimensional SUGRA (see appendix 9.4.1): even though the dimensional reduction of an asymptotically AdS$_{d+1}\times K$ over the compact dimensions K may violate the WCC while satisfying the NCC, inclusion of the compact dimensions restores the WCC. Prima facie, then, it may be tempting to conclude that when working in full ten or eleven dimensional SUGRA, our results immediately imply that complexity of formation is *always* positive. It would then be natural to conclude that $\mathcal{C}(|\psi\rangle\,,|R\rangle)$ should be identified with $\mathcal{C}_F(|\psi\rangle)$, with a reference state $|R\rangle$ that is identically the vacuum $|0\rangle$.

This naive conclusion, however, suffers from several flaws. First, the Complexity=Volume proposal does not admit an obvious generalization allowing the inclusion of compact directions.

There are (at least) two natural candidates: (1) the volume of the maximal volume slice Σ_{full} in the full $\text{AdS}_{d+1} \times K$ spacetime; or (2) the volume of the maximal volume slice Σ_{reduced} extended in the non-compact directions. It is simple (see Sec. 9.2) to show that in general the dimensional reduction of Σ_{full} does not result in Σ_{reduced}, and that consequently

$$\text{vol}[\Sigma_{\text{full}}] \neq \text{vol}[\Sigma_{\text{reduced}} \times K].$$

A maximal volume slice in the full spacetime need not be maximal in the AdS directions, and vice versa.

Our goal here, however, will not be to argue in favor of either (1) or (2) but to demonstrate that *neither* candidate can avoid a negative complexity of formation: the WCC restoration that accompanies the inclusion of compact directions fails to save either candidate from predicting that certain valid spacetimes in AdS/CFT are simpler than the vacuum[1]. As a consequence, since $\mathcal{C}_F$ is not positive semi-definite, it cannot be reinterpreted as the complexity with the vacuum as the reference state. To simplify matters, we will demonstrate this in a special case where the two candidate proposals coincide: a moment of time symmetry. In such a case, Σ_{full} reduces in the AdS directions to Σ_{reduced}, so that any conclusions are free of ambiguities relating to a choice between (1) and (2).

This result may at first appear to contradict our proof in [17]. How can the vacuum-subtracted maximal volume be negative in spacetimes respecting the WCC? The answer is a prime realization of the principle of conservation of misery: while the inclusion of compact directions restores the WCC, it in turn violates our assumption about AdS_{d+1} asymptotics. Thus we have two (mutually exclusive) options: accept violations of WCC in the absence of compact directions, or accept violations of AdS_{d+1} asymptotics. As it turns out, either option leads to states with negative complexity of formation.

In the following, we find a family of AdS_4 initial data supported by scalar tachyons above the Breitenlohner-Freedman bound [22], inherited from a truncation and dimensional reduction over the compact directions of eleven-dimensional SUGRA. In the full asymptotically $\text{AdS}_4 \times S^7$ data, the WCC is satisfied, but upon dimensional reduction the resulting spacetimes violate the WCC and satisfy the NCC. These geometries come in two flavors [2]: with and without boundary sources. For the former, the inclusion of boundary sources yields asymptotically AdS_4 spacetimes undergoing AdS false vacuum decay; such spacetimes, again supported by scalar tachyons, have a negatively divergent complexity of formation that decreases at late times. Among these spacetimes is a novel cosmological wormhole; even though the spacetime connects two asymptotic boundaries (with no dS region in between [25]), we find that the holographic volume complexity is nevertheless *smaller* than that of pure AdS. While spacetimes with negative and divergent $\mathcal{C}_F$ due to boundary sources were previously considered by [26], our examples without boundary sources are quite distinct, and should

[1] By 'valid' here we mean a stricter definition than is typically used (which is often just the requirement of the NEC and global hyperbolicity: here we mean that they are inherited from top down truncations of SUGRA.

[2] For computational facility, the examples with boundary sources are not constructed in an exact dimensional reduction of eleven-dimensional SUGRA, but instead with a slightly modified (but qualitatively similar) scalar potential, which enables analytical solutions. Analogous one-sided spacetimes were considered in $D = 11$ SUGRA reduced to $\text{AdS}_4 \times S^7$ in [23, 24], and we expect all of our qualitative findings in Sec. 9.3 to apply in $D = 11$ SUGRA.

be regarded as the main finding of this paper (although we also expand on examples with boundary sources, more analogous to the ones discussed in [26]). In this case we find initial data with arbitrarily negative (but finite) complexity of formation, both when viewed in eleven and four dimensions. The arbitrarily low complexity in this case is not caused by altering the boundary theory. Instead, it is obtained by a smooth deformation of the CFT state away from vacuum, and the low $\mathcal{C}_V$ is a genuine IR-effect caused by the compact dimensions.

There is however an underlying common denominator to all of our examples: they are constructed by turning on relevant scalar primaries in the CFT. This pattern together with the theorems of [17] suggests potential insights into the landscape of low complexity holographic states. If tachyonic bulk scalars, which are dual to relevant CFT scalar primaries, happen to be the only WEC-violating fields, then the only way to reduce the "distance" – as measured by $\mathcal{C}_V$ – to $|R\rangle$ below the fixed nonzero value $\mathcal{C}_V(|0\rangle)$ is to turn on VEVs for relevant scalars. Other operators will be dual to WEC-respecting fields, and so turning them on will cause $\mathcal{C}_V$ to increase with respect to the vacuum value. Thus, the presence of unstable directions of the IR fixed point correlates with the possibility of reducing complexity below the vacuum value. This could be due to the fact that the vacuum of a potential gapped phase at the end of the RG flow has significantly fewer correlations, simplifying the preparation of the state.

Our particular examples of low complexity spacetimes also provide potential insight into the reference state $|R\rangle$ implicit in the CV proposal. These examples are constructed by creating pockets of approximately constant scalar field at a moment of time-symmetry in the bulk, resulting in an effective AdS radius smaller than the asymptotic value L within the pocket. We find that $\mathcal{C}_F$ becomes arbitrarily negative as the pocket becomes larger: that is, the complexity becomes progressively closer to that of $|R\rangle$ via this reduction of the effective AdS radius in an increasingly large region. Since the limit of small AdS radius is not a well-defined classical geometry, this finding is consistent with the common perspective that $|R\rangle$ is a state without a geometric dual, e.g. a set of factorized qubits.

What is the upshot of our results for the Complexity=Volume proposal? At minimum, there is need for an unambiguous prescription that accounts for contributions from compact dimensions. It is clear that volumes can have a qualitatively different behavior when compact dimensions are included. On a more speculative level, there appears to be a sharp distinction (in the dimensionally-reduced picture) between operators whose dual is WCC-respecting and WCC-violating; acting on the vacuum with the former can only increase complexity; the latter, however, can decrease complexity. It would be interesting to understand this better in the dual CFT, perhaps using the proposed definitions of complexity in [27–30].

The paper is structured as follows. In Sec. 9.2 we present our SUGRA maximal volume asymptotically $\mathrm{AdS}_4 \times S^7$ initial data, together with its dimensional reduction. Then, to ensure that our constructed initial data gives the unique maximal volume slice in the evolved spacetime, we derive general properties of maximal volume slices in type II and $D = 11$ SUGRA in Sec. 9.4.2. Finally, in Sec. 9.3 we turn on boundary sources to study spacetimes undergoing AdS vacuum decay, both one-sided and two-sided. The appendix 9.4 provides technical details omitted in the main text.

9.2 Lower Unbounded Complexity of Formation

We begin by constructing asymptotically $\mathrm{AdS}_4 \times S^7$ examples with negative complexity of formation supported by a well-studied truncation of eleven-dimensional SUGRA compactified on the S^7 [31][3]

$$S = \frac{1}{8\pi G_N}\int_M \mathrm{d}^4x\sqrt{-g}\left[\frac{1}{2}R + \frac{3}{L^2} - \frac{1}{2}(\nabla\phi)^2 - V(\phi)\right], \tag{9.4}$$

with scalar potential

$$V(\phi) = \frac{1}{L^2}\left(1 - \cosh\sqrt{2}\phi\right). \tag{9.5}$$

Since $V(\phi)$ is unbounded below, this theory violates the WEC (and equivalently the WCC). However, it is simple to check that the tachyonic scalar mass about the $\phi = 0$ vacuum is above the BF bound. Furthermore, the null energy condition is satisfied, as always is the case for minimally coupled scalars.

A solution to the equations of motion of (9.4) with four-dimensional line element ds_4^2 lifts to a solution of eleven dimensional SUGRA with geometry

$$\mathrm{d}s^2 = \Delta^{2/3}\mathrm{d}s_4^2 + \frac{4L^2}{\Delta^{1/3}}\sum_{i=1}^{4} X_i^{-1}\left(\mathrm{d}\mu_i^2 + \mu_i^2\mathrm{d}\psi_i^2\right), \tag{9.6}$$

where $\mathrm{d}s_4^2$ is the four-dimensional metric and

$$\begin{aligned}
\mu_i &= (\sin\theta, \cos\theta\sin\varphi, \cos\theta\cos\varphi\sin\xi, \cos\theta\cos\varphi\cos\xi)\,,\\
X &\equiv X_1 = X_2 = e^{-\frac{\phi}{\sqrt{2}}},\\
X_3 &= X_4 = X^{-1},\\
\Delta &= \sum_{i=1}^{4} X_i\mu_i^2.
\end{aligned} \tag{9.7}$$

The angles $(\theta, \varphi, \psi_1, \psi_2, \psi_3, \psi_4)$ run over the range $[0, \pi]$, while $\xi \in [0, 2\pi)$. If we set $\phi = 0$ ($X_i = 1$), then the transverse space just becomes a round S^7 with radius $2L$: turning on the scalar ϕ squashes the S^7.

We now want to construct initial data on a spherically symmetric maximal volume slice Σ which has arbitrarily low complexity of formation in the $(d+1) = 4$ theory (9.4). Furthermore, upon success of this endeavor, we want to investigate whether considering the full volume in eleven-dimensional SUGRA restores positivity or boundedness from below. Let us first note that in the spherically symmetric case, if Σ has embedding coordinates $(t(r), r, \Omega_i)$ where Ω_i are the angles on the 2-sphere, then Σ can be lifted to a slice $\tilde{\Sigma}$ in the eleven-dimensional spacetime with embedding coordinates $(t(r), r, \Omega_i, \theta, \varphi, \xi, \psi_i)$. However, the slice $\tilde{\Sigma}$ is generally not a maximal volume slice (even though Σ is): turning on the scalar ϕ induces volume in the

[3]This theory was used to construct big crunch geometries in [23, 24]. Our boundary conditions will differ from [23, 24], so that the boundary dual will be different. The spacetimes in the next section will more closely resemble the situation in [23, 24].

compact dimensions, so if $\partial_t \phi \neq 0$ on Σ, then we can gain volume in the eleven-dimensional spacetime $(\tilde{M}, \tilde{g})$ by deforming $\tilde{\Sigma}$. However, letting K_{ab} denote the extrinsic curvature, if we take a moment of time symmetry, $K_{ab}[\Sigma] = \partial_t \phi = 0$, then we will also be at a moment of time-symmetry in eleven dimensions, and so $\tilde{\Sigma}$ is also extremal in $(\tilde{M}, \tilde{g})$.

In a moment we will construct explicit initial data, but let us first find an expression for the volume of $\tilde{\Sigma}$. We can take our coordinates on Σ so that

$$\begin{aligned} \mathrm{d}s^2|_{\Sigma} &= B(r)\mathrm{d}r^2 + r^2 \mathrm{d}\Omega^2, \\ \mathrm{d}s^2|_{\tilde{\Sigma}} &= \Delta^{2/3}\left(B(r)\mathrm{d}r^2 + r^2\mathrm{d}\Omega^2\right) + \frac{4L^2}{\Delta^{1/3}} \sum_{i=1}^{4} X_i^{-1}\left(\mathrm{d}\mu_i^2 + \mu_i^2 \mathrm{d}\psi_i^2\right), \end{aligned} \tag{9.8}$$

for some $B(r) > 0$. Integrating out the compact dimensions, we find an effective volume form $\tilde{\boldsymbol{\epsilon}}$ on Σ:

$$\tilde{\boldsymbol{\epsilon}} = (2L)^7 f\left(e^{-\frac{\phi}{\sqrt{2}}}\right) \boldsymbol{\epsilon}, \tag{9.9}$$

where $\boldsymbol{\epsilon}$ is the canonical volume form on Σ induced in the four-dimensional spacetime (M, g) and

$$f(X) = \pi^4 \frac{9\left(1 + X^{2/3}\right)\left(2 + 4X^{2/3} + 8X^{4/3} + 7X^2 + 8X^{8/3} + 4X^{10/3} + 2X^4\right)}{70X^{1/3}(1 + X^{2/3} + X^{4/3})^3}. \tag{9.10}$$

When $\tilde{\boldsymbol{\epsilon}}$ is integrated over Σ, it gives the volume of $\tilde{\Sigma}$. Since $f(X) \geq f(1) = \mathrm{vol}[S^7]$, including the compact dimensions always increases the volume compared to the naive multiplication of the $d = 3$ volume with $(2L)^7 \,\mathrm{vol}[S^7]$. For large $|\phi|$, to leading order

$$\tilde{\boldsymbol{\epsilon}} = \frac{27}{35}(2L)^7 \,\mathrm{vol}[S^7] e^{\frac{|\phi|}{3\sqrt{2}}} \boldsymbol{\epsilon}. \tag{9.11}$$

Thus, for large scalar condensates there is generally an *exponential* difference in $|\phi|$ between the naive $(2L)^7 \,\mathrm{vol}[S^7]\,\mathrm{vol}[\Sigma]$ and the true volume $\mathrm{vol}[\tilde{\Sigma}]$. This clearly demonstrates that compact directions can dramatically modify the volume even if the extremal slice is unchanged.

We now want to pick initial data leveraging the negativity of $V(\phi)$ to minimize the volume of Σ. The solution of the constraint equations for Einstein-Maxwell-Scalar theory on a spherically symmetric maximal volume slice at a moment of time symmetry and with $d = 3$ is [17]

$$\begin{aligned} B(r) &= \left(1 + r^2 - \frac{\omega(r)}{r}\right)^{-1}, \\ \omega(r) &= \frac{1}{2}\int_0^r \mathrm{d}\rho \rho^2 e^{\frac{1}{2}\int_r^\rho \mathrm{d}z \phi'(z)^2}\left[\left(1 + \rho^2\right)\phi'(\rho)^2 + 2 - 2\cosh\sqrt{2}\phi\right], \end{aligned} \tag{9.12}$$

where we pick units of $L = 1$ for brevity. The quantity $\omega(r)$ is a quasi-local mass function, and $\omega(\infty)$ is proportional to the conserved spacetime mass when $\omega(\infty)$ is finite [23].

We now pick the scalar profile on Σ to fall off in such a way so that the evolution of the initial data on Σ does not spoil the AdS asymptotics. Furthermore, we must ensure that $B(r) > 0$ everywhere so that Σ is everywhere spacelike. Beyond these two constraints, we

are free to choose the profile for ϕ.[4] The usual near boundary analysis of the Einstein-Klein-Gordon system constrains the asymptotic behavior of ϕ:

$$\phi(r,x) = r^{-\Delta_-}\left(\phi^{(0)}(x) + \phi^{(2)}(x) r^{-2} + \ldots\right) + r^{-\Delta_+}\left(\psi^{(0)}(x) + \psi^{(2)}(x) r^{-2} + \ldots\right), \tag{9.13}$$

where $\Delta_- = 1$ and $\Delta_+ = 2$. In order to avoid turning on boundary sources and to keep $\omega(\infty)$ finite (so the volume divergence structure agrees with that of pure AdS and so the Balasubramanian-Kraus stress tensor is defined [33]), we take $\phi^{(2n)} = 0$.

Let us construct one-sided initial data with no minimal surfaces, so that one coordinate patch covers the whole of Σ (this happens when $B(r)$ is nowhere divergent [17]). We choose the profile

$$\phi(r) = 1 - e^{-a^2/r^2}, \tag{9.14}$$

which has the requisite $\mathcal{O}(r^{-2})$ behavior needed to keep $\omega(\infty)$ finite. In the appendix we prove that in type II and $D = 11$ SUGRA, (1) any $K = 0$ slice is maximal, and (2) there can only be one maximal slice at a fixed anchoring. Thus, the spacetime obtained by evolving our initial data cannot possess another maximum volume slice with larger volume.

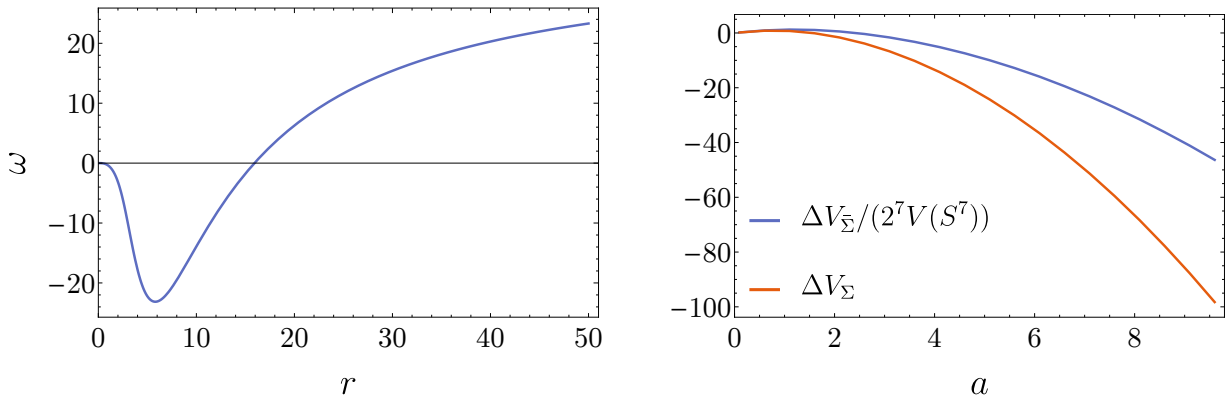
It can be checked that the profile (9.14) results in $0 < B(r) < \infty$, so that our assumption of no minimal surfaces is satisfied. We now proceed to calculate volumes of Σ and $\tilde{\Sigma}$ relative a constant$-t$ slice of AdS_4 and $\mathrm{AdS}_4 \times S^7$, respectively:

$$\begin{aligned}
\Delta V_\Sigma(a) &= \mathrm{vol}[S^2] \int_0^\infty \mathrm{d}r r^2 \left[\frac{1}{\sqrt{1 + r^2 - \frac{\omega(r)}{r}}} - \frac{1}{\sqrt{1+r^2}} \right], \\
\Delta V_{\tilde{\Sigma}}(a) &= \mathrm{vol}[S^2] 2^7 \int_0^\infty \mathrm{d}r r^2 \left[\frac{f\left(e^{-\frac{\phi}{\sqrt{2}}}\right)}{\sqrt{1 + r^2 - \frac{\omega(r)}{r}}} - \frac{\mathrm{vol}[S^7]}{\sqrt{1+r^2}} \right],
\end{aligned} \tag{9.15}$$

(Here the factor 2^7 appears since the round S^7 has radius 2 in units of $L = 1$.)

In Fig. 9.1 we plot the result: we show ΔV_Σ and $\Delta V_{\tilde{\Sigma}}/V(S^7)2^7$ plotted against a, together with the profile $\omega(r)$ for the value $a = 5$. Other values of a give a qualitatively similar shape for $\omega(r)$. We see that the vacuum subtracted volume for Σ becomes negative as we increase a. We find no signs of this decrease stopping for very large values of a. However, if we impose a finite cutoff at $r \sim \frac{1}{\epsilon}$, the decrease will saturate at $a \sim 1/\epsilon$. Either way, we see that turning on an increasingly large condensate of our tachyonic scalar field (or in the CFT language, turning on an increasingly large VEV for a relevant scalar primary) takes us closer to the reference state $|R\rangle$. This is a very distinct behavior which a CFT dual to volume ought to reproduce.

[4] We do not have a guarantee that it is possible to prepare (9.14) via the Euclidean path integral. However, for our conclusions to fail, it would have to be impossible to construct any qualitatively similar scalar condensate at a moment of time symmetry, since our findings do not depend on the particular quantitative details of the condensate (9.14). Any tachyonic scalar condensate should give the same conclusion as long as we have (1) a pocket where ϕ is non-zero and approximately constant, (2) the pocket is of size at least $r \sim L$, and (3) the scalar falls off as slowly as is consistent (9.13) and finite energy. It seems unlikely that such profiles cannot be prepared. In fact, [32] shows perturbatively that such profiles can be prepared using the Euclidean path integral.



(a) (b)

Figure 9.1: (a) $\omega(r)$ for $a = 5$, and (b) the vacuum-subtracted volume as function of a in units of L.

What is the bulk mechanism leveraged in these examples that allow negative $\mathcal{C}_F$? In the WEC-violating case ($d = 3$) it is the absence of a lower bound on the intrinsic Ricci scalar of maximal volume slices. It is illustrative to look at a conjecture of Schoen (which is proven for $d = 3$ [34–36]), which states that[5]

Conjecture 2 ([37]). *Let (Σ, h_0) be a closed hyperbolic Riemannian manifold with constant negative scalar curvature $R[h_0]$. Let h be another metric on Σ with scalar curvature $R[h] \geq R[h_0]$. Then* $\mathrm{vol}[\Sigma, h] \geq \mathrm{vol}[\Sigma, h_0]$.

While this conjecture pertains to compact rather than conformally compact manifolds, there are good reasons to believe it holds for conformally compact manifolds with AdS asymptotics, as discussed in [17]. Now, in our $d = 3$ data the inequality $R[h] \geq R[h_0]$ with h_0 being the metric of hyperbolic space of radius $L = 1$ no longer holds as a consequence of the tachyon condensate, so this is presumably what allows $\mathcal{C}_F < 0$. What about the eleven-dimensional case? Since the WCC allows us to put a lower bound on the intrinsic Ricci scalar of $\tilde{\Sigma}$, it might look like the conditions of Schoen's theorem hold. But this is not so: the comparison manifold must have a hyperbolic metric h_0, and static slices of $\mathrm{AdS}_4 \times S^7$ are not hyperbolic. Type IIA, IIB and eleven-dimensional SUGRA all satisfy (see Appendix 9.4.1)[6]

$$t^a t^b \left(R_{ab}[g] - \frac{1}{2} g_{ab} R[g] \right) \geq 0, \tag{9.16}$$

which through the Gauss-Codazzi equation implies that the intrinsic Ricci scalar of any extremal hypersurface in every solution of these theories is positive. Thus, hyperbolic volume comparison theorems no longer apply.[7] While we do not know of any volume comparison

[5]The conjecture is phrased in a different but equivalent way in [37]. The version stated here can be found in [38].

[6]Thus any AdS vacuum of these theories will satisfy the WCC (9.3) with the relevant AdS radius.

[7]Strictly speaking they could apply for some choice of h_0, but not when we pick h_0 to be a solution of the maximal-volume constraints in our theory, which is the relevant case for $\mathcal{C}_F$.

results for asymptotically $\mathrm{AdS}_{d+1} \times K$ type manifolds satisfying (9.16), it is instructive to note that volume comparison results for manifolds of spherical topology tend to imply *lower volume* when the Ricci scalar is higher [38].[8] This together with our example indicates that, with respect to volume, deformations that mainly affect the compact dimensions behave very different from those that mainly deform the non-compact dimensions.

Finally, let us inquire about the fate of our very low $\mathcal{C}_V$ data upon time-evolution. A guess, in keeping with earlier work on the same tachyonic scalar theory [23] and the spacetimes considered in the next section, would be that it collapses into a big crunch singularity. The negative and unbounded potential $V(\phi) \sim 1 - \cosh\sqrt{2}\phi$ seems to favor such a collapse. However, our data is different from that of [23] and the next section in a significant way: ours has only normalizable modes turned on, so that the ordinary definition of the energy is finite and positive. Furthermore, the source of the decay to a big crunch in [23, 24] was argued to be the presence of a lower unbounded triple trace term in the dual theory Hamiltonian caused by the non-normalizable mode, but here this term is not turned on due to the faster scalar field falloff. It seems likely that our data evolves to eventually form a black hole.

The skeptical reader may at this point refuse to take such unboundedly low $\mathcal{C}_F$ in single-sided spacetimes seriously, pointing out that the CV-proposal was in its original formulation intended to describe wormholes and the volume in the interior of a horizon. Possibly, such a reader may concede, there are some subtleties in one-sided geometries; but surely wormholes – the original motivation for CV – are still safe.

Any such perspective is however about to be disappointed: a small modification of the construction above allows us to build two-boundary geometries supported by tachyonic scalars with unboundedly small vacuum-subtracted volumes. To do so, we modify our construction above by using the following solution to the constraint equations

$$\omega(r) = \frac{1}{2} e^{-\frac{1}{2}\int_{r_0}^{r} \mathrm{d}\rho\rho\phi'(\rho)^2} \left\{ \omega_0 + \int_{r_0}^{r} \mathrm{d}\rho\rho^2 e^{\frac{1}{2}\int_{r_0}^{\rho} \mathrm{d}z\phi'(z)^2} \left[\left(1+\rho^2\right)\phi'(\rho)^2 + 2 - 2\cosh\sqrt{2}\phi \right] \right\}, \tag{9.17}$$

where $1 + r_0^2 - \frac{\omega_0}{r_0} = 0$. For any r_0 we can use a profile similar to (9.14) to make $\mathrm{vol}[\Sigma]$ arbitrarily negative compared to two copies of pure AdS. Thus the phenomenon of lower unbounded $\mathcal{C}_F$ is equally relevant for wormholes.

9.3 Negatively Divergent Complexity of Formation

The clear culprit for negative $\mathcal{C}_F$ in Sec. 9.2 was compact dimensions or tachyonic scalars. While the main point of interest in this article is the effect of including the compact dimensions, the importance of the scalar tachyons above clearly bears some further investigation. In this section we provide additional examples of tachyonic scalars causing unusual volume behavior. Previous work [26] has conducted a near-boundary analysis that found that turning on boundary sources for these tachyons (thus changing the asymptotic structure, in contrast with the previous section in which the asymptotics were unmodified) can result in initial data that has divergent $\dot{\mathcal{C}}_F$. Our results in this section support this conclusions and expand

[8] Note however that in this case, a bound on just the Ricci scalar is not sufficient for volume comparisons. Further bounds on R_{ab} must be satisfied [38].

it further by (1) providing a full spacetime evolution to clarify the physical picture and (2) constructing wormholes with the same properties.

The setting will be unstable asymptotically AdS spacetimes undergoing decay. We will leverage that there are analytical examples of such spacetimes, rather than just initial data. The price we pay is that (1) the theory under consideration does not come from a known realization of AdS/CFT, and (2) the scalar potential is only known numerically. We do not expect this price to be conceptually meaningful: spacetimes that are entirely analogous qualitatively can be constructed numerically directly from the SUGRA potential [23, 24]. Here we prefer to work with an analytically known metric, but we do not expect any of the qualitative features of our analysis to change by the modification of the potential. As emphasized above, we here deviate from the setup of the previous section: the scalar field falloff, which is sufficiently slow that boundary sources are turned on, resulting in a divergent Balasubramanian-Kraus [33] stress tensor. Defining a boundary stress tensor then requires additional counterterms involving the scalar field [23, 39–43]. As we will see, this in turn causes the divergence structure of extremal surface volumes to differ from pure AdS, leading to a UV-divergent $\mathcal{C}_F$ and $\frac{\mathrm{d}\mathcal{C}_V}{\mathrm{d}t}$.

9.3.1 AdS Vacuum Decay

The one-sided spacetimes we consider are given by the one-parameter family of metrics constructed in [44], parametrized by the real positive parameter c. These geometries are covered by two coordinate patches, with patch I having metric

$$\mathrm{d}s_{\mathrm{I}}^2 = \mathrm{d}\xi^2 + a(\xi)^2 \left(-\mathrm{d}\zeta^2 + \cosh^2\zeta \mathrm{d}\Omega^2\right), \tag{9.18}$$

where $\mathrm{d}\Omega^2$ now is the metric of a $d-1$–dimensional sphere, and with

$$a(\xi) = (1+c)\sinh\xi - 2c\sinh\frac{\xi}{2}. \tag{9.19}$$

The second patch, with coordinates (t, ρ, Ω_i), is obtained by the analytic continuation $\xi = it, \tau = i\rho$. Patch I is the causal wedge of the spacetime, with the AdS boundary at $\xi = \infty$, where the scalar field is in the false vacuum (a local maximum). As ξ approaches 0, where the edge of the causal wedge lies, the scalar field approaches the true vacuum value. Patch II is an FRW-region which initially expands with motion away from $t = 0$ and then crunches, with a curvature singularity at $t = \pm t_*$, where $a(it_*) = 0$. Physically, the $\zeta = 0$ slice is a bubble of true vacuum that nucleates inside the false vacuum at a moment of time symmetry. Forward or backward time-evolution results in the subsequent decay of the false AdS vacuum. See Fig. 9.2 for a conformal diagram drawn for the case $c = 1, d = 3$ (see Appendix 9.4.3 for the computation).

This spacetime has past and future cosmological singularities, and in the boundary conformal frame of the static cylinder, the boundary exists only for a finite time. The dual field theory, if it exists, can either be viewed as living on de Sitter space, or it can be seen as a field theory on the Einstein static universe whose evolution terminates in finite time – possibly due to a Hamiltonian that is unbounded below, as discussed in [23].

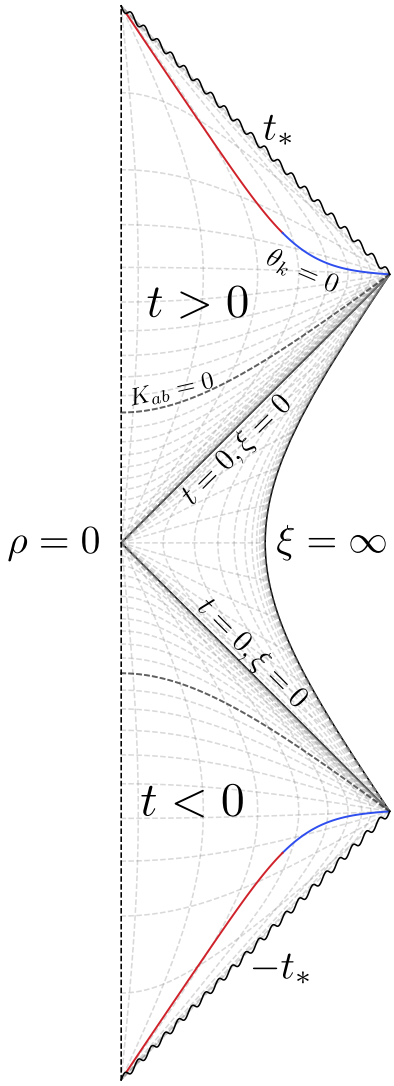


Figure 9.2: Conformal diagram of the spacetime (9.18) with $d = 3$ and $c = 1$. Dashed lines running vertically are hypersurfaces of constant ρ and ξ, with equidistant coordinate spacing. Dashed lines running horizontally are hypersurfaces of constant t and ζ, with equidistant coordinate spacing. The blue line is the spacelike section of a holographic screen, while the red line is the timelike portion. The darker horizontal dashed lines are the constant$-t$ surfaces at which the FRW region transitions between expanding and crunching ($a'(t) = 0$), which are totally geodesic. These are extremal surface barriers.

$\mathcal{C}_F$ is negative and divergent

Consider again the coordinates

$$\mathrm{d}s^2|_\Sigma = \frac{1}{1+r^2-\frac{\omega(r)}{r^{d-2}}}\mathrm{d}r + r^2\mathrm{d}\Omega^2, \tag{9.20}$$

on Σ. Assume now that $\omega(r)$ is divergent at large r, with the leading behavior at large$-r$ given by $\omega(r) \sim \omega_s r^s$ for $0 < s < d$.[9] The leading ω-dependent divergence in the volume then is given by:

$$\begin{aligned}\mathrm{vol}[\Sigma] &\sim \mathrm{vol}[S^{d-1}]\int^{r_{\rm cut}} \mathrm{d}r \frac{r^{d-1}}{\sqrt{1+r^2-\frac{\omega(r)}{r^{d-2}}}} \\ &\sim \mathrm{vol}[S^{d-1}]\int^{r_{\rm cut}} \mathrm{d}r r^{d-2}\left[\frac{1}{2r^{d-s}}\omega_s + \ldots\right] \\ &\sim \omega_s \frac{\mathrm{vol}[S^{d-1}]r_{\rm cut}^{s-1}}{2(s-1)},\end{aligned} \tag{9.21}$$

Comparing with a slice of pure AdS with cutoff at the same area-radius $r_{\rm cut}$, we find

$$\mathrm{vol}[\Sigma] - \mathrm{vol}[\Sigma_{\mathrm{AdS}_4}] = \omega_s \frac{\mathrm{vol}[S^{d-1}]r_{\rm cut}^{s-1}}{2(s-1)} + \text{subleading}. \tag{9.22}$$

We now proceed to show that for our spacetime, we have $s = d - \frac{1}{2}$ and $\omega_s < 0$, giving that $\mathcal{C}_F$ is negative and divergent.

To calculate ω_s, it is useful to know that there is a geometric functional $\omega[\sigma, \Sigma]$ that reduces to $\omega(r)$ when σ is a symmetric codimension-2 spatial surface and Σ spherically symmetric [17]:[10]

$$\omega[\sigma,\Sigma] = \frac{1}{\mathrm{vol}[S^{d-1}]}\left(\frac{A[\sigma]}{\mathrm{vol}[S^{d-1}]}\right)^{\frac{1}{d-1}}\int_\sigma \left[\frac{\mathcal{R}}{(d-1)(d-2)} - \frac{H^2}{(d-1)^2} + \frac{1}{L^2}\right], \tag{9.23}$$

where $H[\sigma]$ is the mean curvature of σ inside Σ and $\mathcal{R}$ the intrinsic Ricci scalar of σ. Let now Σ be the $\zeta = 0$ hypersurface, which is a maximal volume slice since it is a moment of time symmetry. Evaluating $\omega[\sigma, \Sigma]$ for a constant ξ surface σ, we find

$$\omega(\xi) = a(\xi)^{d-2}\left(1 + a(\xi)^2 - a'(\xi)^2\right) = -\frac{1}{2^{d-1}}c(1+c)^2 e^{\left(d-\frac{1}{2}\right)\xi} + \mathcal{O}(e^{(d-1)\xi}). \tag{9.24}$$

If we were to change coordinates to the form (9.20) we would find $r = \mathcal{O}(e^\xi)$, and so indeed we have $s = d - 1/2$, $\omega_s < 0$, showing that $\mathcal{C}_F$ is negative and divergent.

Note that pure AdS-subtracted volume here is somewhat unnatural from the field theory perspective. The scalar field falls off sufficiently slowly so as to turn on a source on the boundary: pure AdS is not a solution of the boundary theory dual to (9.18), so the comparison appears ill-motivated. In this particular setting – though not in the previous section – the negatively divergent $\mathcal{C}_F$ should be viewed as a statement purely about volumes in asymptotically AdS spacetimes, rather than as a statement pertaining to a single field theory. It is in principle possible that the field theory dual to (9.18), if it exists, has a preferred state for the volume subtraction, for which $\mathcal{C}_F$ would be positive.

[9] $s \geq d$ is incompatible with being asymptotically AdS with radius 1.

[10] This is proportional to the so-called Geroch-Hawking mass when $d = 3$.

Complexity change

Let us now compute the leading divergent contribution to the complexity change for spherically symmetric maximal volume slices. To compute the change in complexity we must choose the bulk cutoff carefully. Any given conformal frame induces a unique Fefferman-Graham coordinate system in a neighbourhood of the conformal boundary [45–47]:

$$\mathrm{d}s^2 = \frac{1}{z^2}[\mathrm{d}z^2 + \gamma_{\mu\nu}(z,x)\mathrm{d}x^\mu \mathrm{d}x^\nu], \tag{9.25}$$

where $z = 0$ is the conformal boundary and $\gamma_{\mu\nu}(0,x)$ the chosen conformal representative. Given this coordinate system, we can cut off volumes at $z = \epsilon$.

In the case at hand there are two natural conformal frames; we can either choose the boundary to be dS_d or the static cylinder. With respect to the dS_d conformal frame it can readily be checked that the leading divergence in the volume of the maximal volume slice anchored at a constant ζ is

$$\mathrm{vol}[\Sigma_\zeta] = \frac{\mathrm{vol}[S^{d-1}]\cosh\zeta^{d-1}}{(d-1)\epsilon^{d-1}} + \mathcal{O}\left(\epsilon^{d-3/2}\right). \tag{9.26}$$

This increases to the future and past of $\zeta = 0$ simply because (1) when regulating with a Fefferman-Graham cutoff, the divergence of maximal volume slices is always proportional to the boundary volume in the chosen conformal frame, and (2) the volume of constant ζ slices of de Sitter increases to the future and past.

Next, let us look at the more interesting case of the static cylinder conformal frame. A computation (see appendix 9.4.3) of the leading order complexity change with cutoff adapted to the static cylinder gives

$$\frac{\mathrm{d}\,\mathrm{vol}[\Sigma_t]}{\mathrm{d}t} = -\frac{1}{\epsilon^{d-\frac{3}{2}}}\frac{d-1}{d-\frac{3}{2}}\sqrt{\frac{c^2}{2(c+1)}}\frac{\sin t}{(\cos t)^{3/2}} + \mathcal{O}(\epsilon^{-d+2}), \qquad t \in \left(-\frac{\pi}{2},\frac{\pi}{2}\right). \tag{9.27}$$

This is clearly negative and divergent, and so unlike in well known examples of black holes, the moment of time-symmetry is here a maximum of $\mathcal{C}_V$, rather than a minimum. This is in constrast with the case of the dS conformal frame, and so highlights how extremal hypersurface volume is an observable whose UV-divergence structure depends strongly on the choice of boundary conformal frame. This is consistent with [26]'s near-boundary analysis, which also found a divergent rate of change for tachyonic scalars with boundary sources turned on. In our example we have the additional benefit of knowing the spacetime globally, providing a physical picture of what is happening in the bulk.

The decrease of $\mathcal{C}_V$ is a UV effect, and the volume behind the horizon is admittedly increasing towards the future (before the crunch region that is, which is anyway hidden from all extremal surfaces). However, the volume behind the horizon is not a natural observable to associate to a boundary state at a fixed time, since turning on sources in the future would alter the horizon location, and thus also the volume behind it. Nevertheless, it does seem that the CV proposal needs some modification in the spacetimes considered in this section. One possibility is some generalized volume functional that includes contributions from the scalar fields, which from a dimensional reduction perspective this appears natural.

9.3.2 Decaying Cosmological Wormholes

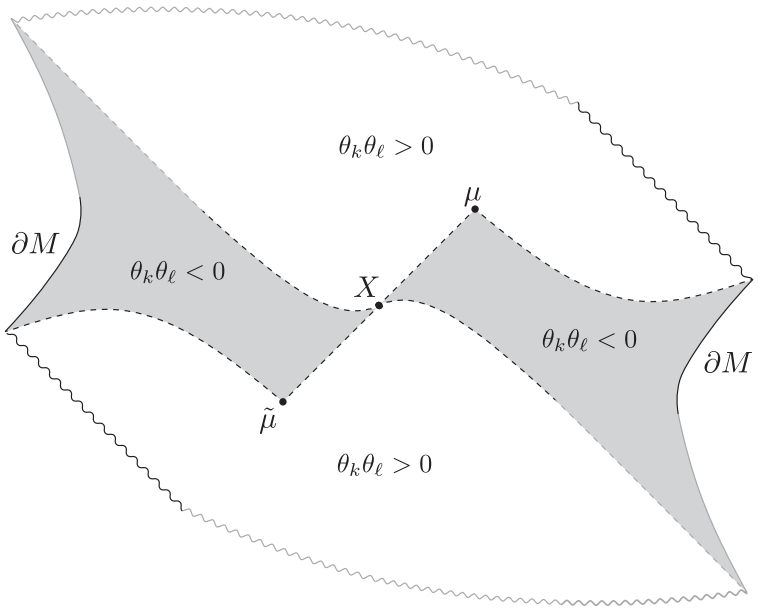


Figure 9.3: Black lines show numerically computed features of a coarse-grained spacetime formed from a marginally trapped (minimar [48]) surface μ on the spacelike section of the holographic screen shown Fig. 9.2. Dashed lines represent apparent horizons, while X is the HRT surface. Grey lines lie outside the range of our numerics, and are pure sketches representing a qualitative image of how the full spacetime could look like. For an illustration with additional details, see Fig. 9.4 in the appendix.

Can we find a wormhole with the same properties as the spacetime considered above – a wormhole with cosmological singularities, negative divergent complexity of formation, and decreasing late time complexity? That would appear to be in some tension with the paradigm of wormhole volume corresponding to increasing complexity; such an example would add urgency in finding an appropriate modification of CV that can accommodate such spacetimes.

It turns out that we can, in fact, build such a spacetime. The procedure is borrowed from [48, 49], which constructs a two-boundary wormhole from a single-sided spacetime containing a marginally trapped surface satisfying certain assumptions. The protocol is roughly as follows: one fixes the data in the exterior of a given marginally trapped surface μ; the rest of the initial data is provided on a stationary null hypersurface fired towards the past interior from μ. This hypersurface eventually develops an extremal surface on it; at this surface, the initial data is CPT conjugated, resulting in a second boundary. The initial data is characteristic; in our case, fixing a spherically symmetric marginally trapped surface means that obtaining the spacetime requires numerical evolution of the spherically symmetric characteristic Einstein-Klein-Gordon equations. We will not expound on the details of the numerics here (although they are surprisingly simple), instead just briefly summarizing our results. Figure 9.3 illustrates one of these wormholes. On the left, we show the marginally trapped surface μ used to build the wormhole embedded in the original spacetime. On the left we see the coarse-grained spacetime in the regions where we are able to obtain it

numerically. We have not obtained any parts of the spacetime in the future or past of X, since this requires evolution past a shockwave, requiring more sophisticated methods than our fairly straightforward technique. The upshot of this solution is that spacetime emergence connecting two asymptotic boundaries via an interior need *not* feature a simple complexity growth with time, emphasizing the necessity of a refinement to the Complexity=Volume proposal that takes into account the different behaviors of deformed theories. We emphasize that this is qualitatively (as well as quantitatively) different from the negative $\mathcal{C}_F$ of the previous section, in which no boundary sources were turned on.

Acknowledgments

It is a pleasure to thank Jan de Boer, Shira Chapman, Sebastian Fischetti, Patrick Jefferson, Lampros Lamprou, Hong Liu, Dan Roberts, and Wati Taylor for discussions. This work is supported in part by NSF grant no. PHY-2011905 and the MIT department of physics. The work of NE was also supported in part by the U.S. Department of Energy, Office of Science, Office of High Energy Physics of U.S. Department of Energy under grant Contract Number DE-SC0012567 (High Energy Theory research) and by the U.S. Department of Energy Early Career Award DE-SC0021886. The work of ÅF is also supported in part by an Aker Scholarship.

9.4 Appendix

9.4.1 The DEC and SEC in type II and $D = 11$ SUGRA

The DEC and the SEC for p-forms

Consider a stress tensor

$$T_{ab} = \frac{1}{p!}\left[pF_a{}^{c_2\ldots c_p}F_{bc_2\ldots c_p} - \frac{k}{2}g_{ab}F^{c_1\ldots c_p}F_{c_1\ldots c_p}\right], \qquad 1 \leq k \leq p, \tag{9.28}$$

where $F_{c_1\ldots c_p}$ is a p-form. For $k = 1$ this is just the stress tensor of a free $(p-1)$ form with curvature $F_{c_1\ldots c_p}$ and action $S = -\frac{1}{2}\int F \wedge \star F$. We first check the DEC. Let u^a, v^b be any two timelike vectors at a point p. We will normalize them to our convenience, since the DEC is independent of the choice of normalization. Choose Riemann normal coordinates at q, so that $g_{\mu\nu}|_q = \eta_{\mu\nu}$ and $u^\mu = (\partial_t)^\mu$, where $\eta_{\mu\nu}$ is the Minkowski metric in Cartesian coordinates. Next, we can always rescale v and perform a rotation of our coordinates so that

$$v^\mu = (\partial_t)^\mu + f(\partial_x)^\mu, \tag{9.29}$$

for some constant $f \geq 0$. With this, we now find at q that

$$v^a u^b T_{ab} = T_{tt} + fT_{tx}. \tag{9.30}$$

We compute that

$$\begin{aligned} p!T_{tt} &= \sum_{\mu_i} \left(pF_t{}^{\mu_2\ldots\mu_p} F_{t\mu_2\ldots\mu_p} + \frac{k}{2} F^{\mu_1\ldots\mu_p} F_{\mu_1\ldots\mu_p} \right) \\ &= \left(p - \frac{k}{2}\right) \sum_{\mu_i} F^2_{t\mu_2\ldots\mu_p} + \frac{k}{2} \sum_{\mu_i, \mu_1 \neq t} F^2_{\mu_1\ldots\mu_p} \\ &\geq \left(p - \frac{k}{2}\right) \sum_{\mu_i} F^2_{t\mu_2\ldots\mu_p} + \frac{k}{2} \sum_{\mu_i} F^2_{x\mu_2\ldots\mu_p}, \end{aligned} \tag{9.31}$$

and

$$p!T_{tx} = \sum_{\mu_i} pF_t{}^{\mu_2\ldots\mu_p} F_{x\mu_2\ldots\mu_p} \quad = p \sum_{\mu_i} F_{t\mu_2\ldots\mu_p} F_{x\mu_2\ldots\mu_p}. \tag{9.32}$$

And so adding up (9.31) and (9.32) we get

$$\begin{aligned} p!T_{ab}u^a v^b &\geq p \sum_{\mu_i} \left[\left(1 - \frac{k}{2p}\right) F^2_{t\mu_2\ldots\mu_p} + \frac{1}{2} f F_{t\mu_2\ldots\mu_p} F_{x\mu_2\ldots\mu_p} + \frac{k}{2} F^2_{x\mu_2\ldots\mu_p} \right] \\ &\quad + \frac{1}{2} pf \sum_{\mu_i} F_{t\mu_2\ldots\mu_p} F_{x\mu_2\ldots\mu_p} \\ &\geq p \sum_{\mu_i} \left[\left(1 - \frac{k}{2p}\right) F^2_{t\mu_2\ldots\mu_p} + \frac{1}{2} f F_{t\mu_2\ldots\mu_p} F_{x\mu_2\ldots\mu_p} + \frac{k}{2} F^2_{x\mu_2\ldots\mu_p} \right]. \end{aligned} \tag{9.33}$$

This is non-negative for each term in the sum. If $F_{t\mu_2\ldots\mu_p} F_{x\mu_2\ldots\mu_p} \geq 0$ this is obvious, since we assumed $1 \leq k \leq p$, giving that each term is manifestly non-negative. So assume $F_{t\mu_2\ldots\mu_p} F_{x\mu_2\ldots\mu_p} < 0$. In this case we get a smaller term if we replace (1) $f \to 2$, (2) $k \to p$ in the first term, and (3) $k \to 1$ in the last term:

$$p!T_{ab}u^a v^b \geq p \sum_{\mu_i} \left[\frac{1}{2} F^2_{t\mu_2\ldots\mu_p} + F_{t\mu_2\ldots\mu_p} F_{x\mu_2\ldots\mu_p} + \frac{1}{2} F^2_{x\mu_2\ldots\mu_p} \right] \geq 0. \tag{9.34}$$

Thus the DEC holds for the stress tensor (9.28). Together with $G_{ab} = 8\pi G_N T_{ab}$, this implies the WCC (9.3).

Next, let us turn to the SEC. Set $8\pi G_N = 1$. Then

$$T \equiv T^a{}_a = g^{ab} G_{ab} = (1 - D/2)\, R \tag{9.35}$$

and so

$$R_{ab} t^a t^b = \left(G_{ab} + \frac{1}{2} g_{ab} R \right) t^a t^b \quad = T_{ab} t^a t^b - \frac{1}{2} R = T_{ab} t^a t^b - \frac{1}{2 - D} T. \tag{9.36}$$

Our stress tensor gives

$$p!T = \frac{1}{2} \left(2p - kD\right) F^{c_1\ldots c_p} F_{c_1\ldots c_p}. \tag{9.37}$$

Choosing again Riemann normal coordinates at our point of interest, we get

$$\begin{aligned} p! R_{\mu\nu} t^\mu t^\nu &= \left(p - \frac{k}{2}\right) \sum_{\mu_i} F^2_{t\mu_2\ldots\mu_p} + \frac{k}{2} \sum_{\mu_i, \mu_1 \neq t} F^2_{\mu_1\ldots\mu_p} - \frac{1}{2}\frac{2p - kD}{2 - D} F_{\mu_1\ldots\mu_p} F^{\mu_1\ldots\mu_p} \\ &\geq \left(p - \frac{k}{2} - \frac{1}{2}\frac{kD - 2p}{D - 2}\right) \sum_{\mu_i} F^2_{t\mu_2\ldots\mu_p} \\ &\geq \left(p - \frac{p}{2} - \frac{1}{2}\frac{pD - 2p}{D - 2}\right) \sum_{\mu_i} F^2_{t\mu_2\ldots\mu_p} \\ &= 0, \end{aligned}$$

where we used that $1 \leq k \leq p$ above. Thus the SEC holds for the stress tensor (9.28).

Type IIB

From [50, 51] we have that the gravitational equations of motion of the bosonic sector of type IIB supergravity in the Einstein frame can be written as

$$\begin{aligned} R_{ab} =& \frac{1}{2}\nabla_a\phi\nabla_b\phi + \frac{1}{2}e^{2\phi}\nabla_a\chi\nabla_b\chi + \frac{1}{96}H_{acdef}H_b{}^{cdef} + \frac{1}{4}e^{-\phi}\left[F_{acd}F_b{}^{cd} - \frac{1}{12}g_{ab}F_{cde}F^{cde}\right] \\ &+ \frac{1}{4}e^{-\phi}\left[L_{acd}L_b{}^{cd} - \frac{1}{12}g_{ab}L_{cde}L^{cde}\right] \end{aligned} \tag{9.38}$$

where H is a five form and F and L are three-forms. The scalar stress tensors are (up to a positive rescaling in the case of χ) just stress tensors of massless scalars and so satisfies the DEC and the SEC. Thus we just need to check that the p-forms. Rewriting the equation in Einstein form we get stress tensors

$$\begin{aligned} T^{(H)}_{ab} &= \frac{1}{5 \times 96}\left(5H_{acdef}H_b{}^{cdef} - \frac{5}{2}g_{ab}H_{cdefg}H^{cdefg}\right) \\ T^{(F)}_{ab} &= \frac{1}{12}e^{-\phi}\left(3F_{acd}F_b{}^{cd} - \frac{1}{2}g_{ab}F_{cde}F^{cde}\right) \\ T^{(L)}_{ab} &= \frac{1}{12}e^{\phi}\left(3L_{acd}L_b{}^{cd} - \frac{1}{2}g_{ab}L_{cde}L^{cde}\right) \end{aligned} \tag{9.39}$$

All of these stress tensors are proportional to (9.28) with a positive coefficient, and so the DEC and SEC holds.

Type IIA and $D = 11$ supergravity

In the Einstein frame, the stress tensors of the bosonic matter in type IIA and eleven-dimensional supergravity is just that of free p-form fields, except for an overall positive factor proportional to an exponential of the dilaton in the case of type IIA [52]. The exact kind of computation as was carried out above shows that the WEC holds classically in these theories.

9.4.2 General properties of $K = 0$ slices in SUGRA

We here derive and highlight some general results on slices of vanishing mean curvature in type IIA/B and $D = 11$ SUGRA. This will justify our statement that the initial data in Sec. 9.2 gives the unique maximal volume slice in the evolved spacetime.

The first observation is the following:

Proposition 1. *The bosonic matter fields of type IIA, IIB and $D = 11$ supergravity in the Einstein frame satisfies the strong energy condition (SEC) and the dominant energy condition (DEC):*

$$\begin{aligned} &SEC: && T_{ab}u^a u^b + \frac{1}{D-2}T^a{}_a \geq 0, && \forall \textit{timelike } u^a, \\ &DEC: && T_{ab}u^a v^a \geq 0, && \forall \textit{timelike } u^a, v^a. \end{aligned} \tag{9.40}$$

This result is known in the literature [53–55], but for convenience we included a derivation above, as the result is often stated without proof. Note that this result applies to the standard bosonic fields, and does not include stringy curvature corrections or additional massive fields. Also, dimensional reduction will in general both break the SEC [55] and the DEC [21] (although for specific types of compactifications they might survive [54]).

The well known fact that the SEC combined with $K = 0$ implies maximality [56, 57] now immediately gives

Proposition 2. *Let (M, g) be any classical solution of type IIA, IIB, or $D = 11$ SUGRA in the Einstein frame. If Σ is a $K = 0$ spacelike hypersurface, possibly with boundary, then Σ is maximal under any variation that leaves its boundary fixed.*

This is a manifestation of the well known fact that the SEC ensures focusing of timelike congruences. The result follows from calculating the second variation of the volume of a spacelike $K = 0$ hypersurface, which reads [56, 57]

$$\delta^2 \,\mathrm{vol}[\Sigma] = -\int_\Sigma (|DN|^2 + N^2 K_{ab}K^{ab} + N^2 R_{ab}n^a n^a) \leq 0, \tag{9.41}$$

where τ^a is the vector field generating the variation of Σ, n^a a unit normal to Σ, D_a the connection on Σ, and $N = \tau \cdot n$. It is assumed that the boundary of Σ is kept fixed in (9.41).

As described above, Proposition 2 will generally not be true in dimensionally reduced spacetime, and so in this case we actually have better control in the full $D = 10, 11$ spacetime. This shows a situation in AdS/CFT where the compact dimensions should be viewed as a resource rather than a nuissance.

Next, [58] has showed that if the SEC holds, then there cannot be two $K = 0$ slices anchored at the same boundary time in an asymptotically AdS spacetime (the proof survives when we also have a compact space). Thus we have the result

Proposition 3. *Let (M, g) be an asymptotically $AdS_{d+1} \times X$ solution of type IIA, IIB, or $D = 11$ SUGRA in the Einstein frame for some compact manifold X. Let Σ be a complete maximal volume slice anchored at boundary time C. Then there is no other maximal volume slice in the domain of dependence of Σ that is anchored at C.*

Proposition 2 and 3 now justifies our assertion from Sec. 9.2 that our initial data Σ is the true maximal volume slice.

Finally, we remark that Proposition 2 and 3 remain true if we add additional SEC-respecting matter, such p-dimensional branes B with action

$$S = -\int_B \mathrm{d}^p x \sqrt{-h} T + \int_B C_p, \tag{9.42}$$

where T is a non-negative potentially field-dependent scalar and C_p a p-form – both independent of the induced metric h_{ab} on B.

9.4.3 AdS Vacuum Decay Computations

Kruskal coordinates

Define

$$\hat{a}(t) = a(it) = (1+c)\sin t - 2c \sin\frac{t}{2}. \tag{9.43}$$

Consider the metric (9.18) in the special case of $c = 1$ and $d = 3$. Define the functions

$$R(\xi) = \int_{\xi_0}^{\xi} \frac{\mathrm{d}\xi'}{a(\xi')} = \frac{1}{3}\log\left[\frac{f(\xi)}{f(\xi_0)}\right], \qquad T(t) = \int_{t_0}^{t} \frac{\mathrm{d}t'}{\hat{a}(t')} = \frac{1}{3}\log\left[\frac{\hat{f}(t)}{\hat{f}(t_0)}\right] \tag{9.44}$$

where

$$\begin{aligned} f(\xi) &= \frac{\cosh\left(\frac{\xi}{4}\right)\sinh\left(\frac{\xi}{4}\right)^3}{\left(1-2\cosh\left(\frac{\xi}{2}\right)\right)^2} = \begin{cases} \frac{\xi^3}{64} + \mathcal{O}(\xi^5) & \xi \ll 1 \\ \frac{1}{16} - \frac{3}{16}e^{-\xi} - \frac{1}{8}e^{-3\xi/2} + \mathcal{O}(e^{-2\xi}) & \xi \gg 1 \end{cases} \\ \hat{f}(t) &= \frac{\cos\left(\frac{t}{4}\right)\sin\left(\frac{t}{4}\right)^3}{\left(1-2\cos\left(\frac{t}{2}\right)\right)^2} = \begin{cases} \frac{t^3}{64} + \mathcal{O}(t^5) & t \ll 1 \\ \frac{1}{4\sqrt{3}(t_*-t)^2} - \frac{1}{8(t_*-t)} + \mathcal{O}(1) & t_* - t \ll 1 \end{cases} \end{aligned} \tag{9.45}$$

where $t_* = \frac{2\pi}{3}$ is the location of the future singularity, ie. the maximal value of t. The constants $t_0 > 0$ and $\xi_0 > 0$ will be chosen later. We have $T \in \mathbb{R}$ with $T = \infty$ the future singularity and $T = -\infty$ the future event horizon, so the coordinate T covers only the future FRW region. We have $R \in (-\infty, R_\partial)$ where $R = -\infty$ is the event horizon and R_∂ the conformal boundary. Define now Kruskal coordinates

$$U = \begin{cases} e^{T(t)-\rho} & \text{Region II} \\ -e^{-\zeta+R(\xi)} & \text{Region I} \end{cases} \qquad V = \begin{cases} e^{T(t)+\rho} & \text{Region II} \\ e^{\zeta+R(\xi)} & \text{Region I} \end{cases}. \tag{9.46}$$

where we now take Region II temporarily to mean the future part only. This gives

$$T(U,V) = \frac{1}{2}\log(UV), \qquad \rho(U,V) = \frac{1}{2}\log(V/U), \qquad U > 0, \tag{9.47}$$

$$\zeta(U,V) = \frac{1}{2}\log(-V/U), \qquad R(U,V) = \frac{1}{2}\log(-VU), \qquad U < 0. \tag{9.48}$$

The future event horizon is now at $U = 0$ and the past event horizon at $V = 0$. The octant $V \geq U \geq 0$ covers the future part of of region I, with $\rho = 0$ at $U = V$. The singularity lies at

$(U > 0, V = \infty)$. The conformal boundary is at $VU = -e^{2R_\partial}$, and region I is covered by the regions $U \leq 0$ and $0 \leq V \leq -\frac{e^{2R_\partial}}{U}$.

Let us now define the functions

$$\hat{t}(X) = T^{-1}\left(\frac{1}{2}\log X\right), \qquad \hat{\xi}(X) = R^{-1}\left(\frac{1}{2}\log(-X)\right), \tag{9.49}$$

so that $t(U,V) = \hat{t}(UV)$ and $\xi(U,V) = \hat{\xi}(UV)$. The domain of $\hat{t}$ is $X \in (0,\infty)$. Since $R \in (-\infty, R_{\max})$, we have that the domain of $\hat{\xi}$ is $X \in (-e^{2R_{\max}}, 0)$. Finally changing the coordinates, we find that the metric is

$$\mathrm{d}s^2 = b(UV)\left[-\mathrm{d}U\mathrm{d}V + \left(\frac{V-U}{2}\right)^2 \mathrm{d}\Omega^2\right], \tag{9.50}$$

where

$$b(UV) = \begin{cases} \frac{\hat{a}(\hat{t}(UV))^2}{\sqrt{U^2V^2}}, & U > 0, V > 0, \\ \frac{a(\hat{\xi}(UV))^2}{\sqrt{U^2V^2}}, & U < 0, V > 0. \end{cases} \tag{9.51}$$

Note that the inverse functions T^{-1} and R^{-1} must be computed numerically, and so the same is also true of b.

In order for $b(X)$ to be continous at $X = 0$ we need to choose ξ_0, t_0 appropriately. For small arguments we have

$$\begin{aligned} \frac{\log(-UV)}{2} &= R(\xi) = \log\left(\frac{\xi}{4}\right) - \frac{1}{3}\log f(\xi_0) + \mathcal{O}(\xi^2), \\ \frac{\log UV}{2} &= T(t) = \log\left(\frac{t}{4}\right) - \frac{1}{3}\log \hat{f}(t_0) + \mathcal{O}(t^2), \end{aligned} \tag{9.52}$$

which at small t and ξ gives the relation

$$\begin{aligned} \xi &= 4f(\xi_0)^{1/3}\sqrt{-UV}, \\ t &= 4\hat{f}(t_0)^{1/3}\sqrt{UV}. \end{aligned} \tag{9.53}$$

Now, near the horizon we have that $\hat{a}(t) = t + \ldots$ and $a(\xi) = \xi + \ldots$, so the function b near the horizon reads

$$b(UV) = \begin{cases} 16\hat{f}(t_0)^{2/3} + \ldots \\ 16 f(\xi_0)^{2/3} + \ldots \end{cases}. \tag{9.54}$$

Thus, continuity of b requires t_0 and ξ_0 to be related by

$$f(\xi_0) = \hat{f}(t_0). \tag{9.55}$$

The conformal diagram

The metric (9.50) is just conformal to Minkowski, and so drawing the conformal diagram is just as for Minkowski, with two exceptions:

- In the region of negative U, $-UV \geq e^{2R_\partial}$ is excised since it lies beyond the conformal boundary.
- The part which corresponds to null infinity in Minkowski (and which is not in the excised region) is here instead a singularity at a finite proper distance.

From these observations, we easily find Fig. 9.2. In our representation have chosen $t_0 = 1$ and rescaled the null coordinates by a convenient overall factor in order to bring the holographic screen closer to the center of the diagram.

The holographic screen

Consider the radial null vectors

$$k^a = \frac{1}{\sqrt{2}\hat{a}(t)}\left[(\partial_t)^a + \frac{1}{\hat{a}(t)}(\partial_\rho)^a\right], \qquad \ell^a = \frac{\hat{a}(t)}{\sqrt{2}}\left[(\partial_t)^a - \frac{1}{\hat{a}(t)}(\partial_\rho)^a\right], \tag{9.56}$$

which are normalized so that $k \cdot \ell = -1$ and so that $k^a \nabla_a k^b = 0$. Calculating the expansions, we find

$$\begin{aligned} \theta_k &= \frac{\sqrt{2}}{\hat{a}(t)^2}\left(\hat{a}'(t) + \frac{1}{\tanh\rho}\right) \\ \theta_\ell &= -\sqrt{2}\left(-\hat{a}'(t) + \frac{1}{\tanh\rho}\right) \end{aligned} \tag{9.57}$$

For times where $\hat{a}'(t) < -1$, we have marginally trapped surfaces at

$$\rho(t) = \operatorname{arctanh}\left(-\frac{1}{\hat{a}'(t)}\right). \tag{9.58}$$

The unnormalized tangents to the screen that are orthogonal to the marginally trapped leaves are

$$\eta^a = (\partial_t)^a + \rho'(t)(\partial_\rho)^a = (\partial_t)^a + \frac{\hat{a}''(t)}{1 - \hat{a}'(t)^2}(\partial_\rho)^a. \tag{9.59}$$

Its norm is given by

$$\eta^2 = -1 + \frac{\hat{a}(t)^2\hat{a}''(t)^2}{(\hat{a}'(t)^2 - 1)^2}, \tag{9.60}$$

which starts out being positive and then transitions to negative at late times.

Timelike geodesics must cross the horizon

Consider a timelike geodesic in region II, which by spherical symmetry can be taken to lie at $\theta = \pi/2$ without loss of generality. The effective Lagrangian for a geodesic is

$$\mathcal{L} = g_{\mu\nu}u^\mu u^\nu = \dot{\xi}^2 - a(\xi)^2\dot{\zeta}^2 + a(\xi)^2\cosh^2\zeta\dot{\varphi}^2, \tag{9.61}$$

Parametrizing by proper time, so that $\mathcal{L} = -1$, gives

$$-\dot{\zeta}^2 + \cosh^2\zeta\dot{\varphi}^2 = -\frac{1 + \dot{\xi}^2}{a^2}. \tag{9.62}$$

Then the geodesic equation for ξ can then be written

$$0 = \ddot{\xi} - aa'\left(-\dot{\zeta}^2 + \cosh^2\zeta\dot{\varphi}^2\right) = \ddot{\xi} + \frac{a'}{a}(1+\dot{\xi}^2). \tag{9.63}$$

Interestingly the angular momentum makes no presence here, so there is no angular moment barrier between the horizon and the conformal boundary. Since $a'(\xi)/a(\xi) > 1$ everywhere, this equation effectively describes a particle subject to friction and a force always pushing in the direction of negative ξ. Thus no geodesic can avoid reaching $\xi = 0$. After this, it enters region II, where it is doomed to encounter the singularity in finite time. Thus, every timelike geodesic ends up in the singularity, and so the singularity is rightfully described as cosmological.

Computing $\dot{\mathcal{C}}_V$

Let us consider the static cylinder conformal frame. Defining $\xi = -\log z$, the metric in the causal wedge becomes

$$\begin{aligned} \mathrm{d}s_I^2 &= \frac{1}{z^2}\left[\mathrm{d}z^2 + f(z)^2\left(\mathrm{d}\zeta^2 + \cosh^2\zeta\mathrm{d}\Omega^2\right)\right] \\ f(z) &= \frac{[1-z]\,[1+c(\sqrt{z}-1)^2+z]}{2} = \frac{1+c}{2} - c\sqrt{z} + \mathcal{O}(z). \end{aligned} \tag{9.64}$$

Next we introduce new coordinates (w,t) through:

$$\begin{aligned} \zeta &= 2\,\mathrm{arctanh}\left(\tan\frac{t}{2}\right), \\ z &= w\frac{c+1}{2\cos t}. \end{aligned} \tag{9.65}$$

To subleading order in w, this transforms the metric into Fefferman-Graham coordinates of the static cylinder:

$$\begin{aligned} \mathrm{d}s_I^2 &= \frac{1}{w^2}\left[\mathrm{d}w^2 + h(w,t)^2\left(-\mathrm{d}t^2 + \mathrm{d}\Omega^2\right) + \mathcal{O}(w)\right], \\ h(w,t) &= 1 - \sqrt{w}\sqrt{\frac{2c^2}{(c+1)\cos t}}, \end{aligned} \tag{9.66}$$

where t lies in the finite range $|t| < \pi/2$. Consider now a hypersurface Σ_{t_0} with embedding coordinates $(w, t(w), \Omega_i)$, and where $t(0) = t_0$. Its volume reads

$$\mathrm{vol}[\Sigma_{t_0}] = \mathrm{vol}[S^{d-1}]\int \mathrm{d}w \frac{h(w,t)^{d-1}}{w^d}\sqrt{1 - h(w,t)t'(w)^2}. \tag{9.67}$$

Expanding near the boundary,

$$t(w) = t_0 + t_1 w + \mathcal{O}(w^{3/2}), \tag{9.68}$$

we find that extremality imposes $t_1 = 0$. Consequentially, the divergence of the volume to subleading order is

$$\begin{aligned}\mathrm{vol}[\Sigma_{t_0}] &= \mathrm{vol}[S^{d-1}] \int_\epsilon \mathrm{d}w \left(w^{-d} - w^{-d+1/2}(d-1)\sqrt{\frac{2c^2}{(c+1)\cos t_0}} + \mathcal{O}(w^{-d+1}) \right) \\ &= \mathrm{vol}[S^{d-1}] \left[\frac{1}{(d-1)} \epsilon^{-d+1} - \frac{d-1}{d-\frac{3}{2}} \epsilon^{-d+\frac{3}{2}} \sqrt{\frac{2c^2}{(c+1)\cos t_0}} + \mathcal{O}(\epsilon^{-d+2}) \right].\end{aligned} \tag{9.69}$$

This implies that the leading order complexity change with cutoff adapted to the static cylinder is given by (9.27).

Plot of the Cosmological Wormhole

Fig. 9.4 shows the domain for which the numerical determination of the coarse grained spacetime has been carried out, together with data on null expansions and area–radii in the geometry.

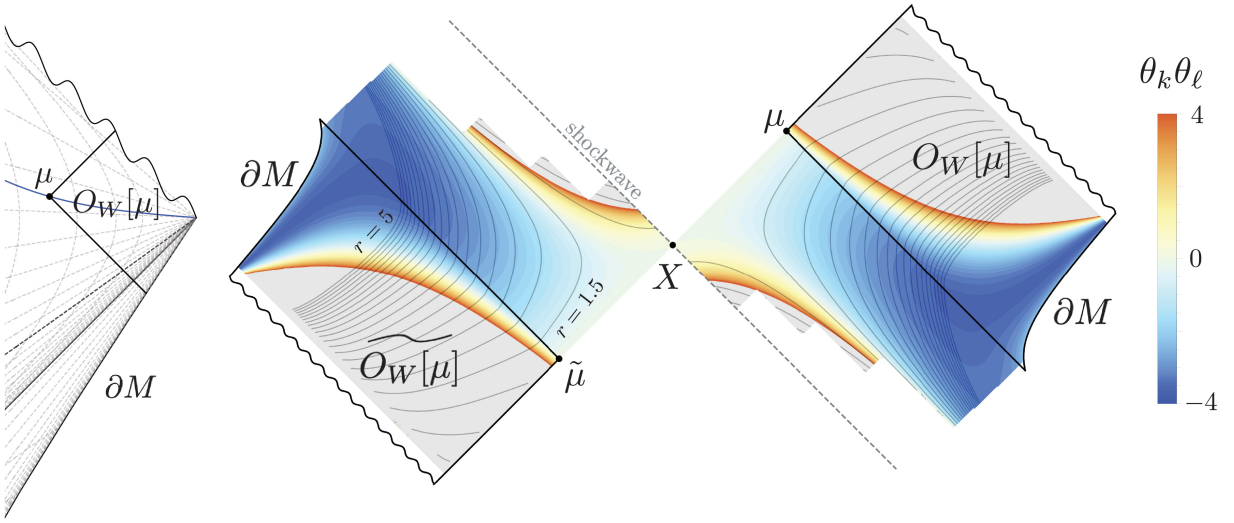


Figure 9.4: To the left, we have zoomed into the outer wedge $O_W[\mu]$ of a marginally trapped surface μ on the spacelike section of the holographic screen in Fig. 9.2. On the right, we show the coarse-grained spacetime corresponding to μ in the regions where we have been able to obtain the metric numerically. The black contour lines show surfaces of constant area radius, saturating at $r = 5$ and with spacings of $\delta r \approx 0.2$. The colored contours show the product of the null expansions $\theta_k\theta_\ell$ for constant$-r$ surfaces, with gray regions corresponding to $\theta_k\theta_\ell > 4$. The shockwave passing through the HRT surface X carries no null energy, but does source a discontinuity in the inaffinity of ℓ^a, which is the null vector along the direction of the shockwave. The quantity $\ell^a \nabla_a \phi$ is discontinuous at X.

References

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

[2] S. Ryu and T. Takayanagi, *Aspects of Holographic Entanglement Entropy*, *JHEP* **0608** (2006) 045, [hep-th/0605073].

[3] V. E. Hubeny, M. Rangamani, and T. Takayanagi, *A Covariant holographic entanglement entropy proposal*, *JHEP* **0707** (2007) 062, [arXiv:0705.0016].

[4] A. C. Wall, *Maximin Surfaces, and the Strong Subadditivity of the Covariant Holographic Entanglement Entropy*, *Class.Quant.Grav.* **31** (2014), no. 22 225007, [arXiv:1211.3494].

[5] B. Czech, J. L. Karczmarek, F. Nogueira, and M. Van Raamsdonk, *The Gravity Dual of a Density Matrix*, *Class.Quant.Grav.* **29** (2012) 155009, [arXiv:1204.1330].

[6] T. Faulkner, A. Lewkowycz, and J. Maldacena, *Quantum corrections to holographic entanglement entropy*, *JHEP* **1311** (2013) 074, [arXiv:1307.2892].

[7] A. Lewkowycz and J. Maldacena, *Generalized gravitational entropy*, *JHEP* **1308** (2013) 090, [arXiv:1304.4926].

[8] X. Dong, *Holographic Entanglement Entropy for General Higher Derivative Gravity*, *JHEP* **01** (2014) 044, [arXiv:1310.5713].

[9] J. Maldacena and L. Susskind, *Cool horizons for entangled black holes*, arXiv:1306.0533.

[10] N. Engelhardt and A. C. Wall, *Quantum Extremal Surfaces: Holographic Entanglement Entropy beyond the Classical Regime*, *JHEP* **01** (2015) 073, [arXiv:1408.3203].

[11] L. Susskind, *Computational Complexity and Black Hole Horizons*, *Fortsch. Phys.* **64** (2016) 24–43, [arXiv:1403.5695]. [Addendum: Fortsch.Phys. 64, 44–48 (2016)].

[12] D. Stanford and L. Susskind, *Complexity and Shock Wave Geometries*, *Phys. Rev. D* **90** (2014), no. 12 126007, [arXiv:1406.2678].

[13] A. R. Brown, D. A. Roberts, L. Susskind, B. Swingle, and Y. Zhao, *Holographic Complexity Equals Bulk Action?*, *Phys. Rev. Lett.* **116** (2016), no. 19 191301, [arXiv:1509.07876].

[14] A. R. Brown, D. A. Roberts, L. Susskind, B. Swingle, and Y. Zhao, *Complexity, action, and black holes*, *Phys. Rev. D* **93** (2016), no. 8 086006, [arXiv:1512.04993].

[15] D. L. Jafferis, A. Lewkowycz, J. Maldacena, and S. J. Suh, *Relative entropy equals bulk relative entropy*, arXiv:1512.06431.

[16] X. Dong, *The Gravity Dual of Renyi Entropy*, *Nature Commun.* **7** (2016) 12472, [arXiv:1601.06788].

[17] N. Engelhardt and Å. Folkestad, *General Bounds on Holographic Complexity*, arXiv:2109.06883.

[18] S. Chapman, H. Marrochio, and R. C. Myers, *Complexity of Formation in Holography*, *JHEP* **01** (2017) 062, [arXiv:1610.08063].

[19] S. Chapman, D. Ge, and G. Policastro, *Holographic Complexity for Defects Distinguishes Action from Volume*, *JHEP* **05** (2019) 049, [arXiv:1811.12549].

[20] A. Bernamonti, F. Galli, J. Hernandez, R. C. Myers, S.-M. Ruan, and J. Simón, *Aspects of The First Law of Complexity*, arXiv:2002.05779.

[21] T. Hertog, G. T. Horowitz, and K. Maeda, *Negative energy density in Calabi-Yau compactifications*, *JHEP* **05** (2003) 060, [hep-th/0304199].

[22] P. Breitenlohner and D. Z. Freedman, *Positive energy in anti-de Sitter backgrounds AND gauged extended supergravity*, *Phys. Lett.* **B115** (1982) 197.

[23] T. Hertog and G. T. Horowitz, *Towards a big crunch dual*, *JHEP* **07** (2004) 073, [hep-th/0406134].

[24] T. Hertog and G. T. Horowitz, *Holographic description of AdS cosmologies*, *JHEP* **04** (2005) 005, [hep-th/0503071].

[25] S. Fischetti, D. Marolf, and A. C. Wall, *A paucity of bulk entangling surfaces: AdS wormholes with de Sitter interiors*, *Class.Quant.Grav.* **32** (2015), no. 6 065011, [arXiv:1409.6754].

[26] M. Moosa, *Divergences in the rate of complexification*, *Phys. Rev. D* **97** (2018), no. 10 106016, [arXiv:1712.07137].

[27] A. Belin, A. Lewkowycz, and G. Sárosi, *Complexity and the bulk volume, a new York time story*, *JHEP* **03** (2019) 044, [arXiv:1811.03097].

[28] N. Chagnet, S. Chapman, J. de Boer, and C. Zukowski, *Complexity for Conformal Field Theories in General Dimensions*, arXiv:2103.06920.

[29] M. Flory and M. P. Heller, *Geometry of Complexity in Conformal Field Theory*, *Phys. Rev. Res.* **2** (2020), no. 4 043438, [arXiv:2005.02415].

[30] M. Flory and M. P. Heller, *Conformal field theory complexity from Euler-Arnold equations*, *JHEP* **12** (2020) 091, [arXiv:2007.11555].

[31] M. Cvetic, M. J. Duff, P. Hoxha, J. T. Liu, H. Lu, J. X. Lu, R. Martinez-Acosta, C. N. Pope, H. Sati, and T. A. Tran, *Embedding AdS black holes in ten-dimensions and eleven-dimensions*, *Nucl. Phys. B* **558** (1999) 96–126, [hep-th/9903214].

[32] D. Marolf, O. Parrikar, C. Rabideau, A. Izadi Rad, and M. Van Raamsdonk, *From Euclidean Sources to Lorentzian Spacetimes in Holographic Conformal Field Theories*, *JHEP* **06** (2018) 077, [arXiv:1709.10101].

[33] V. Balasubramanian and P. Kraus, *A Stress tensor for Anti-de Sitter gravity*, *Commun. Math. Phys.* **208** (1999) 413–428, [hep-th/9902121].

[34] G. Perelman, *The entropy formula for the Ricci flow and its geometric applications*, *arXiv Mathematics e-prints* (Nov., 2002) math/0211159, [math/0211159].

[35] G. Perelman, *Ricci flow with surgery on three-manifolds*, *arXiv Mathematics e-prints* (Mar., 2003) math/0303109, [math/0303109].

[36] I. Agol, N. M. Dunfield, P. A. Storm, and W. P. Thurston, *Lower bounds on volumes of hyperbolic Haken 3-manifolds*, *arXiv Mathematics e-prints* (June, 2005) math/0506338, [math/0506338].

[37] R. M. Schoen, *Variational theory for the total scalar curvature functional for riemannian metrics and related topics*, in *Topics in Calculus of Variations* (M. Giaquinta, ed.), (Berlin, Heidelberg), pp. 120–154, Springer Berlin Heidelberg, 1989.

[38] H. L. Bray, *The Penrose inequality in general relativity and volume comparison theorems involving scalar curvature*. PhD thesis, STANFORD UNIVERSITY, Nov., 1997.

[39] M. Henneaux, C. Martinez, R. Troncoso, and J. Zanelli, *Black holes and asymptotics of 2+1 gravity coupled to a scalar field*, *Phys. Rev. D* **65** (2002) 104007, [hep-th/0201170].

[40] G. T. Horowitz, *Creating naked singularities and negative energy*, *Phys. Scripta T* **117** (2005) 86–91, [hep-th/0312123].

[41] M. Henneaux, C. Martinez, R. Troncoso, and J. Zanelli, *Asymptotically anti-de Sitter spacetimes and scalar fields with a logarithmic branch*, *Phys. Rev. D* **70** (2004) 044034, [hep-th/0404236].

[42] T. Hertog and K. Maeda, *Black holes with scalar hair and asymptotics in $N = 8$ supergravity*, *JHEP* **07** (2004) 051, [hep-th/0404261].

[43] M. Henneaux, C. Martinez, R. Troncoso, and J. Zanelli, *Asymptotic behavior and Hamiltonian analysis of anti-de Sitter gravity coupled to scalar fields*, *Annals Phys.* **322** (2007) 824–848, [hep-th/0603185].

[44] X. Dong and D. Harlow, *Analytic Coleman-De Luccia Geometries*, *JCAP* **1111** (2011) 044, [arXiv:1109.0011].

[45] C. Fefferman and C. R. Graham, *Conformal invariants*, *Elie Cartan et les Mathématiques d'aujourdhui* (Astérisque) p. 95. 1985.

[46] C. Graham and J. M. Lee, *Einstein metrics with prescribed conformal infinity on the ball*, *Advances in Mathematics* **87** (1991), no. 2 186–225.

[47] C. R. Graham and E. Witten, *Conformal anomaly of submanifold observables in AdS / CFT correspondence*, *Nucl. Phys. B* **546** (1999) 52–64, [hep-th/9901021].

[48] N. Engelhardt and A. C. Wall, *Coarse Graining Holographic Black Holes*, *JHEP* **05** (2019) 160, [arXiv:1806.01281].

[49] N. Engelhardt and A. C. Wall, *Decoding the Apparent Horizon: Coarse-Grained Holographic Entropy*, *Phys. Rev. Lett.* **121** (2018), no. 21 211301, [arXiv:1706.02038].

[50] H. Lu, C. N. Pope, and T. A. Tran, *Five-dimensional N=4, SU(2) x U(1) gauged supergravity from type IIB*, *Phys. Lett. B* **475** (2000) 261–268, [hep-th/9909203].

[51] D. Cassani, G. Dall'Agata, and A. F. Faedo, *Type IIB supergravity on squashed Sasaki-Einstein manifolds*, *JHEP* **05** (2010) 094, [arXiv:1003.4283].

[52] M. J. D. Hamilton, *The field and Killing spinor equations of M-theory and type IIA/IIB supergravity in coordinate-free notation*, arXiv:1607.00327.

[53] J. M. Maldacena and C. Nunez, *Supergravity description of field theories on curved manifolds and a no go theorem*, *Int. J. Mod. Phys. A* **16** (2001) 822–855, [hep-th/0007018].

[54] G. W. Gibbons, *Thoughts on tachyon cosmology*, *Class. Quant. Grav.* **20** (2003) S321–S346, [hep-th/0301117].

[55] P. K. Townsend and M. N. R. Wohlfarth, *Accelerating cosmologies from compactification*, *Phys. Rev. Lett.* **91** (2003) 061302, [hep-th/0303097].

[56] D. Brill and F. Flaherty, *Isolated maximal surfaces in spacetime*, *Communications in Mathematical Physics* **50** (1976), no. 2 157–165.

[57] J. E. Marsden and F. J. Tipler, *Maximal hypersurfaces and foliations of constant mean curvature in general relativity*, *Physics Reports* **66** (1980), no. 3 109–139.

[58] J. Couch, S. Eccles, T. Jacobson, and P. Nguyen, *Holographic Complexity and Volume*, *JHEP* **11** (2018) 044, [arXiv:1807.02186].

Chapter 10

Maximal Entangling Rates from Holography

Abstract: We prove novel speed limits on the growth of entanglement, equal-time correlators, and spacelike Wilson loops in spatially uniform time-evolving states in strongly coupled CFTs with holographic duals. These bounds can also be viewed as quantum weak energy conditions. Several of the speed limits are valid for regions of arbitrary size and with multiple connected components, and our findings imply new bounds on the effective entanglement velocity of small subregions. In 2d CFT, our results prove a conjecture by Liu and Suh for a large class of states. We also bound spatial derivatives of entanglement and correlators. Key to our findings is a momentum-entanglement correspondence, showing that entanglement growth is computed by the momentum crossing the HRT surface. In our setup, we prove a number of general features of boundary-anchored extremal surfaces, such as a sharp bound on the smallest radius that a surface can probe, and that the tips of extremal surfaces cannot lie in trapped regions. Our methods rely on novel global GR techniques, including a delicate interplay between Lorentzian and Riemannian Hawking masses. While our proofs assume the dominant energy condition in the bulk, we provide numerical evidence that our bounds are true under less restrictive assumptions.

10.1 Introduction

Entanglement is one of the key features unique to quantum mechanics, and its effects are ubiquitous in modern physics. It is now clear that entanglement and entanglement entropy is a central quantity across a diverse range of fields, such as quantum many-body physics [1–3], quantum information theory [4–6], quantum gravity [7–21], and quantum field theories and their RG flows [22–30].

A central question relevant to all of the above subjects is how entanglement behaves dynamically. In this paper, we address the following questions: are there general bounds on the entanglement entropy in time-dependent states? Does there exist speed limits on how fast it can grow? The latter question is relevant to understanding how rapidly quantum

information can propagate, how long it takes a many-body system to thermalize, or, in quantum gravity, for constraining the dynamics of spacetime itself.

While calculating entanglement entropies is notoriously hard, many lessons have been learned over the last two decades. Quantum quenches in particular have received considerable interest. In a quantum quench, the Hamiltonian is abruptly changed, or a source is turned on over a small time interval δt. In either case, there is an abrupt injection of energy into the system, kicking the state out of equilibrium. The subsequent approach to equilibrium can then be computed in various setups. In the seminal paper by Calabrese and Cardy [23], the entanglement entropy S_R of an interval R of length ℓ in a $(1+1)$-dimensional conformal field theory (CFT) after a uniform quench was computed, and for large times and interval lengths, it was found to behave as

$$S_R(t) - S_R(t=0) = \begin{cases} 2s_{\text{th}}t & t < \ell/2 \\ s_{\text{th}}\ell & t \geq \ell/2 \end{cases}, \tag{10.1}$$

where s_{th} the thermal entropy density of the final state. Linear growth of entanglement for large regions R after uniform quenches has also been found in higher dimensional holographic CFTs [31–34]. In particular, after local equilibration and before late time saturation, the entanglement entropy of a region R after a quench was found to behave as [33, 34]

$$S_R(t) - S_R(t=0) = v_E s_{\text{th}} \text{Area}[\partial R] t + \dots \tag{10.2}$$

with v_E the so-called entanglement velocity, which satisfies $v_E \leq 1$.

While quenches provide useful insights on entanglement dynamics, they do not cover all kinds of states, and it would be useful to have more general constraints. Some such results do exist. Consider two quantum systems $A \cup a$ and $B \cup b$ coupled by an interaction Hamiltonian H acting only on A and B. In [35] (building on [36]) it was proven that

$$\left| \frac{\mathrm{d}S_{A\cup a}}{\mathrm{d}t} \right| \leq \eta \|H\| \log d, \tag{10.3}$$

where $d = \min\{\dim A, \dim B\}$, and where η is an order 1 constant. While this bound has broad generality for finite-dimensional systems, it is not useful in QFT, where d is infinite. Even if we UV-regulate to make d finite, $\|H\|$ is infeasible to compute. Furthermore, the bound is state-independent, and it is natural to suspect there exists stronger bounds that depend on the conserved charges of the state.

A bound more useful in QFT was conjectured [33, 34], based on the findings in holographic quenches. It was proposed that a normalized instantaneous entanglement growth $\mathfrak{R}$ in relativistic QFT satisfies the bound

$$\mathfrak{R} \equiv \frac{1}{\text{Area}[\partial R] s_{\text{th}}} \left| \frac{\mathrm{d}S_R}{\mathrm{d}t} \right| \leq 1. \tag{10.4}$$

In [37] relativistic QFT was used to prove that $\mathfrak{R} \leq 1$ for large convex regions R in spatially uniform states, neglecting contributions to S_R not scaling with volume.[1] However, it was

[1] A proof was also given in [38] for half-planes, taking linear growth of entanglement as an assumption. For quenches, it was proven for large regions holographically in [39], together with many other properties of quenches.

found in [34] that the largest values for $\mathfrak{R}$ were obtained for intermediate sized regions, where it could exceed v_E (for $d > 2$), and where existing proofs of $\mathfrak{R} \leq 1$ do not apply. Thus, the validity of (10.4) for general regions is still an open question.[2]

In this work, for holographic CFTs with large coupling and large$-N$ (large effective central charge), we prove novel bounds that imply $\mathfrak{R} \leq 1$ for a large class of situations not covered by [37, 38]. We also prove several bounds that to our knowledge have not been previously discussed, including growth bounds on correlators and Wilson loops. We will see that our growth bounds can be seen as new types of quantum energy conditions, valid for uniform states. We also derive absolute bounds on entanglement entropy and equal-time correlators.

Let us now summarize our results. Consider first a 2d CFT on $S^1 \times \mathbb{R}$ or Minkowski space in a homogeneous and isotropic state undergoing time-evolution. Let t label the timeslices on which the state is uniform. Let R be a union of n finite intervals of any size. Assuming an energy condition and certain falloff conditions on the matter fields in the bulk, which we assume for all bounds presented in the following, we prove that

$$\left|\frac{\mathrm{d}S_R}{\mathrm{d}t}\right| \leq n\sqrt{\frac{8\pi c}{3}\big(\langle T_{tt}\rangle - \langle T_{tt}\rangle_{\text{vacuum}}\big)}, \tag{10.5}$$

where c is the central charge and $\langle T_{tt}\rangle$ the CFT energy density one-point function, which is the same everywhere in a uniform state. If we work with uncharged states, (10.5) implies that $\mathfrak{R} \leq 1$. Thus, for the $2d$ theories under consideration, we have given a proof of $\mathfrak{R} \leq 1$ to regions of arbitrary finite size and with any number of connected components.[3]

Next, consider $d \geq 2$-dimensional holographic CFTs on Minkowski space, again in a time-evolving uniform state. Taking R to be either a single ball or strip of characteristic size ℓ, we prove that

$$\left|\frac{\mathrm{d}S_R}{\mathrm{d}t}\right| \leq \kappa \mathrm{Vol}[R]\,\langle T_{tt}\rangle\left[1 + \mathcal{O}\left(\frac{\ell^d\,\langle T_{tt}\rangle}{c_{\text{eff}}}\right)\right], \tag{10.6}$$

where κ is an $O(1)$ numerical constant given in (10.31), and which depends on d and the shape of R. c_{eff} is the effective central charge, to be defined in the following. For small regions this bound is much stronger than $\mathfrak{R} \leq 1$.[4] If β is the effective inverse temperature at which the thermal energy density equals $\langle T_{tt}\rangle$, we get

$$\mathfrak{R} \leq \mathcal{O}(\ell/\beta) \ll 1. \tag{10.7}$$

We also prove a higher-dimensional analogue of (10.5), although the proof is more limited.

[2] In fact, since generic QFTs can have state-dependent divergences in S_R [40], $\partial_t S_R$ can be divergent in some theories, and so (10.4) cannot be true in all relativistic QFTs. This means that a generalization of the proofs of [37, 38] to include contributions not scaling with volume impossible without more input on the theories under consideration.

[3] For holographic CFTs, (10.5) also improves on a bound proven for single intervals of any size in all $2d$ CFTs by [37], which can be written as $\mathfrak{R} \leq \coth\left(\pi\ell\sqrt{\frac{\pi c}{6\langle T_{tt}\rangle}}\right)$.

[4] See [41] for a discussion of a different definition of $\mathfrak{R}$, where s_{th} is replaced by the vacuum-subtracted entanglement entropy per volume in the final state. With this definition, $\mathfrak{R}$ is $O(1)$ for small subregions, but it can exceed 1.

We prove for states that are somewhat more general than quench states that

$$\left|\frac{\mathrm{d}S_R}{\mathrm{d}t}\right| \leq \frac{1}{4}\mathrm{Area}[\partial R] c_{\mathrm{eff}} \left[\frac{16\pi}{(d-1)c_{\mathrm{eff}}} \langle T_{tt}\rangle\right]^{\frac{d-1}{d}}, \tag{10.8}$$

where R either is a single ball, or the union of any number of strips. Considering a neutral state, (10.8) translates into $\mathfrak{R} \leq 1$. While our proof of (10.8) applies to a smaller class of states, we give substantial numerical evidence that (10.8) holds more generally for all uniform states.

For strips, we also prove bounds on the entanglement entropy itself. For R_ℓ a strip of width ℓ at fixed time t, we prove that the vacuum subtracted entropy $\Delta S(\ell)$ satisfies

$$\partial_\ell \Delta S(\ell) \geq 0, \tag{10.9}$$

which in particular implies $\Delta S \geq 0$.

In the special dimensions of $d = 2, 3, 4$, we prove additional bounds. Assuming the geodesic approximation for correlators [42], we prove that the equal-time two-point function of a scalar operator O of large scaling dimension Δ in $d = 2$ satisfies

$$\left|\frac{\mathrm{d}}{\mathrm{d}t} \log \langle O(x)O(0)\rangle_{\rho(t)}\right| \leq \sqrt{\frac{96\pi\Delta^2}{c}\left(\langle T_{tt}\rangle - \langle T_{tt}\rangle_{\mathrm{vac}}\right)}, \tag{10.10}$$

where $\rho(t)$ is the state under consideration. This bound is saturated in the global CFT_2 quenches studied in [43, 44] (for any Δ and c). We also prove a tighter bound when x is small:

$$\left|\frac{\mathrm{d}}{\mathrm{d}t} \log \langle O(x)O(0)\rangle_{\rho(t)}\right| \leq \frac{12\pi\Delta|x|}{c}\left(\langle T_{tt}\rangle - \langle T_{tt}\rangle_{\mathrm{vac}}\right)[1 + \ldots], \tag{10.11}$$

where dots indicate $\mathcal{O}\left(x^2 \langle T_{tt}\rangle / c\right)$ corrections. We furthermore prove a bound on the correlator itself. Letting $x > 0$, we have

$$\frac{\mathrm{d}}{\mathrm{d}x} \ln \langle O(x)O(0)\rangle_{\rho(t)} \leq \frac{\mathrm{d}}{\mathrm{d}x} \ln \langle O(x)O(0)\rangle_{\mathrm{vacuum}} = -\frac{2\Delta}{x}, \tag{10.12}$$

which shows that for the states covered by our assumptions, correlations between heavy scalars must die off faster than in the vacuum.

When $d = 3, 4$, we prove bounds on Wilson loops $\mathcal{W}(C)$ of spacelike circles C, assuming we can compute these using classical worldsheets in the bulk. Assuming $\mathcal{N} = 4$ SYM with gauge group $SU(N)$ and 't Hooft coupling λ on the boundary, we show that[5]

$$\left|\frac{\mathrm{d}}{\mathrm{d}t} \log \langle \mathcal{W}(C)\rangle_{\rho(t)}\right| \leq \mathrm{Length}[C]\sqrt{\frac{2\lambda}{3N^2}\langle T_{tt}\rangle}, \qquad d = 4. \tag{10.13}$$

In $d = 3$, we prove a similar result, but for the more restricted set of states which includes global quenches (see (10.132)). For small Wilson loops, we also have stricter bounds, which we give in the main text (see (10.127)).

[5]For other potential $d = 4$ holographic CFTs, our result can be written in terms of the effective central charge and effective coupling – see main text.

How are these bounds proven? Let us give the broad picture, restricting to the time-derivative of the entanglement entropy of a strip for concreteness. For CFTs dual to classical Einstein gravity, the von Neumann entropy of the reduced state ρ_R on a subregion R is given by the HRT formula [10, 11, 45], which says that

$$S_R = \frac{\text{Area}[X]}{4G_N}, \tag{10.14}$$

where X is the HRT surface in the gravitational bulk, which roughly means a codimension 2 spacelike surface that has stationary area under perturbations of X in the bulk interior. Bounding $\partial_t S_R$ in uniform states now corresponds to bounding $\partial_t \text{Area}[X_t]$, where X_t is a one-parameter family of HRT surfaces living in general time-dependent spacetimes with planar symmetry. Key to our proofs then is carrying out the analysis locally on a planar symmetric spatial slice Σ that contains X_t. We then show that the change in entanglement entropy is given by

$$\frac{\mathrm{d}S_R}{\mathrm{d}t} = \int_X GP, \tag{10.15}$$

where P is the matter momentum density in a direction orthogonal to HRT surface, and G essentially a propagator that only depends on the smallest radius probed by X, and not any other details of the spacetime. We thus see that the flux of matter falling out of the entanglement wedge is directly responsible for the increase of entanglement entropy. The formula (10.15) can be seen as momentum-entanglement correspondence, analogue to the momentum-complexity correspondence proposed in [46] and given a precise form in [47–49]. To further leverage this formula to get our proofs, we study two quasilocal masses and find in certain dimensions the integral in (10.15) is exactly encoded in the difference between these two quasilocal masses at infinity. A detailed analysis of the monotonicity properties of these masses under various flows then lets us prove our final bounds, essentially using a combination of Lorentzian and Riemannian inverse mean curvature flows. We emphasize that beyond our assumed symmetries, we do not need to assume a particular form of the spacetimes we are considering, and we are certainly not restricted to quenches for our most general bounds.

Along the way we derive various general properties of the HRT surfaces of strips and spheres in planar symmetric spacetimes. For example, for $d = 2$ we prove that the radius r_0 of the tip of the HRT surface of a strip of width ℓ satisfies

$$r_0 \geq \frac{2L^2}{\ell}, \tag{10.16}$$

where L the AdS radius. We prove similar bounds in higher dimensions. We also prove that the tip of an HRT surface of a sphere or a strip can never lie in a trapped region in spacetime. The same is shown for boundary anchored extremal surfaces of dimension $q + 1$ anchored at q-spheres.

This paper is organized as follows. In Sec. 10.2 we set up our assumptions and prove all our entanglement growth bounds for strip subregions R. In Sec. 10.3 we prove the entanglement growth bounds for ball shaped regions R and furthermore derive general properties $(q + 1)$-dimensional extremal surfaces anchored at q-dimensional spheres on the boundary, leading

to our results for correlators and Wilson loops. In Sec. 10.4 we prove bounds on spatial derivatives of the entanglement entropy of strips and equal-time two-point correlators in $d = 2$. In Sec. 10.5, for a subset of our bounds, we give significant numerical evidence that the dominant energy condition, which was assumed for our proofs, can be replaced by less restrictive assumptions. Finally, in Sec. 10.6, we conclude with a discussion of the implications of our findings, together with future directions. For a reader only wanting to understand the results without getting into the details of the proofs, it is possible to only read sections 10.2.1, 10.3.1, 10.4, 10.5, and 10.6.

10.2 Maximal Entanglement Rates for Strips

10.2.1 Setup and summary of results

Consider a d-dimensional holographic CFT in Minkowski space dual to classical Einstein gravity. Consider now some general time-evolving state $\rho(t)$ possessing a geometric dual, and having spatially homogeneous and isotropic one-point functions for local operators dual to bulk fields, such as the CFT stress tensor T_{ij}. Homogeneity and isotropy ensures that the dual asymptotically AdS_{d+1} spacetime $(\mathcal{M}, g)$ has planar symmetry. We allow $\rho(t)$ to live on either one or two copies of Minkowski space, so that the dual spacetime can have either one or two asymptotic boundaries. For a single system, we allow $\rho(t)$ to be mixed.[6]

Our goal in this section is to use the HRT entropy formula in this setup to derive a speed limit on the growth of the entanglement for a strip, and in some cases the union of any number of strips, provided they all live on the same connected component of the conformal boundary. In Sec. 10.3 we will generalize to spherical subregions, and to Wilson loops and two-point correlators. However, we will present the results on entanglement growth for spherical regions in this section, since they naturally are presented together with the results for strips.

Before presenting our results, let us set up our assumptions. We will assume that our spacetimes are AdS-hyperbolic, meaning that we can foliate $(\mathcal{M}, g_{ab})$ with spacelike hypersurfaces Σ_t that all have the same topology and are geodesically complete as Riemannian manifolds. These represent moments of time. Next, letting L be the asymptotic AdS radius, we assume that $(\mathcal{M}, g_{ab})$ satisfies the Einstein equations

$$R_{ab} - \frac{1}{2} g_{ab} R - \frac{d(d-1)}{2L^2} g_{ab} = 8\pi G_N \mathcal{T}_{ab}, \tag{10.17}$$

and that the dominant energy condition (DEC) holds for the bulk stress tensor $\mathcal{T}_{ab}$, meaning that

$$\mathcal{T}_{ab} u^a v^b \geq 0 \qquad \text{for all timelike } u^a, v^b. \tag{10.18}$$

Next, we assume that the Balasubramanian-Kraus [51] boundary stress tensor $\langle T_{ij} \rangle$ is finite. When it is finite, it corresponds to the one-point function of the CFT stress tensor. To specify falloff assumptions more explicitly, let Ω be any defining function, meaning any

[6] Allowing two-sided spacetimes means that automatically allow mixed states on a single CFT, since we can always find a purification dual to a wormhole, simply by gluing a second CPT-conjugate copy of the spacetime to itself along the HRT surface [50].

function on the conformal compactification of $\mathcal{M}$ such that the pullback of $\Omega^2 g_{ab}|_{\partial\mathcal{M}}$ to the conformal boundary is a Lorentzian metric. We then require that the bulk stress tensor satisfies

$$\mathcal{T}_{ab}u^a v^b \sim o\left(\Omega^d\right), \qquad \forall \text{ unit vectors } v^a, u^a, \tag{10.19}$$

near the conformal boundary $\partial\mathcal{M}$. In the radial coordinate r introduced below, this means the stress tensor in an orthonormal basis falls off as $o(r^{-d})$. Matter fields with falloffs sufficiently slow to require modifications of the definition of the spacetime mass are not covered by our results.[7] To avoid having to repeat the same assumptions in every theorem, let us define the following:

Definition 11. *We say that an $AAdS_{d+1}$ spacetime $(\mathcal{M}, g_{ab})$ is regular if it is AdS-hyperbolic, has falloffs (10.19), and g_{ab} is C^2.*

For index conventions, we will take $a, b, \ldots$ to be abstract spacetime indices, and $\alpha, \beta, \ldots$ to be abstract indices on spacelike hypersurfaces Σ. We take $\mu, \nu, \ldots$ to be coordinate indices on Σ. Other indices should be clear in the context. Furthermore, whenever intrinsic tensors on submanifolds are written with spacetime indices, we mean the pushforward/pullback to spacetime using the embedding map.

To describe the boundary regions covered by our results, we select a Minkowski conformal frame on the conformal boundary with coordinates

$$\mathrm{d}s^2|_{\partial\mathcal{M}} = -\mathrm{d}t^2 + L^2(\mathrm{d}\phi^2 + \mathrm{d}\boldsymbol{x}^2), \qquad \phi \in \mathbb{R}, \quad \boldsymbol{x} \in \mathbb{R}^{d-2}, \tag{10.20}$$

where the constant t-slices are the ones on which we have uniform one-point functions for local operators. For $d = 2$ we can allow ϕ to be periodically identified, in which case we say that $\mathcal{M}$ has spherical symmetry. If $\partial\mathcal{M}$ has two connected components, we focus on a particular one. We define $R_{t'}$ to be the one-parameter family of boundary regions given by

$$-\frac{\ell}{2L} \le \phi \le \frac{\ell}{2L}, \qquad t = t', \tag{10.21}$$

which just corresponds to a strip or interval of length ℓ at time t'. In this section, when we talk about strips or refer to a one-parameter family, we always mean the family (10.21). We will abbreviate $R_{t=0} \equiv R$, and define

$$\mathrm{Area}[\partial R_t] = \mathrm{Area}[\partial R] = L^{d-2}\int_{\mathbb{R}^{d-2}} \mathrm{d}^{d-2}\boldsymbol{x}, \qquad d > 2, \tag{10.22}$$

while for $d = 2$, we have $\mathrm{Area}[\partial R] = 2$. For $d > 2$ this is of course divergent, but since it always appears as an overall prefactor it causes no difficulties.

Next, the HRT formula [10, 11, 45] states that the von Neumann entropy of the reduced CFT state on R_t, $\rho_R(t) \equiv \mathrm{tr}_{R^c}\,\rho(t)$, is given by

$$S_R(t) = -\,\mathrm{tr}\left[\rho_R(t)\ln\rho_R(t)\right] = \frac{\mathrm{Area}[X_t]}{4G_N}, \tag{10.23}$$

[7] In this case, depending on how slow the falloffs are, subleading divergences in the entropy might become state dependent [40], in which case $\mathfrak{R} \le 1$ cannot remain true. See discussion in Sec. 10.6.

where X_t is the minimal codimension-2 spacelike surface in $(\mathcal{M}, g_{ab})$ that is (1) a stationary point of the area functional (i.e. extremal), (2) anchored at ∂R_t on the conformal boundary ($\partial X_t = \partial R_t$), and (3) homologous to R_t. The latter means that there exists spacelike hypersurface Σ with $\partial\Sigma = X_t \cup R_t$, where we here mean the boundary in the conformal completion. We will use the gravitational description to derive an upper bound on

$$\left|\frac{\mathrm{d}}{\mathrm{d}t}\left(\frac{\mathrm{Area}[X_t]}{4G_N}\right)\right| \tag{10.24}$$

purely in terms of quantities that have a known interpretation in the CFT. While $\mathrm{Area}[X_t]$ is formally divergent, since we (1) work with spacetimes with falloffs (10.19) and (2) $\mathrm{Area}[\partial R_t]$ is time-independent, (10.24) is in fact finite up to the $\mathrm{Area}[\partial R]$ prefactor.

Let us now summarize our main results, which are broadly divided into two categories. The first class of bounds scales like $\mathrm{Area}[\partial R]$, and they are strongest when R is large. The second class of bounds scales like $\mathrm{Vol}[R]$, and they are consequently the strongest for small subregions. For intermediate sized regions, where the entanglement entropy is about the enter the volume-scaling regime, we expect the two types of upper bounds to be roughly comparable.

First, for a three-dimensional bulk, we obtain the following

Theorem 23. *Let $(\mathcal{M}, g_{ab})$ be a regular asymptotically AdS_3 spacetime with planar or spherical symmetry satisfying the DEC. Assume that X_t is the HRT surface of a finite interval R_t. Then*

$$\left|\frac{\mathrm{d}}{\mathrm{d}t}\left(\frac{\mathrm{Area}[X_t]}{4G_N}\right)\right| \leq \sqrt{\frac{8\pi c}{3}\big(\langle T_{tt}\rangle - \langle T_{tt}\rangle_{\mathrm{vac}}\big)}, \tag{10.25}$$

where $c = \frac{3L}{2G_N}$.

Since the HRT surface of a union of strips is just the union of HRT surfaces of a collection of individual strips, this bound immediately implies that if R is a union of n intervals contained in a single moment of time on one of the connected components of $\partial\mathcal{M}$, then

$$\left|\frac{\mathrm{d}}{\mathrm{d}t}S_R\right| \leq n\sqrt{\frac{8\pi c}{3}\big(\langle T_{tt}\rangle - \langle T_{tt}\rangle_{\mathrm{vac}}\big)}. \tag{10.26}$$

While we are not able to give a general proof of the analogue of Theorem 23 in higher dimensions, we prove a generalization in thin-shell spacetimes:

Theorem 24. *Let $(\mathcal{M}, g_{ab})$ be an asymptotically $AdS_{d+1\geq 3}$ spacetime with planar symmetry satisfying the DEC. Assume that X_t is the HRT surface of a region R_t corresponding to either a finite width strip or a ball. Assume that the bulk matter consists of $U(1)$ gauge fields and a thin shell of matter:*

$$\mathcal{T}_{ab} = \mathcal{T}^{\mathrm{shell}}_{ab} + \mathcal{T}^{\mathrm{Maxwell}}_{ab}, \tag{10.27}$$

where $\mathcal{T}^{\mathrm{shell}}_{ab}$ has delta function support on a codimension-1 worldvolume that is timelike or null, and with $\mathcal{T}^{\mathrm{shell}}_{ab}$ separately satisfying the DEC. Assume $(\mathcal{M}, g_{ab})$ is regular, except we do not require g_{ab} to be C^2 at the shell. Then

$$\left|\frac{\mathrm{d}}{\mathrm{d}t}\left(\frac{\mathrm{Area}[X_t]}{4G_N}\right)\right| \leq \frac{1}{4}\mathrm{Area}[\partial R]c_{\mathrm{eff}}\left[\frac{16\pi}{(d-1)c_{\mathrm{eff}}}\langle T_{tt}\rangle\right]^{\frac{d-1}{d}}, \tag{10.28}$$

where $c_{\text{eff}} = L^{d-1}/G_N$.

This theorem applies to thin-shell Vaidya spacetimes and charged generalizations. These spacetimes (and related setups) have been studied extensively [31–34, 39, 41, 52–92] as holographic models CFT quenches. However, more general cases than Vaidya are allowed, where the shell might correspond to some brane in the bulk, propagating in a timelike direction. Using the duality between radius and scale in the CFT, thin shell spacetimes correspond to CFT states where all dynamics is happening at a single scale (that evolves with time). We also should note that (10.28) holds if R is a union of any number of strips on the same conformal boundary, due to the fact that the HRT surface of n strips is just equal to n HRT surfaces of n (generally different) strips. One can hope that this might also be true for multiple spheres, but this does not follow from our current analysis. Also, while we do not have a proof, we conjecture that (10.28) is valid in all DEC respecting regular planar symmetric AAdS_{d+1} spacetimes, and we provide strong numerical evidence for this in Sec. 10.5.

Also, note that c_{eff} can be defined purely in CFT in terms of a universal prefactor of the sphere vacuum entanglement entropy [10, 45], or in terms of the renormalized entanglement entropy [93, 94]. So our final bounds on $|\partial_t S|$ make no reference to the bulk.

The previous two results give upper bounds scaling like $\text{Area}[\partial R]$. Now let us turn to bounds scaling like $\text{Vol}[R]$. We prove the following bound on small regions, valid for all $d \geq 2$:

Theorem 25. *Let $(\mathcal{M}, g_{ab})$ be a regular asymptotically $AdS_{d+1\geq 3}$ spacetime with planar symmetry satisfying the DEC. Assume that X_t is the HRT surface of a region R_t corresponding to either a strip or a ball. Let ℓ be either the strip width or ball radius, and assume that*

$$\frac{\ell^d \langle T_{tt}\rangle}{c_{\text{eff}}} \ll 1. \tag{10.29}$$

Then

$$\left|\frac{\mathrm{d}}{\mathrm{d}t}\left(\frac{\text{Area}[X_t]}{4G_N}\right)\right| \leq \kappa_d \text{Vol}[R]\,\langle T_{tt}\rangle \left[1 + \mathcal{O}\left(\frac{\langle T_{tt}\rangle\,\ell^d}{c_{\text{eff}}}\right)\right]. \tag{10.30}$$

where

$$\kappa_d = \frac{\Gamma\left(\frac{1}{2(d-1)}\right)}{\Gamma\left(\frac{d}{2(d-1)}\right)} \begin{cases} 2\pi & R \text{ is an interval and } d = 2, \\ \frac{\sqrt{\pi}}{d-1} & R \text{ is a strip and } d > 2, \\ 2\sqrt{\pi} & R \text{ is a ball and } d > 2. \end{cases} \tag{10.31}$$

Next, for thin shell spacetimes, volume-type bounds can be proven exactly for subregions of any size, at the cost of a slightly larger prefactor:

Theorem 26. *Consider the same setup as in Theorem 24. Then*

$$\left|\frac{\mathrm{d}}{\mathrm{d}t}\left(\frac{\text{Area}[X_t]}{4G_N}\right)\right| \leq \kappa'_d \text{Vol}[R]\,\langle T_{tt}\rangle\,, \tag{10.32}$$

with

$$\kappa'_d = d^{-\frac{d}{2(d-1)}} \frac{\Gamma\left(\frac{1}{2(d-1)}\right)}{\Gamma\left(\frac{d}{2(d-1)}\right)} \begin{cases} 2\pi & R \text{ is an interval and } d = 2, \\ \sqrt{\frac{4\pi}{(d-1)}} & R \text{ is a strip and } d > 2, \\ \sqrt{16\pi(d-1)} & R \text{ is a ball and } d > 2. \end{cases} \tag{10.33}$$

For a strip, a few values of the prefactors are

$$\kappa_d = \begin{cases} 2\pi & d = 2 \\ 2.62\ldots & d = 3 \\ 2.43\ldots & d = 4 \\ 2 & d = \infty \end{cases}, \qquad \kappa'_d = \begin{cases} 2\pi & d = 2 \\ 3.25\ldots & d = 3 \\ 3.34\ldots & d = 4 \\ 4 & d = \infty. \end{cases} \tag{10.34}$$

We will now outline the strategy used to obtain these bounds. First, we observe that there exists exactly one homology hypersurface Σ_t that both contains X_t, and which respects the planar symmetry of $(\mathcal{M}, g_{ab})$. Then we show that the location of X_t on Σ_t can be solved for exactly in terms of the intrinsic geometry on Σ_t. Together with the DEC, this fact allows us to lower bound the radius of the tip of the HRT surface. Next, we use the fact that since X_t is extremal, the first order variation of its area is a pure boundary term located at ∂X_t [95], and we show that this boundary term is simply given by a particular component of the extrinsic curvature of Σ_t as $r \to \infty$. Then we work out the form of Einstein constraint equations on Σ_t, and show that the relevant extrinsic curvature component can be written as an integral of the matter flux over the HRT surface. Finally, essentially relying on inverse mean curvature flow of Lorentzian and Riemannian Hawking masses, and their monotonicity properties under these flows, we bound the integrated matter flux across the HRT surface from above in terms of the mass of the spacetime.

Now, before we dive in, we should clarify the meaning of radii in planar symmetric spacetimes. Since we have planar symmetry, spacetime has a two-parameter foliation where each leaf is a codimension-2 spacelike plane that has the usual flat intrinsic metric. When we talk about a plane, we always mean one of these leafs. These planes can all be assigned an "area radius" r, and it is possible to view r as a scalar function on spacetime which is not tied to any coordinate. Nevertheless, unlike in spherical symmetry, there is an overall scaling ambiguity in this function, since the non-compactness of the planes means we cannot normalize r to some area – there is no "unit plane". However, if we choose some Minkowski conformal frame on the boundary, we can fix the overall normalization of r by demanding that the defining function Ω that takes us to the chosen conformal frame is $\Omega = r/L$. We will implicitly assume such a choice, and refer to *the* radius of a plane.

10.2.2 An explicit solution for the HRT surface location

Without loss of generality, we will bound the time-derivative at $t = 0$ and use the shorthands $X_{t=0} = X$ and $R_{t=0} = R$. Since R is a strip contained in a canonical time slice of Minkowski, and since the ambient spacetime has planar symmetry, there exists a homology hypersurface $\hat{\Sigma}$ of X respecting the planar symmetry – see Figure 10.1. We can pick coordinates on $\hat{\Sigma}$ so that its induced metric $H_{\alpha\beta}$ reads

$$H_{\mu\nu}\mathrm{d}y^\mu \mathrm{d}y^\nu = B(r)\mathrm{d}r^2 + r^2(\mathrm{d}\phi^2 + \mathrm{d}\boldsymbol{x}^2), \quad r \in [r_0, \infty), \quad \phi \in [-\Phi(r), \Phi(r)], \tag{10.35}$$

where $\phi = \Phi(r)$ is the coordinate embedding function of (half of) the HRT surface in $\hat{\Sigma}$, as illustrated in Figure 10.1. r_0 is the smallest value of r probed by the HRT surface, corresponding to its tip. $\hat{\Sigma}$ can naturally be extended to include all $\phi \in \mathbb{R}$ by planar symmetry,

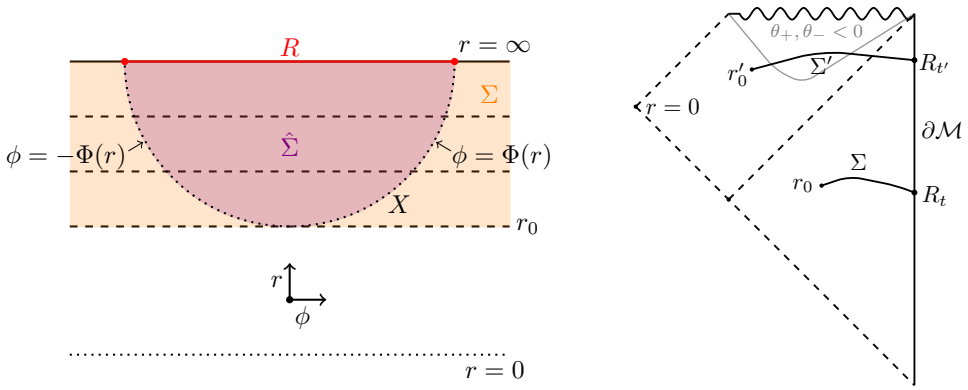


Figure 10.1: Left: the planar symmetric homology hypersurface $\hat{\Sigma}$ with respect to the HRT surface X. Σ is the extended homology hypersurface, whose boundary is the plane at $r = r_0$. Dashed lines are planes – i.e. constant r surfaces. Right: example conformal diagram indicating possible embeddings of two extended homology hypersurfaces Σ and Σ'. The grey line is an apparent horizon, with vanishing outwards null expansion, $\theta_+ = 0$.

and this choice turns out to be convenient for us. We denote the corresponding hypersurface as Σ, and refer to it as the extended homology hypersurface.[8] See Figure 10.1. The boundary of Σ (in the bulk proper) is a plane of radius r_0.

Relying on the formulas derived in the remainder of this section, we prove the following Lemma in appendix 10.7.5:

Lemma 12. *Let Σ be the extended homology hypersurface of an HRT surface X anchored at a strip region given by (10.21). Then a single coordinate system of the form*

$$\mathrm{d}s^2 = B(r)\mathrm{d}r^2 + r^2\mathrm{d}\boldsymbol{x}^2 \tag{10.36}$$

is enough to cover all of Σ. Furthermore, X has only one turning point, meaning that embedding function $r(\phi)$ is monotonically increasing for $\phi \geq 0$.

This means one function $\Phi(r)$ contains all the information about the embedding of X in Σ – we do not need multiple branches. It also means that Σ cannot have any locally stationary planes – that is – no planes of vanishing mean curvature, where $B(r)$ would blow up. The means we never need to worry about patching across coordinate systems when working on Σ. Geometrically, it implies that Σ has no "throats".

Now, taking $(r, \boldsymbol{x})$ to be coordinates on X, the induced metric on X reads

$$\mathrm{d}s^2|_X = \left[B(r) + r^2\Phi'(r)^2\right]\mathrm{d}r^2 + r^2\mathrm{d}\boldsymbol{x}^2. \tag{10.37}$$

Since X is an extremal surface, its area is stationary under all variations, including under variations within Σ. Enforcing this gives an ODE for $\Phi(r)$ in terms of $B(r)$. To find it, we

[8] $\hat{\Sigma}$ and Σ are unique. Planar symmetry means that $(\mathcal{M}, g_{ab})$ can be foliated by planes, and every point $p \in X$ lies in some plane in this foliation. Demanding planar symmetry of Σ requires that the full leaf intersected by p is included in Σ, and so we have a one-parameter family of codimension-2 surfaces picked out by X, which thus fully specifies Σ, and similarly for $\hat{\Sigma}$.

compute the mean curvature $\mathcal{K}$ of X viewed as a submanifold of Σ and demand it to be zero. This gives the equation (see appendix 10.7.1 for a computation)

$$rB\Phi'' + (d-1)r^2\left(\Phi'\right)^3 + \Phi'\left(dB - \frac{r}{2}B'\right) = 0. \tag{10.38}$$

The relevant boundary conditions are

$$\Phi'(r_0) = \infty, \quad \Phi(r_0) = 0, \tag{10.39}$$

where the former says that r_0 is the radius of the plane tangent to the tip of the HRT surface (i.e. where $\frac{\mathrm{d}r}{\mathrm{d}\phi} = 0$), while the latter implements that $\phi = 0$ corresponds to the center of the strip. It turns out that equation (10.38) can be integrated, and the solution with the correct boundary condition is

$$\Phi(r) = \int_{r_0}^{r} \mathrm{d}\rho \frac{\sqrt{B(\rho)}}{\rho\sqrt{(\rho/r_0)^{2d-2}-1}}. \tag{10.40}$$

This gives the location of the HRT surface within Σ explicitly in terms of the geometry of Σ. We now use this solution to determine the Einstein constraint equations on Σ, and to derive a formula for the rate of change of the entanglement growth.

10.2.3 A momentum-entanglement correspondence

Since X_t is extremal, its first order variation reduces to a pure boundary term given by (see for example the appendix of [95, 96]):

$$\frac{\mathrm{dArea}[X_t]}{\mathrm{d}t}\Big|_{t=0} = \int_{\partial X} N^a \eta_a, \tag{10.41}$$

where η^a is the translation vector generating the flow of ∂X_t at conformal infinity at $t = 0$, while N^a is the normal to ∂X that is also tangent to X, and that points towards the conformal boundary. In writing this formula, we implicitly assume that it is evaluated with some near-boundary cutoff that is subsequently removed. As is well known, given some choice of boundary conformal frame, a canonical choice of cutoff exists [97–99], which in our case reduces to a cutoff in the radial coordinate r. With a cutoff adapted to the Minkowski conformal frame and the falloffs (10.19), (10.41) is finite, even though $\mathrm{Area}[X_t]$ diverges.

Now we write (10.41) in a more useful form. We will give all the main steps, but relegate tedious but straight forward computations to the appendix.

Using the planar symmetry of Σ, the extrinsic curvature $K_{\alpha\beta}$ of Σ is given by

$$K_{\mu\nu}\mathrm{d}y^\mu \mathrm{d}y^\nu = K_{rr}(r)\mathrm{d}r^2 + K_{\phi\phi}(r)\left[\mathrm{d}\phi^2 + \mathrm{d}\boldsymbol{x}^2\right], \tag{10.42}$$

where we take the extrinsic curvature to be defined with respect to the future directed normal. Using this, we show in appendix (10.7.3), retracing the steps of [100], that

$$\frac{\mathrm{dArea}[X_t]}{\mathrm{d}t}\Big|_{t=0} = -\frac{\mathrm{Area}[\partial R]}{L^{d-2}} \lim_{r\to\infty} r^{d-3} K_{\phi\phi}. \tag{10.43}$$

Physically, $\lim_{r\to\infty} r^{d-3}K_{\phi\phi}$ measures the boost angle at which X hits the conformal boundary, or rather, the subleading part of the angle, since extremality implies that X hits $\partial\mathcal{M}$ orthogonally. This can be seen by studying extremal surfaces in a near-boundary expansion. Thus, we see that the entanglement growth is, up to a factor, identically given by the (subleading) boost angle at which the HRT surface hits the boundary. The same was found for maximal volume slices in [100].

Next we want to find a more explicit expression for $\lim_{r\to\infty} r^{d-3}K_{\phi\phi}$. To do this, we need to use the Einstein constraint equations, which read

$$\begin{aligned} \mathcal{R} + K^2 - K^{\alpha\beta}K_{\alpha\beta} + \frac{d(d-1)}{L^2} &= 16\pi G_N \mathcal{T}_{ab}t^a t^b, \\ D_\alpha K^\alpha{}_\beta - D_\beta K &= 8\pi G_N \mathcal{T}_{ab}t^a e^b_\beta, \end{aligned} \tag{10.44}$$

where $\mathcal{R}$ is the Ricci scalar of the metric on Σ, t^a the future unit normal to Σ, $K = H^{\alpha\beta}K_{\alpha\beta}$, and e^a_α a set of tangent vectors to Σ. To write these equations in coordinate form, it is convenient to introduce the function $\omega(r)$ as

$$B(r) = \frac{1}{\frac{r^2}{L^2} - \frac{\omega(r)}{r^{d-2}}}. \tag{10.45}$$

We will call $\omega(r)$ the Riemannian Hawking mass.[9] It will play a central role in our work. Whether or not $\omega(\infty)$ is proportional to the spacetime mass for some general spacelike hypersurface Σ depends on the behavior of the extrinsic curvature Σ at large r. It turns out that for $d \geq 3$, and with Σ being the extended homology hypersurface of an HRT surface, it has the property that it is proportional to the CFT energy density:

$$\langle T_{tt}\rangle = \frac{d-1}{16\pi G_N L^{d-1}}\omega(\infty), \qquad d \geq 3. \tag{10.46}$$

For $d = 2$, the right hand side is a lower bound on $\langle T_{tt}\rangle - \langle T_{tt}\rangle_{\text{vac}}$, where the vacuum energy must be subtracted when we allow ϕ to be periodic. We will explain these facts in Sec. 10.2.5.

It is also convenient to redefine $K_{rr}(r)$ in terms of a function $F(r)$ which is the rr-component of the extrinsic curvature in an orthonormal basis

$$K_{rr}(r) \equiv B(r)F(r). \tag{10.47}$$

In terms of these functions, the constraint equations in coordinate form read

$$(d-1)\frac{\omega'(r)}{r^{d-1}} = 2\mathcal{E}(r) - \frac{(d^2-3d+2)}{r^4}K_{\phi\phi}(r)^2 - \frac{2(d-1)}{r^2}F(r)K_{\phi\phi}(r), \tag{10.48}$$

$$K'_{\phi\phi}(r) - \frac{K_{\phi\phi}(r)}{r} = rF(r) - \frac{r^2}{d-1}\mathcal{J}(r), \tag{10.49}$$

where we introduced the notation

$$\begin{aligned} \mathcal{E} &= 8\pi G_N \mathcal{T}_{ab}t^a t^b, \\ \mathcal{J} &= 8\pi G_N \mathcal{T}_{ab}(\partial_r)^a t^b. \end{aligned} \tag{10.50}$$

[9] For $d = 3$ it is also known as the Geroch-Hawking mass [101–104], and its was used to prove the Riemannian Penrose inequality [104].

These are (proportional to) the energy density and radial momentum density of the matter with respect to the frame t^a. $\mathcal{J} > 0$ corresponds to matter falling into the bulk towards smaller r. From (10.19) and the fact that $B(r) \sim \mathcal{O}(r^{-1})$, we find that

$$\mathcal{E} \sim o\left(1/r^d\right), \qquad \mathcal{J} \sim o\left(1/r^{d+1}\right), \tag{10.51}$$

where we use that $\frac{1}{\sqrt{B}}(\partial_r)^a$ is a unit vector.

To turn (10.48) and (10.49) into a closed system, we will eliminate $F(r)$. We do this by imposing extremality of X in the direction of t^a. To do this, note that the inwards (outwards) null expansion θ_+ (θ_-) of X can be written as (see for example the appendix of [100])

$$\sqrt{2}\theta_\pm[X] = \pm\mathcal{K}[X] + K - n^\alpha n^\beta K_{\alpha\beta}, \tag{10.52}$$

where n^α is the outwards normal to X within Σ,[10] and where we remind that $\mathcal{K}[X]$ is the mean curvature of X within Σ. Extremality means $\theta_+ = \theta_- = 0$, which implies that $\mathcal{K} = 0$ and

$$K|_X = n^\alpha n^\beta K_{\alpha\beta}|_X. \tag{10.53}$$

This equation holds at $\phi = \Phi(r)$, which by planar symmetry means it holds everywhere on Σ. Writing out this equation in coordinates, carried out in appendix 10.7.2, we find

$$F + \frac{(d-2)K_{\phi\phi}}{r^2} + (d-1)K_{\phi\phi}\frac{(\Phi')^2}{B} = 0. \tag{10.54}$$

Plugging in the solution for $\Phi(r)$, given in (10.40), we get that

$$F(r) = -\frac{K_{\phi\phi}(r)}{r^2}\left(\frac{(d-2)r^{2d-2} + r_0^{2d-2}}{r^{2d-2} - r_0^{2d-2}}\right), \tag{10.55}$$

which upon insertion into the constraints, gives a closed system of ODEs

$$\frac{\omega'(r)}{r^{d-1}} = \frac{2}{d-1}\mathcal{E}(r) + \frac{K_{\phi\phi}^2}{r^4}h_1(r), \tag{10.56}$$

$$K'_{\phi\phi}(r) + \frac{K_{\phi\phi}}{r}h_2(r) = -\frac{r^2}{d-1}\mathcal{J}(r), \tag{10.57}$$

where

$$h_1(r) = \frac{(d-2)\left(r/r_0\right)^{2d-2} + d}{\left(r/r_0\right)^{2d-2} - 1}, \qquad h_2(r) = \frac{(d-3)\left(r/r_0\right)^{2d-2} + 2}{\left(r/r_0\right)^{2d-2} - 1}. \tag{10.58}$$

Now, $F(r)$ is a component of the extrinsic curvature in an orthonormal basis, so it must be finite at r_0. Using this to fix an integration constant, we find that the solutions of (10.56) and (10.57) are

$$K_{\phi\phi}(r) = -\frac{r^2}{(d-1)\sqrt{(r/r_0)^{2d-2} - 1}}\int_{r_0}^{r} \mathrm{d}\rho\, \mathcal{J}(\rho)\sqrt{(\rho/r_0)^{2d-2} - 1}, \tag{10.59}$$

$$\omega(r) = \omega(r_0) + \int_{r_0}^{r}\mathrm{d}\rho\left[\rho^{d-5}h_1(\rho)K_{\phi\phi}(\rho)^2 + \frac{2\rho^{d-1}}{d-1}\mathcal{E}(\rho)\right], \tag{10.60}$$

[10] We have here taken the outwards and inwards null vectors, k_+^a and k_-^a, respectively, to be $k_\pm^a = 2^{-1/2}(t^a \pm n^a)$.

where $K_{\phi\phi}(r_0) = 0$ since $\mathcal{J}(\rho)$ must be bounded.[11] Inserting (10.59) into (10.43) and multiplying by $(4G_N)^{-1}$, we get that

$$\left.\frac{\mathrm{d}S_R}{\mathrm{d}t}\right|_{t=0} = \frac{\mathrm{Area}[\partial R]}{4G_N L^{d-2}(d-1)} \int_{r_0}^{\infty} \mathrm{d}r \mathcal{J}(r) \sqrt{r^{2d-2} - r_0^{2d-2}}. \tag{10.61}$$

Since $\mathcal{J} > 0$ corresponds to a flux of energy density towards decreasing r, we see that matter falling out of the entanglement wedge and deeper into the bulk is directly responsible for the increase of entanglement. Conversely, outgoing matter is responsible for decrease in entanglement. We can also rewrite this formula in a covariant way. In appendix 10.7.3 we show that

$$\left.\frac{\mathrm{d}S_R}{\mathrm{d}t}\right|_{t=0} = \int_X G\mathcal{T}_{ab} n^a t^a \tag{10.62}$$

where n^a is the outwards unit normal to X that is tangent to Σ, and

$$G(r) = \frac{2\pi r^d}{(d-1) r_0^{d-1}}. \tag{10.63}$$

The formulas (10.43), (10.59)–(10.63) are the main results of this section. These results together with the theorems proven in the following section are crucial pieces to our proven bounds.

10.2.4 Geometric constraints on the HRT surface

In this section, we prove the following

Theorem 27. *Let $(\mathcal{M}, g_{ab})$ be a regular asymptotically $AdS_{d+1\geq 3}$ spacetime with planar symmetry satisfying the DEC. Let X be the HRT surface of a strip R of width ℓ, and let be r_0 be the smallest radius probed by X. Then*

$$\frac{L^2}{r_0} \leq \frac{\Gamma\left(\frac{1}{2(d-1)}\right)}{2\sqrt{\pi}\Gamma\left(\frac{d}{2(d-1)}\right)} \ell. \tag{10.64}$$

Furthermore, if $r_{0,\mathrm{vac}}$ is the smallest radius probed by the HRT surface X_0 of a strip of width ℓ in pure AdS_{d+1}, then

$$r_0 \geq r_{0,\mathrm{vac}}. \tag{10.65}$$

We now give the proof assuming that $\omega(r_0) \geq 0$, and then we will spend most of the rest of this section proving this assertion.

Proof. By Lemma 14, proven below, we have that $\omega(r_0) \geq 0$. Furthermore, the DEC implies that $\mathcal{E}$ is positive. Hence, (10.60) gives that $\omega(r)$ is everywhere positive. But this means that

$$B(r) = \frac{1}{\frac{r^2}{L^2} - \frac{\omega(r)}{r^{d-2}}} \geq \frac{L^2}{r^2}, \tag{10.66}$$

[11] For thin-shell spacetimes $\mathcal{J}(\rho)$ can be a delta function, but we can safely assume this delta function does not have support exactly at $r = r_0$.

which allows us to lower bound the strip width as follows:

$$\begin{aligned}\ell &= 2L\Phi(\infty) = 2L\int_{r_0}^{\infty} \mathrm{d}\rho \frac{\sqrt{B(\rho)}}{\rho\sqrt{(\rho/r_0)^{2d-2}-1}} \\ &\geq 2L^2 \int_{r_0}^{\infty} \mathrm{d}\rho \frac{1}{\rho^2\sqrt{(\rho/r_0)^{2d-2}-1}} \\ &= \frac{2L^2}{r_0}\frac{\sqrt{\pi}\Gamma\left(\frac{d}{2(d-1)}\right)}{\Gamma\left(\frac{1}{2(d-1)}\right)}.\end{aligned} \tag{10.67}$$

Finally, if we are in pure AdS, we must have have that the spacetime mass is vanishing, implying that $\omega(\infty) = 0$, and so by $\omega'(r) \geq 0$ and the fact that $\omega(r_0) \geq 0$, we must have $\omega(r) = 0$ everywhere. But that means that the above inequalities become equalities, giving $\frac{L^2}{r_0} \geq \frac{L^2}{r_{0,\mathrm{vac}}}$, which implies (10.65). □

Now we turn to proving that $\omega(r_0)$ is non-negative. The crucial tool is a planar-symmetric AdS_{d+1} version of the Lorentzian Hawking mass [105], which we define for a planar surface σ as

$$\mu[\sigma] = \frac{r^d}{L^2} - \frac{2r^d\theta_+\theta_-}{k_+\cdot k_-(d-1)^2}, \tag{10.68}$$

where k^+ and k^- are the outwards and inwards null vectors orthogonal to σ, respectively, and $\theta_\pm$ the corresponding null expansions. In [106], generalizing the results of [107] to planar symmetry and AAdS_{d+1} spacetimes, it was shown that the DEC implies that $\mu[\sigma]$ is monotonically non-decreasing when σ is moving in an outwards spacelike direction, provided we are in a normal region of spacetime, meaning that $\theta_+ \geq 0, \theta_- \leq 0$ when we take k^a_+ and k^a_- to be future directed.[12]

Furthermore, it is useful to rewrite the Riemannian Hawking mass $\omega(r)$ in a different way. ω can be thought of as a function of a planar surface σ together with a hypersurface Σ containing it, and in [100] it is shown that we can write ω as

$$\omega[\sigma,\Sigma] = \frac{r^d}{L^2} - \frac{\mathcal{K}[\sigma]^2}{(d-1)^2}, \tag{10.69}$$

where $\mathcal{K}$ is the mean curvature of σ in Σ. Using (10.52), which assumes the normalization $k_+\cdot k_- = -1$, we see that $2\theta_+\theta_- = (K - n^\alpha n^\beta K_{\alpha\beta})^2 - \mathcal{K}^2$, and so we get the following relation between the Hawking masses

$$\mu[\sigma] = \omega[\sigma,\Sigma] + \frac{r^d}{(d-1)^2}(K - n^\alpha n^\beta K_{\alpha\beta})^2. \tag{10.70}$$

With this in hand, we prove the following Lemma.

[12] This monotonicity is a planar-symmetric Lorentzian version of the monotonicity the Riemannian Hawking mass under inverse mean curvature flow, which has been used to prove Riemannian Penrose inequalities [101–104]. A Lorentzian flow with a monotonic Lorentzian Hawking mass for compact surfaces in three dimensions, without any symmetry assumptions, was studied in [108].

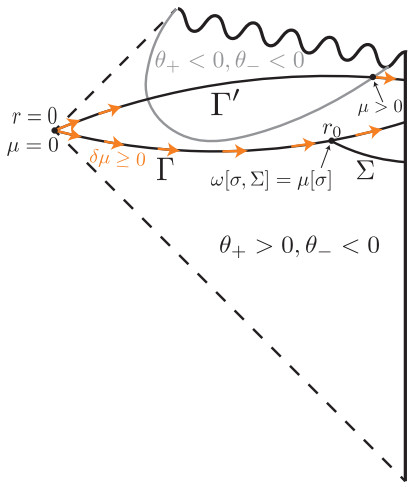


Figure 10.2: Example of two complete hypersurfaces Γ and Γ'. The Lorentzian Hawking mass is vanishing at $r = 0$ and positive at marginally trapped surfaces, given by the planes contained in the gray line. μ is monotonically non-decreasing along spacelike outwards flows in the untrapped region, where $\theta_+ \geq 0, \theta_- \leq 0$. At the boundary σ of the extended homology hypersurface Σ, the Riemannian Hawking mass ω with respect to Σ agrees with the Lorentzian Hawking mass μ.

Lemma 13. *Let Γ be a complete planar symmetric hypersurface with one conformal boundary. Let σ_r be a one-parameter family of planes in Γ with radius r, and with $r \in (0, \epsilon]$ for any $\epsilon > 0$. Then*

$$\lim_{r \to 0} \mu[\sigma_r] \geq 0. \tag{10.71}$$

Proof. Let us pick coordinates

$$\mathrm{d}s^2|_\Gamma = \left[\frac{r^2}{L^2} - \frac{\omega(r)}{r^{d-2}}\right]^{-1} \mathrm{d}r^2 + r^2 \mathrm{d}\boldsymbol{x}^2 \tag{10.72}$$

on Γ in a neighborhood of $r = 0$. Since Γ is complete and we only have one conformal boundary, arbitrarily small r must be part of Γ. Since Γ is spacelike, we must have $\omega(r) \leq r^d/L^2$, which means that $\omega(r) \sim \mathcal{O}(r^d)$ at small r. Now, from (10.70) we see that $\mu[\sigma_r] \geq \omega[\sigma_r, \Gamma]$ and so

$$\mu[\sigma_r] \geq \mathcal{O}(r^d). \tag{10.73}$$

Taking $r \to 0$ proves our assertion. □

Now we are ready to prove that $\omega(r_0) \geq 0$, together with the fact that the tip of the HRT surface cannot lie in a trapped region.

Lemma 14. *Let $(\mathcal{M}, g_{ab})$ be a planar-symmetric regular asymptotically AdS_{d+1} spacetime. Let X be the HRT surface of a strip. Then the tip of X lies in an untrapped region of spacetime, meaning the future null expansions of the plane σ tangent to X at the tip satisfies*

$$\theta_+[\sigma] \geq 0, \quad \theta_-[\sigma] \leq 0. \tag{10.74}$$

Furthermore, if the DEC holds and $(\mathcal{M}, g)$ is regular, the Riemannian Hawking mass of σ is non-negative:

$$\omega[\sigma, \Sigma] = \omega(r_0) \geq 0. \tag{10.75}$$

Proof. Let Σ be the unique planar symmetric extended homology hypersurface containing X. Let σ be the boundary of Σ in the bulk, having radius r_0. Its null expansion is

$$\sqrt{2}\theta_{\pm}[\sigma] = \pm\mathcal{K}[\sigma] + K - r^{\alpha}r^{\beta}K_{\alpha\beta}, \tag{10.76}$$

where $r^{\alpha} = \frac{1}{\sqrt{B}}(\partial_r)^{\alpha}$. An explicit computation gives

$$\begin{aligned} \mathcal{K}[\sigma] &= \frac{d-1}{r_0\sqrt{B(r_0)}}, \\ K &= \frac{1}{B}K_{rr} + \frac{K_{\phi\phi}(d-1)}{r^2}, \end{aligned} \tag{10.77}$$

and so we find

$$\sqrt{2}\theta_{\pm}[\sigma] = \pm\frac{d-1}{r_0\sqrt{B(r_0)}} - \frac{K_{\phi\phi}(r_0)(d-1)}{r_0^2}. \tag{10.78}$$

From (10.59) we have that $K_{\phi\phi}(r_0) = 0$, and so we get that

$$\pm\theta_{\pm} \geq 0, \tag{10.79}$$

proving the first assertion.

Next, since $K_{\phi\phi}(r_0) = 0$ we see that $2\theta_+\theta_-|_{\sigma} = -\mathcal{K}^2|_{\sigma}$, implying that $\mu[\sigma] = \omega[\sigma, \Sigma]$. Now, since our spacetime is AdS-hyperbolic, we can embed σ in a complete hypersurface with planar symmetry Γ, see Figure 10.2. Since σ lies in an untrapped region of spacetime, and since Γ is spacelike, $\mu[\sigma]$ is monotonically non-increasing as we deform σ in inwards along Γ while preserving its planar symmetry. Since the g_{ab} is C^2, $\theta_{\pm}$ are continuous, and so as we deform σ inwards, one of two things happen. Either we hit a marginally trapped surface, where $\theta_+\theta_- = 0$ and where μ is manifestly positive, or we approach $r = 0$, where we again have that μ is non-negative by Lemma 13. See Figure 10.2. But since μ is non-increasing along this deformation, and since it ends up somewhere non-negative, we must have $\mu[\sigma] \geq 0$. But $\mu[\sigma] = \omega[\sigma, \Sigma]$, completing the proof. □

We have illustrated the fact that the tip cannot lie in a trapped region of spacetime in Figure 10.1 – the tip cannot lie behind the gray line. Note that the proof of this fact does not rely on the DEC. This result improves on the findings of [109] in the special case where we have planar symmetry. In [109], they showed without any symmetry assumptions that the tip of an HRT surface in a $(2+1)$–dimensional spacetime can never lie in the so-called umbral region, which is a special subset of the trapped region that lies behind regular holographic screens [109, 110]. They also showed this result with planar symmetry in all dimensions. Here we extend this result to show that the whole trapped region is forbidden, although our result is more limited in that it always requires planar symmetry and a strip (or spherical) boundary region. Note also that this result does not forbid X to probe inside trapped regions – it is only the tip that is forbidden to lie there (see Figure 10.1). For example, for early times after a quench, the HRT surface will have portions threading through the trapped region [33, 34].

10.2.5 Proofs

Proof of $d = 2$ bound

We are now ready to prove Theorem 23. Evaluating the Lorentzian Hawking mass on a sphere at large r in a planar symmetric AAdS_{d+1} spacetime with falloffs (10.19), we get that

$$\langle T_{tt}\rangle = \frac{d-1}{16\pi G_N L^{d-1}}\mu(\infty). \tag{10.80}$$

This is valid also for $d = 2$, except if ϕ is periodically identified, we must replace the left hand side with $\langle T_{tt}\rangle - \langle T_{tt}\rangle_{\text{vac}}$. It can be seen to be true by evaluating $\mu(\infty)$ near the boundary in the usual Fefferman-Graham expansion [97–99]. Now, from (10.70) and (10.77) we have that

$$\mu(r) = \omega(r) + r^{d-4}K_{\phi\phi}(r)^2. \tag{10.81}$$

From (10.59), we see that $K_{\phi\phi}$ has asymptotic falloff $K_{\phi\phi} \sim \mathcal{O}(r^{3-d})$. Thus, we get that for $d \geq 3$, $\mu(\infty) = \omega(\infty)$, while for $d = 2$, we have

$$\mu(\infty) = \omega(\infty) + \left(\lim_{r\to\infty} r^{-1}K_{\phi\phi}\right)^2. \tag{10.82}$$

Since $\omega(\infty) \geq 0$ by the DEC, when $d = 2$ we obtain

$$\left|\lim_{r\to\infty} r^{-1}K_{\phi\phi}\right| \leq \sqrt{\mu(\infty)} = \sqrt{16\pi G_N L\,\langle T_{tt}\rangle}. \tag{10.83}$$

Using that $\mathrm{Area}[\partial R_t] = 2$, and combining (10.83) and (10.43) then yields

$$\left|\frac{\mathrm{d}S_R}{\mathrm{d}t}\right|_{t=0} \leq \frac{1}{2G_N}\sqrt{16\pi G_N L\,\langle T_{tt}\rangle} = \sqrt{\frac{8\pi c}{3}\,\langle T_{tt}\rangle}, \tag{10.84}$$

where we used the known Brown-Henneaux expression for the central charge: $c = \frac{3L}{2G_N}$ [111]. This proves Theorem 23.

Proof of bound for small ℓ

Now let us consider the result for small subregions, given by Theorem 25. The following Lemma is what we need:

Lemma 15. *Let $(\mathcal{M}, g_{ab})$ be a regular asymptotically $AdS_{d+1\geq 3}$ spacetime with planar symmetry satisfying the DEC. Let X be the HRT surface of a strip R of width ℓ, and let be r_0 be the smallest radius probed by X. Assume that*

$$\frac{\ell^d\,\langle T_{tt}\rangle}{c_{\text{eff}}} \ll 1. \tag{10.85}$$

Then

$$\left|\lim_{r\to\infty} r^{d-3}K_{\phi\phi}\right| \leq \frac{L}{2r_0}\omega(\infty)\left[1 + \mathcal{O}\left(\frac{\ell^d\,\langle T_{tt}\rangle}{c_{\text{eff}}}\right)\right]. \tag{10.86}$$

Proof. Let us for convenience define $W = -\lim_{r\to\infty} r^{d-3} K_{\phi\phi}$, and assume without loss of generality that $W > 0$ (otherwise, just reverse the time direction). Using the solutions (10.59) and (10.60), we have that

$$\begin{aligned}\frac{W}{\omega(\infty)} &= \frac{1}{d-1}\frac{\int_{r_0}^{\infty} \mathrm{d}r \mathcal{J}(r)\sqrt{r^{2d-2} - r_0^{2d-2}}}{\omega(r_0) + \int_{r_0}^{\infty} \mathrm{d}r \left[r^{d-5} h_1(r) K_{\phi\phi}(r)^2 + \frac{2r^{d-1}}{d-1}\mathcal{E}(r)\right]} \\ &\leq \frac{\int_{r_0}^{\infty} \mathrm{d}r r^{d-1}\mathcal{J}(r)}{2\int_{r_0}^{\infty} \mathrm{d}r r^{d-1}\mathcal{E}(\rho)}.\end{aligned} \tag{10.87}$$

The DEC requires that

$$0 \leq 8\pi G_N \mathcal{T}_{ab}\left[t^a \pm \frac{1}{\sqrt{B}}(\partial_r)^a\right] t^b = \mathcal{E} \pm \frac{1}{\sqrt{B}}\mathcal{J}, \tag{10.88}$$

and so we have that

$$\mathcal{E} \geq \frac{1}{\sqrt{B}}|\mathcal{J}|. \tag{10.89}$$

Writing B in terms of ω, and enforcing the DEC, we get

$$\frac{W}{\omega(\infty)} \leq \frac{\int_{r_0}^{\infty} \mathrm{d}r r^{d-1}\mathcal{J}(r)}{\frac{2}{L}\int_{r_0}^{\infty} \mathrm{d}r r^d \sqrt{1 - \frac{\omega(r)L^2}{r^d}}|\mathcal{J}(r)|}. \tag{10.90}$$

Let us now for a moment assume that we are perturbatively close to the vacuum, where ϵ is a perturbative parameter parametrizing the magnitude of $\omega(\infty)$. By monotonicity and positivity of $\omega(r)$, $\omega(r) \sim \mathcal{O}(\epsilon)$ as well, and so the $\omega(r)$ appearing in the square root gives higher order contributions:

$$\begin{aligned}\frac{W}{\omega(\infty)} &\leq \frac{\int_{r_0}^{\infty} \mathrm{d}r r^{d-1}|\mathcal{J}(r)|}{\frac{2}{L}\left[\int_{r_0}^{\infty} \mathrm{d}r r^d |\mathcal{J}(r)| - \frac{L^2}{2}\int_{r_0}^{\infty} \mathrm{d}r \omega(r)|\mathcal{J}(r)| + \ldots\right]} \\ &= \frac{L}{2}\frac{\int_{r_0}^{\infty} \mathrm{d}r r^{d-1}|\mathcal{J}(r)|}{\int_{r_0}^{\infty} \mathrm{d}r r^d |\mathcal{J}(r)|}\left[1 + \frac{L^2}{2}\frac{\int_{r_0}^{\infty} \mathrm{d}r \omega(r)|\mathcal{J}|}{\int_{r_0}^{\infty} \mathrm{d}r r^d |\mathcal{J}|} + \ldots\right] \\ &\leq \frac{L}{2r_0}\left[1 + \frac{L^2}{r_0^d}\omega(\infty) + \ldots\right] \\ &\leq \frac{L}{2r_0}\left[1 + \frac{L^{2d}}{r_0^d}\frac{\mu(\infty)}{L^{2d-2}} + \ldots\right] \\ &= \frac{L}{2r_0}\left[1 + \frac{16\pi\eta_d}{d-1}\frac{\ell^d \langle T_{tt}\rangle}{c_{\text{eff}}} + \ldots\right]\end{aligned} \tag{10.91}$$

where η_d is the $O(1)$ number coming from using (10.64). We see that the effective expansion parameter is the dimensionless quantity $\frac{\ell^d \langle T_{tt}\rangle}{c_{\text{eff}}}$. So the expansion is not really in small mass, which is dimensionful, but in small strip width relative to the inverse energy density per CFT degree of freedom. □

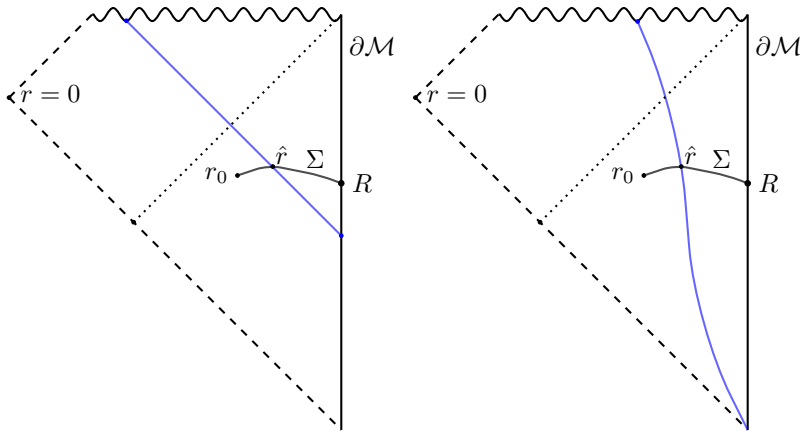


Figure 10.3: Examples of thin-shell spacetimes, where the blue lines correspond to the shells. The left space is dual to a uniform quench, where matter is thrown in from the boundary, while the right is a spacetime with a brane in the bulk interior.

From (10.43) and (10.64), we get, up to the perturbative corrections,

$$\begin{aligned}\left|\frac{\mathrm{d}S_R}{\mathrm{d}t}\right| &\leq \frac{\mathrm{Area}[\partial R]}{4G_N L^{d-2}}\frac{L}{2r_0}\omega(\infty) \leq \frac{\sqrt{\pi}}{d-1}\frac{\Gamma\left(\frac{1}{2(d-1)}\right)}{\Gamma\left(\frac{d}{2(d-1)}\right)}\ell\mathrm{Area}[\partial R]\,\langle T_{tt}\rangle \\ &= \mathrm{Vol}[R]\,\langle T_{tt}\rangle \begin{cases} 2\pi & d=2 \\ \frac{\sqrt{\pi}}{d-1}\frac{\Gamma\left(\frac{1}{2(d-1)}\right)}{\Gamma\left(\frac{d}{2(d-1)}\right)} & d>2 \end{cases},\end{aligned} \tag{10.92}$$

where we used (10.64) and (10.46) in the second inequality. This also holds for $d=2$, since $\omega(\infty)\leq\mu(\infty)$, provided we replace $\langle T_{tt}\rangle \to \langle T_{tt}\rangle - \langle T_{tt}\rangle_{\mathrm{vac}}$ if ϕ is compact. Also, note that for $d=2$ we have that $\ell\mathrm{Area}[\partial R] = 2\mathrm{Vol}[R]$. This completes the proof of Theorem 25 for strip regions.

Proof of bounds in thin-shell spacetimes

We now turn our attention to thin-shell spacetimes, where we will be able to establish that a bound of the form $|\partial_t S| \leq \#\mathrm{Vol}[R]\,\langle T_{tt}\rangle$ holds for any ℓ. Furthermore, in this class of spacetimes we will prove our conjectured generalization of Theorem 23 to $d>2$, i.e. Theorem 24.

Consider a spacetime where the matter consists of a single thin shell of matter that separately satisfies the DEC, together with a possible contribution from any number of $U(1)$ gauge fields:

$$\begin{aligned}\mathcal{E} &= \kappa\delta(r-\hat{r}) + \mathcal{E}^{\mathrm{Maxwell}}, \\ \mathcal{J} &= \eta\delta(r-\hat{r}),\end{aligned} \tag{10.93}$$

for some $\eta,\kappa,\hat{r}>r_0$. See Figure 10.3. Here we used that in planar symmetry, Maxwell fields give no contribution to the radial momentum density $\mathcal{J}$ (see for example Sec. 3 of [100]). In fact, we can add to the $U(1)$ gauge fields any matter that has a positive contribution to $\mathcal{E}$ but no contribution to $\mathcal{J}$.

The DEC, through (10.89), imposes that $\mathcal{J}$ only can have support at $\hat{r}$. Without loss of generality, we take $\eta > 0$. Let us in this section also use our scaling freedom in r to set $r_0 = L$ and choice of units to set $L = 1$.

Define again $W = -\lim_{r\to\infty} r^{d-3} K_{\phi\phi}$. Plugging (10.93) into (10.59), the solution for $K_{\phi\phi}$ is

$$K_{\phi\phi}(r) = -\frac{r^2}{d-1}\eta\sqrt{\frac{\hat{r}^{2d-2}-1}{r^{2d-2}-1}}\theta(r-\hat{r}), \tag{10.94}$$

and so

$$\eta = \frac{(d-1)W}{\sqrt{\hat{r}^{2d-2}-1}}, \tag{10.95}$$

which gives

$$K_{\phi\phi}(r) = -\frac{r^2 W}{\sqrt{r^{2d-2}-1}}\theta(r-\hat{r}). \tag{10.96}$$

Next, let us solve for the contribution to $\omega(r)$ from the squared extrinsic curvature term in (10.60):

$$\begin{aligned} Q(r) &\equiv \int_1^r \mathrm{d}\rho \rho^{d-5} K_{\phi\phi}(\rho)^2 h_1(\rho) \\ &= \theta(r-\hat{r}) W^2 \int_{\hat{r}}^r \mathrm{d}\rho \rho^{d-1} \frac{h_1(\rho)}{[\rho^{2d-2}-1]} \\ &= W^2\theta(r-\hat{r})\left[\frac{\hat{r}^d}{\hat{r}^{2d-2}-1} - \frac{r^d}{r^{2d-2}-1}\right]. \end{aligned} \tag{10.97}$$

To proceed, we need to understand what happens to ω as we cross the shock. Restricting attention to a small neighborhood of $\hat{r}$, where we can treat explicit occurrences of r not appearing in delta functions as constant, the equation for ω reads

$$(d-1)\frac{\omega'(r)}{\hat{r}^{d-1}} = 2\mathcal{E}^{\text{shell}} + \ldots, \tag{10.98}$$

where the terms indicated with dots will make no contribution to the discontinuity. Remembering that the DEC implies that $\sqrt{B}\mathcal{E} \geq |\mathcal{J}|$, imposing the DEC on the shell means that

$$\mathcal{E}^{\text{shell}} \geq \sqrt{\hat{r}^2 - \frac{\omega(r)}{\hat{r}^{d-2}}}\eta\delta(r-\hat{r}). \tag{10.99}$$

Inserting (10.99) into (10.98), dividing by the prefactor of the delta function, and integrating from $\hat{r}-\varepsilon$ to $\hat{r}+\varepsilon$ for some small positive ε, we find

$$\sqrt{\hat{r}^2 - \frac{\omega_-}{\hat{r}^{d-2}}} - \sqrt{\hat{r}^2 - \frac{\omega_+}{\hat{r}^{d-2}}} \geq \frac{\hat{r}}{d-1}\eta + \mathcal{O}(\varepsilon), \tag{10.100}$$

where we defined $\omega_\pm = \omega(\hat{r}\pm\varepsilon)$. We only have a sensible solution when $B(r)$ is real and positive everywhere, which requires

$$\frac{1}{d-1}\eta \leq \sqrt{1 - \frac{\omega_-}{\hat{r}^d}}. \tag{10.101}$$

Solving for ω_- from (10.100) and inserting our expression for η, we get that

$$\omega_+ \geq \omega_- + \frac{\hat{r}^d}{\sqrt{\hat{r}^{2d-2}-1}} W \left[\sqrt{1 - \frac{\omega_-}{\hat{r}^d}} - \frac{W}{\sqrt{\hat{r}^{2d-2}-1}} \right]. \tag{10.102}$$

Using this and (10.81), the Lorentzian Hawking mass at infinity has the lower bound

$$\begin{aligned} \mu(\infty) &= \omega(\infty) + \delta_{d2} W^2 \\ &\geq \omega_- + \frac{\hat{r}^d}{\sqrt{\hat{r}^{2d-2}-1}} W \left[\sqrt{1 - \frac{\omega_-}{\hat{r}^d}} - \frac{W}{\sqrt{\hat{r}^{2d-2}-1}} \right] + Q(\infty) + \delta_{d2} W^2 \\ &= \omega_- + \frac{\hat{r}^d}{\sqrt{\hat{r}^{2d-2}-1}} W \sqrt{1 - \frac{\omega_-}{\hat{r}^d}}, \end{aligned} \tag{10.103}$$

where δ_{ij} is the Kronecker delta. Thus, for any real n, we have that

$$\frac{W^n}{\mu(\infty)} \leq \frac{W^n}{\omega_- + \frac{\hat{r}^d}{\sqrt{\hat{r}^{2d-2}-1}} W \sqrt{1 - \frac{\omega_-}{\hat{r}^d}}} \equiv U_n, \tag{10.104}$$

together with the constraints

$$0 \leq \omega_- \leq \hat{r}^d, \tag{10.105}$$

$$W \leq \sqrt{\hat{r}^{2d-2}-1} \sqrt{1 - \frac{\omega_-}{\hat{r}^d}}. \tag{10.106}$$

Our goal will now be to upper bound U_n for all legal triplets $(W, \hat{r}, \omega_-)$ for $n = 1$ and $n = \frac{d}{d-1}$, which turns out to be values that will give interesting growth bounds.

Note first that we have

$$\partial_{\omega_-}^2 U_n \geq 0, \tag{10.107}$$

so any local extremum of U_n with respect to ω_- is a minimum. Thus, for any given W and $\hat{r}$, U_n is maximized when ω_- is on the boundary of its domain. First, take $\omega_- = 0$. Then, assuming that $1 \leq n \leq \frac{d}{d-1}$,

$$U_n = \frac{W^{n-1}\sqrt{\hat{r}^{2d-2}-1}}{\hat{r}^d} \leq \frac{\left[\hat{r}^{2d-2}-1\right]^{\frac{n}{2}}}{\hat{r}^d} \leq \frac{\hat{r}^{n(d-1)}}{\hat{r}^d} \leq 1, \tag{10.108}$$

where we used (10.106) in the second inequality. For $n = 1$, we get the stronger bound

$$U_1 \leq \frac{\sqrt{\hat{r}^{2d-2}-1}}{\hat{r}^d} \leq \sqrt{\frac{d-1}{d^{\frac{d}{d-1}}}} \equiv \alpha_d, \tag{10.109}$$

where the upper bound is found by maximizing with respect to $\hat{r}$. Next, let us look at the maximal value for ω_-, where we have the equality

$$W = \sqrt{\hat{r}^{2d-2}-1} \sqrt{1 - \frac{\omega_-}{\hat{r}^d}}. \tag{10.110}$$

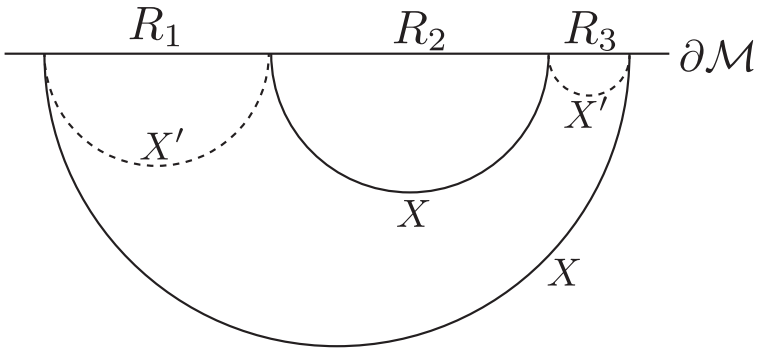


Figure 10.4: Possible HRT surfaces X and X' of the region $R_1 \cup R_3$, projected onto a timeslice.

Neglecting the first ω_- in the denominator of U_n and using $W \leq \sqrt{\hat{r}^{2d-2}-1}$, we get

$$U_n \leq \frac{\left[\hat{r}^{2d-2}-1\right]^{n/2}}{\hat{r}^d}. \tag{10.111}$$

But this is just the expression bounded earlier, and so (10.108) and (10.109) holds generally. Restoring factors of L, r_0, we have the following true bounds

$$W \leq L^{\frac{d-2}{d}}\omega(\infty)^{\frac{d-1}{d}}, \tag{10.112}$$

$$W \leq \alpha_d \frac{L}{r_0}\omega(\infty). \tag{10.113}$$

Redoing the steps in (10.92) with the numerical factor from in (10.113), we get Theorem 26 for strip regions. Next, inserting (10.112) into (10.43), we find

$$\left|\frac{\mathrm{d}S_R}{\mathrm{d}t}\right| \leq \frac{\mathrm{Area}[\partial R]}{4G_N L^{d-2}} L^{\frac{d-2}{d}} \left[\frac{16\pi G_N L^{d-1}}{d-1}\langle T_{tt}\rangle\right]^{\frac{d-1}{d}} = \frac{1}{4}\mathrm{Area}[\partial R] c_{\mathrm{eff}} \left[\frac{16\pi}{(d-1)c_{\mathrm{eff}}}\langle T_{tt}\rangle\right]^{\frac{d-1}{d}}, \tag{10.114}$$

proving Theorem 24 for strip regions.

10.2.6 Multiple strips and mutual information

Our results not scaling with Vol$[R]$ can be generalized to regions R consisting of n disjoint finite strips by simply applying the same argument to each connected component of the HRT surface separately. For $d = 2$ this results in

$$\left|\frac{\mathrm{d}S_R}{\mathrm{d}t}\right| \leq n\sqrt{\frac{8\pi c}{3}\langle T_{tt}\rangle}. \tag{10.115}$$

It is easy to see that (10.114) also holds true for n strips. No modification is needed, since Area$[\partial R]$ implicitly contains the factor of n present in the $d = 2$ case.

For the bounds scaling like volume, the behavior is different, since the upper bound depends on the connectivity properties of the entanglement wedge. Consider for example

$d = 2$ and the three intervals R_1, R_2, R_3 in Figure 10.4, and let $R = R_1 \cup R_3$ be the region under consideration. We then see that

$$\left|\frac{\mathrm{d}S_R}{\mathrm{d}t}\right| \leq \kappa \langle T_{tt}\rangle \begin{cases} \mathrm{Vol}[R] & \text{the entanglement wedge is disconnected,} \\ \mathrm{Vol}[R] + 2\mathrm{Vol}[R_2] & \text{the entanglement wedge is connected,} \end{cases} \tag{10.116}$$

where κ is the relevant numerical prefactor of either Theorem 25 or 26. We get this result by adding the different volume factors from each connected component of the HRT surface. Similar games can be played for n strips in d dimensions.

Next, lets us consider $d = 2$ and the mutual information between two subsystems R_1 and R_2 consisting of n_1 and n_2 finite intervals, respectively. We then have

$$\begin{aligned} \Big|\partial_t I(R_1, R_2)\Big| &= |\partial_t S_{R_1} + \partial_t S_{R_2} - \partial_t S_{R_1 R_2}| \leq |\partial_t S_{R_1}| + |\partial_t S_{R_2}| + |\partial_t S_{R_1 R_2}| \\ &\leq (2n_1 + 2n_2)\sqrt{\frac{8\pi c}{3}\langle T_{tt}\rangle}. \end{aligned} \tag{10.117}$$

Using (10.28), the generalization to higher d is obvious.

10.3 Maximal Rates for Balls, Wilson Loops and Correlators

10.3.1 Setup and summary of results

In this section we will consider extremal surfaces X_t of dimension $q+1$ anchored at q-dimensional spheres ∂R_t at time t on the conformal boundary, where extremal means that all the mean curvatures of X_t are zero.[13] We take the spheres to have radius $\mathcal{R}$. For $q = 0$, ∂R_t just consists of two points, and X_t is a one-parameter family of geodesics. For $q = d-2$, X_t is a one-parameter family of HRT surfaces anchored at spheres. For $q = 1$, X_t are two-dimensional spacelike worldsheets anchored at circles.

As before we are working with planar symmetric spacetimes, subject to the same assumptions described in Sec. 10.2.1. The logical steps will be mostly identical to Sec. 10.2, but with extra technicalities coming from the curvature of ∂R_t. Note that since we now have submanifolds of varying dimensions, we will use the symbol $\|\cdot\|$ to indicate the measure of the surface in the natural induced volume form. For quantities on the conformal boundary, $\|\cdot\|$ means with respect to the induced metric from the Minkowski conformal frame. We will use Length[], Area[] and Vol[] to refer to the measure of surfaces of dimension 1, codimension 2, and codimension 1, respectively.

To describe the relevant subregions in our results, let $\boldsymbol{z}$ be Cartesian coordinates in the direction transverse to the sphere ∂R_t. We now choose coordinates for our Minkowski conformal frame on the boundary to be

$$\mathrm{d}s^2 = -\mathrm{d}t^2 + L^2\left(\mathrm{d}\phi^2 + \phi^2 \mathrm{d}\Omega_q^2 + \mathrm{d}\boldsymbol{z}^2\right), \tag{10.118}$$

[13] This means that the generalized volume of X_t is stationary under perturbations with compact support.

with $\mathrm{d}\Omega_q^2$ the metric of a round unit q-sphere, and with the constant$-t$ slices the ones on which one-point functions of local operators are constant. For $q = 0$ there is no $\mathrm{d}\Omega_q^2$–term, while for $q = d - 2$ there is no $\mathrm{d}\boldsymbol{z}^2$ term. ϕ is a dimensionless radial coordinate on the boundary, and $R_{t'}$ is given by

$$0 \leq \phi \leq \frac{\mathcal{R}}{L}, \qquad t = t', \qquad \boldsymbol{z} = 0. \tag{10.119}$$

We now have

$$\|\partial R_t\| = \|\partial R\| = \Omega_q \mathcal{R}^q, \tag{10.120}$$

where Ω_q is the volume of a unit q-sphere.

Let us now summarize the results proven in this section. For entanglement entropy, we will prove the parts Theorems 24, 25 and 26 that refer to spherical ∂R. For extremal surfaces of other dimensionalities, the following theorem applies to the most general class of spacetimes and subregions:

Theorem 28. *Let $(\mathcal{M}, g_{ab})$ be a regular asymptotically AdS_{d+1} spacetime with planar symmetry satisfying the DEC. Assume that d is even, and let be X_t be an extremal surface of dimension $d/2$, anchored on the conformal boundary at the sphere ∂R_t. Then*

$$\left|\frac{\mathrm{d}}{\mathrm{d}t}\|X_t\|\right| \leq \|\partial R\| L^{\frac{d}{2}} \sqrt{\frac{16\pi}{c_{\text{eff}}(d-1)} \langle T_{tt}\rangle}. \tag{10.121}$$

Of course, for $d = 2$, this just reduces to Theorem 23. For $d = 4$ this can be converted to the growth bound on circular Wilson loops, given by (10.13). Next, for surfaces X_t anchored at small spheres on the boundary, we get the following:

Theorem 29. *Let $(\mathcal{M}, g_{ab})$ be a regular asymptotically AdS_{d+1} spacetime with planar symmetry satisfying the DEC. Let be X_t be an extremal surface of dimension $q + 1$, anchored on the conformal boundary at the sphere ∂R_t having radius $\mathcal{R}$. Assume that*

$$q \geq \frac{d-2}{2} \tag{10.122}$$

and

$$\frac{\mathcal{R}^d \langle T_{tt}\rangle}{c_{\text{eff}}} \ll 1. \tag{10.123}$$

Then

$$\left|\frac{\mathrm{d}}{\mathrm{d}t}\|X_t\|\right| \leq \eta_{d,q} \|\partial R\| L^{q+1} \mathcal{R}^{d-q-1} \frac{\langle T_{tt}\rangle}{c_{\text{eff}}} \left[1 + \mathcal{O}\left(\frac{\mathcal{R}^d \langle T_{tt}\rangle}{c_{\text{eff}}}\right)\right], \tag{10.124}$$

where

$$\eta_{d,q} = \frac{8\pi}{d-1} \left[\frac{\Gamma\left(\frac{1}{2(q+1)}\right)}{\sqrt{\pi}\Gamma\left(\frac{q+2}{2(q+1)}\right)}\right]^{d-q-1}. \tag{10.125}$$

Using well known dictionary entries, described in Sec. 10.3.6, this converts to growth bounds on the entanglement of small balls, small circular Wilson loops, and heavy two-point functions at small separations. Specifically, for the latter two, we get

$$\left|\frac{\mathrm{d}}{\mathrm{d}t}\log|\left\langle O(x)O(0)\right\rangle_{\rho(t)}|\right| \leq \frac{8\pi\Delta}{c_{\text{eff}}}|x|\left\langle T_{tt}\right\rangle\left[1+\mathcal{O}\left(\frac{|x|^2\left\langle T_{tt}\right\rangle}{c_{\text{eff}}}\right)\right], \quad d=2, \tag{10.126}$$

and

$$\left|\frac{\mathrm{d}}{\mathrm{d}t}\log|\left\langle \mathcal{W}(C)\right\rangle_{\rho(t)}|\right| \leq \frac{8\pi\sqrt{\lambda_{\text{eff}}}}{(d-1)c_{\text{eff}}}\eta_{d,1}\mathcal{R}^{d-1}\left\langle T_{tt}\right\rangle\left[1+\mathcal{O}\left(\frac{\mathcal{R}^d\left\langle T_{tt}\right\rangle}{c_{\text{eff}}}\right)\right], \quad d\in\{3,4\}, \tag{10.127}$$

where $\sqrt{\lambda_{\text{eff}}}=L^2/\ell_{\text{string}}^2$ is the effective 't Hooft coupling, and ℓ_{string} the bulk string length. Finally, for thin-shell spacetimes, we prove the following:

Theorem 30. *Let $(\mathcal{M},g_{ab})$ be an asymptotically AdS_{d+1} spacetime with planar symmetry satisfying the DEC. Assume that X_t is an extremal surface anchored at a boundary sphere of dimension*

$$q\geq\frac{d-2}{2}. \tag{10.128}$$

Next, assume that the bulk matter consists of $U(1)$ gauge fields and a thin shell of matter:

$$\mathcal{T}_{ab}=\mathcal{T}_{ab}^{\text{shell}}+\mathcal{T}_{ab}^{\text{Maxwell}}, \tag{10.129}$$

where $\mathcal{T}_{ab}^{\text{shell}}$ has delta function support on a codimension-1 worldvolume that is timelike or null, and that separately satisfies the DEC. Assume $(\mathcal{M},g)$ is regular, except we do not require g_{ab} to be C^2 at the shell. Then

$$\left|\frac{\mathrm{d}}{\mathrm{d}t}\|X_t\|\right| \leq \|\partial R\|L^{q+1}\left[\frac{16\pi}{(d-1)c_{\text{eff}}}\left\langle T_{tt}\right\rangle\right]^{\frac{q+1}{d}} \tag{10.130}$$

and

$$\left|\frac{\mathrm{d}}{\mathrm{d}t}\|X_t\|\right| \leq \kappa_{d,q}\|\partial R\|L^{q+1}\mathcal{R}^{d-q-1}\frac{\left\langle T_{tt}\right\rangle}{c_{\text{eff}}} \tag{10.131}$$

with $\kappa_{d,q}$ given by (10.192).

The main application of (10.130) is to bound Wilson loops in $d=3$, where we get

$$\left|\frac{\mathrm{d}}{\mathrm{d}t}\log|\left\langle \mathcal{W}(C)\right\rangle_{\rho(t)}|\right| \leq \frac{\sqrt{\lambda_{\text{eff}}}\text{Length}[C]}{2\pi}\left[\frac{8\pi}{c_{\text{eff}}}\left\langle T_{tt}\right\rangle\right]^{2/3}. \tag{10.132}$$

Let us now turn to the proofs.

10.3.2 An implicit solution for the extremal surface location

As earlier, let Σ be the extended planar symmetric homology hypersurface containing X. For the exact same reason as earlier, there is a unique choice of Σ. We can now pick coordinates on Σ given by

$$\mathrm{d}s^2|_\Sigma=H_{\mu\nu}\mathrm{d}y^\mu\mathrm{d}y^\nu=B(r)\mathrm{d}r^2+r^2\left(\mathrm{d}\phi^2+\phi^2\mathrm{d}\Omega_q^2+\mathrm{d}\boldsymbol{z}^2\right). \tag{10.133}$$

Again, one such coordinate system covers all of Σ, as shown in appendix 10.7.5.

We take our intrinsic coordinates on X to be (r, Ω^i), where Ω^i are coordinates on the sphere. The embedding coordinates of X in Σ reads

$$X^\mu = (r, \phi = \Phi(r), \Omega^i, \boldsymbol{z} = 0), \tag{10.134}$$

where the symmetries of the problem dictate $\boldsymbol{z} = 0$. The induced metric on X is

$$\mathrm{d}s^2|_X = \left[B(r) + r^2\Phi'(r)^2\right]\mathrm{d}r^2 + r^2\Phi(r)^2\mathrm{d}\Omega_q^2. \tag{10.135}$$

Now we must implement the condition that X is extremal, which requires us to compute all its mean curvatures and demand them to be vanishing. To do this, let n_a^I be an orthonormal basis of normal forms to X that are tangent to Σ, labeled by I. Let t_a be the future timelike normal orthogonal to Σ. A complete basis of mean curvatures of X now is

$$\begin{aligned} \mathcal{K}^I &= h^{ab}\nabla_a n_b^I, \\ \mathcal{K}^0 &= h^{ab}\nabla_a t_b, \end{aligned} \tag{10.136}$$

where $h^{ab} = g^{ab} + t^a t^b - \delta^{IJ} n_I^a n_J^b$. All of these quantities must vanish. Considering the $\mathcal{K}^I$ corresponding the $\boldsymbol{z}$ directions, we just get 0 by our symmetries. Letting $I = n$ denote the remaining normal direction in Σ, we get by direct computation that (see appendix (10.7.6) for some of the required ingredients)

$$\begin{aligned} \mathcal{K}^n = \frac{1}{r\sqrt{B}\left[B + r^2(\Phi')^2\right]^{3/2}}\Bigg[& r^2 B\Phi'' + (q+1)r^3(\Phi')^3 \\ & + \left((q+2)B - \frac{1}{2}rB'\right)r\Phi' - \frac{qB}{\Phi}\left(B + r^2(\Phi')^2\right)\Bigg]. \end{aligned} \tag{10.137}$$

If it was not for the last term, we would reproduce (10.38) by setting $q = d - 2$. The new term is caused by the curvature of ∂R. Now, with this last term, we no longer have an explicit analytical solution (when $q > 0$). However, we can find an implicit solution that lets us proceed. Define

$$\chi(r) = \frac{qB}{\Phi\Phi'}\left(\frac{B}{\Phi'^2} + r^2\right), \tag{10.138}$$

so that our equation for extremality reads

$$(q+1)r^3(\Phi')^3 + \left((q+2)B - \frac{1}{2}rB'\right)r\Phi' + r^2 B\Phi'' - (\Phi')^3\chi(r) = 0. \tag{10.139}$$

Imposing $\Phi(r_0) = 0$, where r_0 is the tip of the extremal surface, we have the implicit solution

$$\Phi(r) = \int_{r_0}^{r}\mathrm{d}\rho \frac{\sqrt{B(\rho)}}{\rho\sqrt{\mathcal{C}\rho^{2q+2}h(\rho) - 1}}, \tag{10.140}$$

where

$$h(r) = 1 - \frac{2}{\mathcal{C}}\int_{r_0}^{r}\mathrm{d}\rho\chi(\rho)\rho^{-6-2q}, \tag{10.141}$$

for some $\mathcal{C}$ that is fixed by imposing $\Phi'(r_0) = \infty$. Assuming $h(r_0)$ is finite, we get that $\mathcal{C} = r_0^{-2q-2}$. This is indeed correct, even though $\chi(r_0)$ looks superficially divergent. Since we are near a minimum of $r(\Phi)$ we have that $r = r_0 + \mathcal{O}(\Phi^2)$ near r_0, and so for r close to r_0 we get $\Phi = \alpha\sqrt{r - r_0}$ for some α. Even though Φ goes to zero at r_0 we find that

$$\chi(r) \sim \mathcal{O}(1) \tag{10.142}$$

near r_0, and so $h(r_0) = 1$. Next, reality of $\Phi(r)$ demands that $h(\rho) \geq (r_0/r)^{2q+2}$, while positivity of Φ and Φ' ensures that $h(r) \leq 1$, and so in total we know that[14]

$$\Phi(r) = \int_{r_0}^{r} \mathrm{d}\rho \frac{\sqrt{B(\rho)}}{\rho\sqrt{(\rho/r_0)^{2q+2}h(\rho) - 1}}, \qquad 0 < (r_0/r)^{2q+2} < h(r) \leq 1. \tag{10.143}$$

10.3.3 The relation between the Hawking masses

Take $K_{\alpha\beta}$ to be the extrinsic curvature of the extended homology hypersurface. Like in the case of the strip, we have that

$$\mu(r) = \omega(r) + r^{d-4}K_{\phi\phi}(r)^2, \tag{10.144}$$

which follows from the same computation as in the previous section, together with (10.270) in appendix 10.7.6. We have that $K_{\phi\phi} \sim \mathcal{O}(1/r^{q-1})$, as becomes clear in the next section. Thus, at large r we have

$$\mu(r) = \omega(r) + \mathcal{O}\left(r^{d-2-2q}\right). \tag{10.145}$$

Consequently, $\omega(\infty)$ is proportional to spacetime mass if and only if

$$q > \frac{d-2}{2}. \tag{10.146}$$

If $2q = d - 2$, $\omega(\infty)$ is smaller than $\mu(\infty)$ by some finite number. For $2q < d - 2$, we get $\omega(\infty) = -\infty$ by (10.145) and the fact that $\mu(\infty)$ is finite and positive. We will see below that this comes out of the constraint equations, since exactly when $2q < d-2$, $\omega(r)$ is neither positive nor monotonically increasing. We will not be able to say anything about the case $2q < d - 2$.

10.3.4 $\partial_t\|X_t\| \leq$ momentum on X_t

The time-derivative of the generalized volume satisfies [95, 96]

$$\frac{\mathrm{d}}{\mathrm{d}t}\|X_t\| = \int_{\partial X_t} N^a \eta_a, \tag{10.147}$$

where $\eta^a = (\partial_t)^a$ generates the deformation of ∂X_t, while N^a is the unit vector that is (1) tangent to X_t, (2) orthogonal to ∂X_t, and (3) pointing towards the conformal boundary. A computation in appendix 10.7.4 shows that (10.147) can be written as

$$\frac{\mathrm{d}}{\mathrm{d}t}\|X_t\|\Big|_{t=0} = \frac{\|\partial R\|}{L^q} \lim_{r\to\infty} r^{q-1}K_{\phi\phi}(r). \tag{10.148}$$

[14] By the same kind of analysis as in appendix 10.7.5, we cannot have additional turning points where $\Phi'(r)$ diverges, and so we have strict inequality in the lower bound.

Now we again reach the stage where we must write the Einstein constraint equations as a closed system, which requires us to impose extremality in the timelike direction.

First, note that from the planar symmetry of Σ, if $\boldsymbol{x}$ are Cartesian coordinates on the plane containing R, then we have that the extrinsic curvature of Σ reads

$$\begin{aligned} K_{\mu\nu}\mathrm{d}y^\mu \mathrm{d}y^\nu &= K_{rr}(r)\mathrm{d}r^2 + K_{\phi\phi}(r)\left(\mathrm{d}\boldsymbol{x}^2 + \mathrm{d}\boldsymbol{z}^2\right) \\ &= K_{rr}(r)\mathrm{d}r^2 + K_{\phi\phi}(r)\left(\mathrm{d}\phi^2 + \phi^2\mathrm{d}\Omega_q^2 + \mathrm{d}\boldsymbol{z}^2\right). \end{aligned} \tag{10.149}$$

Thus, the components of the extrinsic curvature with indices in the sphere directions reads

$$K_{ij} = K_{\phi\phi}(r)\phi^2 w_{ij}, \tag{10.150}$$

where w_{ij} is the unit metric on the q-sphere. Define again $F(r)$ through the relation $K_{rr}(r) = F(r)B(r)$. Computing $\mathcal{K}^0 = 0$, using (10.150), and solving for $F(r)$ (see Appendix 10.7.6), we get, after substituting our expression for $\Phi(r)$, that

$$F(r) = -\frac{K_{\phi\phi}(r)}{r^2}H(r), \tag{10.151}$$

where we for convenience defined the function

$$H(r) = \frac{q(r/r_0)^{2q+2}h(r) + 1}{(r/r_0)^{2q+2}h(r) - 1}. \tag{10.152}$$

Since $\partial_h H < 0$ and $h(r) \leq 1$, we get the lower bound

$$H(r) \geq \frac{q(r/r_0)^{2q+2} + 1}{(r/r_0)^{2q+2} - 1} \equiv H_L(r). \tag{10.153}$$

The constraint equations (10.48) and (10.49) are unchanged, except now the expression for $F(r)$ is different. Plugging it in we get

$$(d-1)\frac{\omega'(r)}{r^{d-1}} = 2\mathcal{E} + \frac{(d-1)}{r^4}K_{\phi\phi}(r)^2\left[2H(r) - d + 2\right], \tag{10.154}$$

$$K'_{\phi\phi} + \left[H(r) - 1\right]\frac{K_{\phi\phi}}{r} = -\frac{r^2}{d-1}\mathcal{J}(r). \tag{10.155}$$

Now, using the lower bound $H_L(r)$, let us note the following:

$$2H(r) - d + 2 \geq \frac{(2q - d + 2)(r/r_0)^{2q+2} + d - 1}{(r/r_0)^{2q+2} - 1}. \tag{10.156}$$

This is positive definite for all r only when $q \geq \frac{d-2}{2}$, so for geodesics ($q = 0$), we only have monotonicity of $\omega(r)$ when $d = 2$. But this is just the case studied in the previous section. For ($q = 1$), which is the relevant case for Wilson loops, we have monotonicity of $\omega(r)$ only for $d \leq 4$. For an HRT surface we have $q = d - 2$, and so we have monotonicity in all dimensions. It is in fact quite surprising that we have monotonicity for any q whatsoever, since when looking at the Einstein constraint equations in covariant form, monotonicity of

the Riemannian Hawking mass is only manifest on hypersurfaces that have vanishing mean curvature.

Let us assume $2q \geq d-2$ going forward, and let us bound $K_{\phi\phi}$ and μ at infinity. Fixing an integration constant by demanding that $F(r_0) =$ finite, the solution to the momentum constraint is

$$K_{\phi\phi}(r) = -\frac{1}{d-1}\int_{r_0}^{r} \mathrm{d}\rho \rho^2 \mathcal{J}(\rho) e^{-\int_\rho^r \mathrm{d}\frac{1}{z}(H(z)-1)}. \tag{10.157}$$

We have that

$$\begin{aligned} |K_{\phi\phi}(r)| &\leq \frac{1}{d-1}\int_{r_0}^{r} \mathrm{d}\rho \rho^2 |\mathcal{J}(\rho)| e^{-\int_\rho^r \mathrm{d}\frac{1}{z}(H_L(z)-1)} \\ &= \frac{r^2}{(d-1)\sqrt{(r/r_0)^{2q+2}-1}}\int_{r_0}^{r} \mathrm{d}\rho |\mathcal{J}(\rho)| \sqrt{(\rho/r_0)^{2q+2}-1}. \end{aligned} \tag{10.158}$$

We see from this expression that $K_{\phi\phi} \sim \mathcal{O}(1/r^{q-1})$. Also, in this last expression, if we replace $|\mathcal{J}| \to -\mathcal{J}$, we just get the solution of (10.155) with $H(r)$ replaced by $H_L(r)$. We will use this fact later.

Inserting (10.158) in (10.148), we finally get

$$\left|\frac{\mathrm{d}}{\mathrm{d}t}\|X_t\|\right| \leq \frac{(d-1)\|\partial R\|}{L^q}\int_{r_0}^{\infty} \mathrm{d}\rho |\mathcal{J}(\rho)| \sqrt{\rho^{2q+2}-r_0^{2q+2}}. \tag{10.159}$$

Unlike for an HRT surface anchored at a strip, we are here only able to write an inequality.

Next, let us turn to the second ingredient: the mass. Rewriting (10.154) in terms of $\mu(r)$, we get

$$(d-1)\frac{\mu'(r)}{r^{d-1}} = 2\mathcal{E} + \frac{(d-1)}{r^4}K_{\phi\phi}(r)^2\left[2H(r)+d-6\right] + \frac{2(d-1)}{r^3}\frac{\mathrm{d}}{\mathrm{d}r}K_{\phi\phi}^2. \tag{10.160}$$

After an integration by parts and using $H(r) \geq H_L(r)$, we get that

$$\mu(\infty) \geq \mu(r_0) + \frac{2}{d-1}\int_{r_0}^{\infty} \mathrm{d}\rho \rho^{d-1}\mathcal{E}(\rho) + \int_{r_0}^{\infty} \mathrm{d}\rho \rho^{d-5} K_{\phi\phi}(r)^2\left[2H_L(r)+d\right] \tag{10.161}$$

where

$$2H_L(r)+d = \frac{(d+2q)(r/r_0)^{2q+2}+(d-2)}{(r/r_0)^{2q+2}-1} \geq 0. \tag{10.162}$$

Possessing now an upper bound on $\partial_t\|X_t\|$ and a lower bound on mass, we next need an upper bound on L^2/r_0.

10.3.5 Constraints on boundary anchored extremal surfaces

It turns out that generalizations of Lemmata 13 and 14 remain true for the surfaces considered in this section. The proof of Lemma 13 is unchanged, while from the discussion in appendix 10.7.5 and 10.7.6, together with the proof of Lemma 14, we get the following constraints on the tip of X:

Lemma 16. *Let $(\mathcal{M}, g_{ab})$ be an asymptotically $AdS_{d+1\geq 3}$ spacetime with planar symmetry. Let X be a $(q+1)$-dimensional extremal surface anchored at a q-sphere on the conformal boundary. Then the tip of X lies in an untrapped region. Furthermore, if $(\mathcal{M}, g_{ab})$ is regular and satisfies the DEC, then $\omega(r_0) \geq 0$, where r_0 is the radius of the tip of X.*

With this in hand, we readily obtain the spherical dimension–$(q+1)$ version of Theorem 27:

Theorem 31. *Let $(\mathcal{M}, g_{ab})$ be a regular planar-symmetric asymptotically $AdS_{d+1\geq 3}$ spacetime satisfying the DEC. Let X be a dimension $q+1$ extremal surface anchored at a sphere of radius $\mathcal{R}$. Let be r_0 be the radius of the plane tangent to the tip of X. Then if*

$$q \geq \frac{d-2}{2}, \tag{10.163}$$

we have

$$\frac{L^2}{r_0} \leq \frac{\Gamma\left(\frac{1}{2(q+1)}\right)}{\sqrt{\pi}\Gamma\left(\frac{q+2}{2(q+1)}\right)}\mathcal{R}. \tag{10.164}$$

Proof. For $2q \geq d-2$, (10.154) implies $\omega'(r) \geq 0$. Combining with $\omega(r_0) \geq 0$, we get $\omega(r) \geq 0$. Using now $h(r) < 1$ and that $\omega(r) \geq 0$ implies $B(r) \geq L/r$, we get

$$\begin{aligned}\mathcal{R} = L\Phi(\infty) &= L\int_{r_0}^{\infty} \mathrm{d}r \frac{\sqrt{B(r)}}{r\sqrt{(r/r_0)^{2q+2}h(r)-1}} \\ &\geq L^2 \int_{r_0}^{\infty} \mathrm{d}r \frac{1}{r^2\sqrt{(r/r_0)^{2q+2}-1}} = \frac{L^2}{r_0}\frac{\sqrt{\pi}\Gamma\left(\frac{q+2}{2(q+1)}\right)}{\Gamma\left(\frac{1}{2(q+1)}\right)}.\end{aligned} \tag{10.165}$$

□

10.3.6 Proofs

Proof of bounds in $d = 2$ and $d = 4$

Consider the special dimension

$$q = \frac{d-2}{2}, \tag{10.166}$$

which can only happen when d is even. As seen previously, we have that $\omega(r) \geq 0$ in this case. Furthermore, (10.144) becomes

$$\mu(\infty) = \omega(\infty) + \left[\lim_{r\to\infty} r^{q-1}K_{\phi\phi}\right]^2, \tag{10.167}$$

and so

$$\left|\lim_{r\to\infty} r^{q-1}K_{\phi\phi}(\infty)\right| \leq \sqrt{\mu(\infty)}, \tag{10.168}$$

which gives

$$\left|\frac{\mathrm{d}}{\mathrm{d}t}\|X_t\|\right|_{t=0} \leq \frac{\|\partial R\|}{L^q}\sqrt{\mu(\infty)} \quad = \|\partial R\| L^{q+1}\sqrt{\frac{16\pi}{c_{\text{eff}}(d-1)}\langle T_{tt}\rangle}. \tag{10.169}$$

This proves Theorem 28. For $q = 0, d = 2$, this is just the formula we used to derive (10.25).

The above result implies a bound on correlators that can be computed using the geodesic approximation. The geodesic approximation says that the two-point function of a CFT scalar operator O of large scaling dimension $\Delta \gg 1$ can be computed as

$$\langle O(\boldsymbol{x})O(0)\rangle_{\rho(t)} = \eta e^{-\frac{\Delta}{L}\|X_t\|_{\text{reg}}}, \tag{10.170}$$

where η is some constant, and $\|X_t\|_{\text{reg}}$ is the regularized distance of a geodesic anchored at the points $(t, \boldsymbol{x})$ and $(t, 0)$ on the conformal boundary. We here adopted the Schrödinger picture. Combining (10.169) and (10.170) and taking $d = 2$, we get[15]

$$\left|\frac{\mathrm{d}}{\mathrm{d}t}\log\langle O(x)O(0)\rangle_{\rho(t)}\right| \leq \sqrt{\frac{96\pi\Delta^2}{c}\langle T_{tt}\rangle}, \qquad d = 2. \tag{10.171}$$

Next, for $d = 4$, (10.169) holds for $q = 1$, where the X_t are two-dimensional worldsheets anchored at circles on the boundary. If $\mathcal{W}(C)$ is a Wilson loop of a circle $C = S^1$, we have that [112, 113]

$$\langle \mathcal{W}(C)\rangle_{\rho(t)} = \eta e^{-\frac{1}{2\pi\alpha'}\|X_t\|}, \tag{10.172}$$

where $\alpha' = \ell^2_{\text{string}}$ and η again some constant. Combining (10.169) and (10.172) we get

$$\left|\frac{\mathrm{d}}{\mathrm{d}t}\log\langle \mathcal{W}(C)\rangle_{\rho(t)}\right| \leq \text{Length}[C]\sqrt{\frac{4\lambda_{\text{eff}}}{3\pi c_{\text{eff}}}\langle T_{tt}\rangle}, \qquad d = 4, \tag{10.173}$$

where $\text{Length}[C] = 2\pi\mathcal{R}$. With the precise dictionary for the duality between type IIB supergravity on $\text{AdS}_5 \times S^5$ and $\mathcal{N} = 4$ super Yang-Mills with gauge group $SU(N)$ and 't Hooft coupling λ, given by

$$\frac{G_N}{R^3} = \frac{\pi}{2N^2}, \quad \sqrt{\lambda} = \frac{L^2}{\alpha'}, \tag{10.174}$$

(10.173) can be written as (10.13).

Proof of bounds for small $\mathcal{R}$

Next we prove bounds that are strong at small radii $\mathcal{R}$. We have:

Lemma 17. *Consider the same assumptions as in Theorem 31. Assume furthermore that* $\mathcal{R}^d \langle T_{tt}\rangle / c_{\text{eff}} \ll 1$ *and* $2q \geq d - 2$. *Then*

$$\left|\lim_{r\to\infty} r^{d-3}K_{\phi\phi}\right| \leq \frac{L}{2r_0^{d-q-1}}\omega(\infty)\left[1 + \mathcal{O}\left(\frac{\mathcal{R}^d\langle T_{tt}\rangle}{c_{\text{eff}}}\right)\right]. \tag{10.175}$$

[15] Of course, we could have derived this in Sec. 10.2, given that that geodesics coincide with the HRT surface when $d = 2$.

Proof. Define $W = -\lim_{r\to\infty} r^{q-1} K_{\phi\phi}$, and assume without loss of generality that $W > 0$. Using (10.59) and (10.154) and the exact same logic as in the proof of Lemma 15, we get

$$\begin{aligned}\frac{W}{\omega(\infty)} &\leq \frac{L\int_{r_0}^{\infty} \mathrm{d}r r^{q+1}|\mathcal{J}(r)|}{2\int_{r_0}^{\infty} \mathrm{d}r r^{d}\sqrt{1-\frac{\omega(r)L^2}{r^d}}|\mathcal{J}(r)|} \\ &\leq \frac{L\int_{r_0}^{\infty} \mathrm{d}r r^{q+1}|\mathcal{J}(r)|}{2\int_{r_0}^{\infty} \mathrm{d}r r^{d}|\mathcal{J}(r)|}\left[1+\frac{L^2\omega(\infty)}{2r_0^d}+\ldots\right] \\ &\leq \frac{L}{2r_0^{d-q-1}}\left[1+\mathcal{O}\left(\frac{\mathcal{R}^d\langle T_{tt}\rangle}{c_{\text{eff}}}\right)\right].\end{aligned} \tag{10.176}$$

□

Inserting now (10.175) into (10.148), we get

$$\begin{aligned}\left|\frac{\mathrm{d}}{\mathrm{d}t}\|X_t\|\right|_{t=0} &\leq \frac{8\pi\|\partial R\|L^{1+q}}{(d-1)c_{\text{eff}}}\frac{L^{2(d-q-1)}}{r_0^{d-q-1}}\langle T_{tt}\rangle \\ &\leq \frac{\|\partial R\|L^{1+q}}{c_{\text{eff}}}\eta_{d,q}\mathcal{R}^{d-q-1}\langle T_{tt}\rangle\end{aligned} \tag{10.177}$$

where

$$\eta_{d,q} = \frac{8\pi}{d-1}\left[\frac{\Gamma\left(\frac{1}{2(q+1)}\right)}{\sqrt{\pi}\Gamma\left(\frac{q+2}{2(q+1)}\right)}\right]^{d-q-1}. \tag{10.178}$$

We can convert this to bounds on two-point functions and circular Wilson loops. Combining (10.177) with (10.170) and (10.172), we get the bounds (10.126) and (10.127).

Finally, with $q = d-2$ and $d > 2$, the entanglement entropy of small spheres is bounded as

$$\left|\frac{\mathrm{d}S_R}{\mathrm{d}t}\right| \leq \frac{2\sqrt{\pi}\Gamma\left(\frac{1}{2(d-1)}\right)}{\Gamma\left(\frac{d}{2(d-1)}\right)}\text{Vol}[R]\langle T_{tt}\rangle + \ldots \tag{10.179}$$

where we used that $\text{Vol}[R] = \text{Area}[\partial R]\mathcal{R}/(d-1)$. The proves the part of Theorem 25 where ∂R is a sphere.

Proof of bounds for thin-shell spacetimes

Finally, let us prove our thin-shell results valid all for $\mathcal{R}$, assuming $2q \geq d-2$. Since we already have general bounds for two-point correlators in $d = 2$ and Wilson loops in $d = 4$, this section is mostly relevant for entanglement in medium or large balls in general d, and for medium and large Wilson loops in $d = 3$. We consider the same setup as in Sec. 10.2.5, and use the same notation. Again, we choose $r_0 = L = 1$.

Now, let us consider the solutions (10.154) and (10.155) with the replacement $H(r) \to H_L(r)$. As discussed in Sec. 10.3.4, this gives a smaller value for $\mu(\infty)$ and larger value for $\left|\lim r^{q-1}K_{\phi\phi}\right|$ if $\mathcal{J}$ has a fixed sign, which is the case here. Since we will consider bounds of

the form $\lim r^{q-1}K_{\phi\phi} \leq \#\mu(\infty)^n$ for $n > 0$, the bounds we obtain with this replacement will be valid for the original spacetime.

Now, with a delta function shock, solution for $K_{\phi\phi}$ reads

$$K_{\phi\phi}(r) = -\frac{r^2}{d-1}\eta\sqrt{\frac{\hat{r}^{2q+2}-1}{r^{2q+2}-1}}\theta(r-\hat{r}) = -\frac{r^2 W}{\sqrt{r^{2q+2}-1}}\theta(r-\hat{r}). \tag{10.180}$$

From (10.161) we get that the contribution to $\mu(\infty)$ from the extrinsic curvature reads

$$\begin{aligned} Q(\infty) &\equiv \int_{r_0}^{\infty} \mathrm{d}\rho \rho^{d-5} K_{\phi\phi}(r)^2 \frac{(d+2q)r^{2q+2}+(d-2)}{r^{2q+2}-1} \\ &= W^2 \frac{\hat{r}^d}{\hat{r}^{2q+2}-1}. \end{aligned} \tag{10.181}$$

The analysis of how the DEC changes across the shock is unchanged from the strip case, except for a few exponents, and we find

$$\omega_+ = \omega_- + \frac{\hat{r}^d}{\sqrt{\hat{r}^{2q+2}-1}} W\left[\sqrt{1-\frac{\omega_-}{\hat{r}^d}} - \frac{W}{\sqrt{\hat{r}^{2q+2}-1}}\right]. \tag{10.182}$$

By the same logic as in Sec. 10.2.5 we get,

$$\frac{W^n}{\mu(\infty)} \leq \frac{W^n}{\omega_- + W\frac{\hat{r}^d}{\sqrt{\hat{r}^{2q+2}-1}}\sqrt{1-\frac{\omega_-}{\hat{r}^d}}} \equiv U_n \tag{10.183}$$

where

$$0 \leq \omega_- \leq \hat{r}^d, \tag{10.184}$$

$$W \leq \sqrt{r^{2q+2}-1}\sqrt{1-\frac{\omega_-}{\hat{r}^d}}. \tag{10.185}$$

Again, it now suffices to take ω_- at the boundary of its allowed domain. With $\omega_- = 0$ and $1 \leq n \leq \frac{d}{q+1}$ we get

$$U_n = \frac{W^{n-1}\sqrt{\hat{r}^{2q-2}-1}}{\hat{r}^d} \leq \frac{[r^{2q+2}-1]^{n/2}}{\hat{r}^d} \leq 1, \tag{10.186}$$

and for $n = 1$ we get the stronger bound

$$U_1 \leq \sqrt{\frac{q+1}{d}}\left(\frac{d}{d-1-q}\right)^{\frac{q+1-d}{2(q+1)}} \equiv \alpha_{d,q}. \tag{10.187}$$

For saturation of (10.185), neglecting the first ω_- in the denominator of U_n and using that $W \leq \sqrt{\hat{r}^{2q+q}-1}$, we get

$$U_n \leq \frac{[r^{2q+2}-1]^{n/2}}{\hat{r}^d} \leq 1. \tag{10.188}$$

Restoring factors of L, r_0, we have the following general bounds

$$W \leq \alpha_{q,d} \frac{L}{r_0^{d-q-1}} \omega(\infty), \tag{10.189}$$

$$W \leq L^{\frac{2q+2-d}{d}} \omega(\infty)^{\frac{q+1}{d}}. \tag{10.190}$$

Combining (10.189) with (10.148) and (10.163) now gives that

$$\left| \frac{\mathrm{d}}{\mathrm{d}t} \|X_t\| \right| \leq \kappa_{d,q} \|\partial R\| L^{1+q} \mathcal{R}^{d-q-1} \frac{\langle T_{tt} \rangle}{c_{\text{eff}}}, \tag{10.191}$$

where

$$\kappa_{d,q} = 2\alpha_{d,q}\eta_{d,q} = \frac{16\pi}{d-1}\sqrt{\frac{q+1}{d}} \left(\frac{d}{d-1-q} \right)^{\frac{q+1-d}{2(q+1)}} \left[\frac{\Gamma\left(\frac{1}{2(q+1)}\right)}{\sqrt{\pi}\Gamma\left(\frac{q+2}{2(q+1)}\right)} \right]^{d-q-1}. \tag{10.192}$$

This shows that the type of bounds derived in the small $\mathcal{R}$ limit holds in thin shell spacetimes for all $\mathcal{R}$, at price of a larger prefactor. For the entanglement entropy of balls, we get

$$\left| \frac{\mathrm{d}S_R}{\mathrm{d}t} \right| \leq 4\sqrt{\frac{\pi(d-1)}{d^{\frac{d}{d-1}}}} \frac{\Gamma\left(\frac{1}{2(d-1)}\right)}{\Gamma\left(\frac{d}{2(d-1)}\right)} \mathrm{Vol}[R] \langle T_{tt} \rangle . \tag{10.193}$$

For Wilson loops we get

$$\left| \frac{\mathrm{d}}{\mathrm{d}t} \log |\langle \mathcal{W}(C) \rangle_{\rho(t)}| \right| \leq \frac{\sqrt{\lambda_{\text{eff}}}}{c_{\text{eff}}} \mathcal{R}^{d-1} \langle T_{tt} \rangle \begin{cases} \frac{\sqrt{128\pi}\Gamma(1/4)}{3^{3/4}\Gamma(3/4)} \approx 26 & d=3 \\ \frac{8\Gamma(1/4)^2}{3\Gamma(3/4)^2} \approx 23 & d=4 \end{cases}. \tag{10.194}$$

Next, consider (10.190). This gives us that

$$\left| \frac{\mathrm{d}}{\mathrm{d}t} \|X_t\| \right|_{t=0} \leq \|\partial R\| L^{q+1} \left[\frac{16\pi}{(d-1)c_{\text{eff}}} \langle T_{tt} \rangle \right]^{\frac{q+1}{d}} \tag{10.195}$$

For $q = d-2$, corresponding to entanglement for ball subregions, this just gives (10.28), verifying that it holds for spheres as well, completing the proof of the part of Theorem 26 concerning spherical ∂R.

For $q = 0, d = 2$ and $q = 1, d = 4$ we just reproduce the bounds of Sec. 10.3.6, which are anyway proven with weaker assumptions there. For $q = 1, d = 3$ we get a new bound on Wilson loops, given by (10.132).

10.4 Bounding Spatial Derivatives

The technology we have developed to bound time derivatives also lets us bound spatial derivatives of extremal surface areas for strips.

Consider a one parameter family of strips R_ℓ of variable width ℓ at some fixed boundary time, given by (10.21) with t' now held fixed. Let X_ℓ be the corresponding one-parameter family of HRT surfaces. A computation in appendix 10.7.3 gives that

$$\frac{\mathrm{d}}{\mathrm{d}\ell}\mathrm{Area}[X_\ell] = \frac{\mathrm{Area}[\partial R]}{L^{d-1}} r_0^{d-1}. \tag{10.196}$$

For a strip, we thus see that the depth of the HRT surface tip uniquely determines $\partial_\ell S$. Using our lower bound on r_0 given by (10.64), we now immediately get the following:

Theorem 32. *Let $(\mathcal{M}, g_{ab})$ be a regular asymptotically $AdS_{d+1\geq 3}$ spacetime with planar symmetry satisfying the DEC. If X_ℓ is the HRT surface of a strip R_ℓ of width ℓ, then*

$$\frac{\mathrm{d}}{\mathrm{d}\ell}\left[\frac{\mathrm{Area}[X_\ell]}{4G_N}\right] \geq \frac{c_{\mathrm{eff}}}{4\ell^{d-1}}\mathrm{Area}[\partial R]\left[\frac{2\sqrt{\pi}\Gamma\left(\frac{d}{2(d-1)}\right)}{\Gamma\left(\frac{1}{2(d-1)}\right)}\right]^{d-1}. \tag{10.197}$$

The lower bound is equal to $\partial_\ell S_{\mathrm{vacuum}}$, and so we get

$$\frac{\mathrm{d}}{\mathrm{d}\ell}\Delta S[\rho_{R_\ell}] \geq 0 \tag{10.198}$$

where ΔS is the vacuum subtracted entropy. Since we get the vacuum entanglement entropy in the limit $\ell \to 0$, this implies that

$$\Delta S \geq 0. \tag{10.199}$$

It is easy to see that (10.198) and (10.199) applies to a subregion R corresponding to a union of any number of finite width strips, with ∂_ℓ now interpreted as the derivative with respect to increasing width of one or more of the connected components.

For $d = 2$, we also get a bound on correlators of heavy scalar single trace primaries. Working at a fixed moment of time with a homogeneous state ρ, the combination of (10.196), (10.64) and (10.170) for $x > 0$ gives

$$\frac{\mathrm{d}}{\mathrm{d}x}\ln\langle\mathcal{O}(x)\mathcal{O}(0)\rangle_\rho \leq \frac{\mathrm{d}}{\mathrm{d}x}\ln\langle\mathcal{O}(x)\mathcal{O}(0)\rangle_{\mathrm{vacuum}} = -\frac{2\Delta}{x} \tag{10.200}$$

which means that correlations must die of faster than the vacuum for the states and operators covered by our assumptions. This in particular implies that

$$\langle\mathcal{O}(x)\mathcal{O}(0)\rangle_\rho \leq \langle\mathcal{O}(x)\mathcal{O}(0)\rangle_{\mathrm{vacuum}}, \tag{10.201}$$

since we just get the vacuum correlator as $x \to 0$.

10.5 Evidence for Broader Validity of Bounds

In the previous sections, we have shown that the DEC allows us to prove several general bounds on the growth of entanglement, correlators and Wilson loops. However, the proofs crucially relied on the dominant energy condition. While the dominant energy condition

holds in type IIA, IIB and eleven-dimensional supergravity (see for example the appendix of [114]), it is typically violated after dimensional reduction [115]. The prototypical example is a scalar field dual to a relevant CFT operator. This field has negative mass squared, leading to DEC violation.

Even though our proofs assumed the DEC, we will now provide strong evidence for a subset of the bounds that they hold when the DEC is violated in reasonable ways. That is, we provide evidence in scalar theories that violate the DEC, but which have proven positive mass theorems [116, 117] (so pure AdS is stable) and respect the null energy condition (NEC). This is evidence that our bounds are true even in CFTs with DEC-violating bulks, since the NEC and a stable vacuum are both necessary conditions for sensible bulk theories.[16]

In fact, we provide evidence not just that

$$\mathfrak{R} \leq 1, \tag{10.202}$$

but also that when $d > 2$,

$$\mathfrak{R} \leq v_E^{(\text{SAdS})} + \delta v_E < 1 \tag{10.203}$$

for some small δv_E that seems to depend on the scalar potential. Here

$$v_E^{(\text{SAdS})} = \sqrt{\frac{d}{d-2}} \left(\frac{d-2}{2(d-1)} \right)^{\frac{d-1}{d}} \tag{10.204}$$

is the entanglement velocity computed in a quench in holography, with the final state being neutral, dual to the AdS-Schwarzschild type black brane [33, 34].

The theories we will consider are neutral scalars minimally coupled to gravity,

$$8\pi G_N \mathcal{L} = \frac{1}{2}R - \frac{d(d-1)}{2L^2} - \frac{1}{2}|\mathrm{d}\phi|^2 - V(\phi), \tag{10.205}$$

where V is negative somewhere, leading to violation of the DEC (but not the NEC). These theories are common in consistent truncations and dimensional reductions of type IIA, IIB, and eleven-dimensional supergravity [118–121]. We consider these theories because, for standard forms of minimally coupled bosonic matter, neutral scalars appear to pose the biggest risk to our bounds. This is because gauge fields give no direct contribution to $\mathcal{J}$, and they have a manifestly positive contributions to the mass (they respect the DEC).

For free theories where $V = \frac{1}{2}m^2\phi^2$, in order to maximize the chance of violating our bounds, we choose potentials that are close to "maximally negative", meaning we pick m^2 just slightly above the Breitenlohner-Freedman [122, 123] bound:

$$m^2 L^2 \geq m_{\text{BF}}^2 L^2 \equiv -(d/2)^2. \tag{10.206}$$

It is known that if $m^2 < m_{\text{BF}}^2$, AdS is unstable, and so these theories cannot be dual to CFTs with a Hamiltonian that is bounded from below. Additionally, to have an example of an interacting potential, in $d = 3$ we consider a top down potential that becomes exponentially negative for large $|\phi|$.

[16] If we choose completely arbitrary bulk scalar theories, we have no positive mass theorem, so we can have $\langle T_{tt} \rangle < 0$ in a homogeneous state, and the bounds are clearly violated. But in these situations there is no sensible holographic dual.

The numerical method

Let us now explain our procedure. For a given $V(\phi)$ and spacetime dimension, we will construct an n-parameter family of initial data, parametrized by coefficients $\{f_i\}_{i=1}^n$. The data will be provided on an extended homology hypersurface of some HRT surface. Then we will define the function $\mathcal{A}(\{f_i\})$ to be equal to the ratio

$$\left|\lim_{r\to\infty} r^{d-3}K_{\phi\phi}\right| / \mu(\infty)^{\frac{d-1}{d}} \tag{10.207}$$

in the initial dataset specified by parameters $\{f_i\}$. Different initial datasets correspond to different moments of time in different spacetimes (with different sizes of R). The value of $\mathcal{A}$ in some particular initial dataset corresponds to the instantaneous entanglement velocity $\mathfrak{R}$ in that configuration, and we will do a numerical maximization of $\mathcal{A}$ with respect to the parameters $\{f_i\}$. If we find that $\mathcal{A}$ is upper bounded, and that the upper bound is $\mathcal{A}_{\max}$, we have provided evidence that

$$\left|\frac{\mathrm{d}S_R}{\mathrm{d}t}\right| \leq \frac{1}{4}\mathcal{A}_{\max}\mathrm{Area}[\partial R]c_{\mathrm{eff}}\left[\frac{16\pi}{(d-1)c_{\mathrm{eff}}}\langle T_{tt}\rangle\right]^{\frac{d-1}{d}}. \tag{10.208}$$

If $\mathcal{A}$ is not upper bounded, or if $\mathcal{A}_{\max} > 1$, we have a counterexample to $\mathfrak{R} \leq 1$ in the theory under consideration.

We will also evaluate the function $\mathcal{B}(\{f_i\})$, which we define as the value of

$$\frac{|\partial_t S_R|}{\mathrm{Vol}[R]\langle T_{tt}\rangle} = \frac{4\pi}{d-1}\frac{L}{\ell}\frac{\left|\lim r^{d-3}K_{\phi\phi}\right|}{\mu(\infty)} \times \begin{cases} 2 & d=2 \\ 1 & d>2 \end{cases} \tag{10.209}$$

for any given initial dataset. By the same logic as earlier, if $\mathcal{B}$ is upper bounded by $\mathcal{B}_{\max}$, we have evidence that

$$\left|\frac{\mathrm{d}}{\mathrm{d}t}S[\rho_R(t)]\right| \leq \mathcal{B}_{\max}\mathrm{Vol}[R]\langle T_{tt}\rangle. \tag{10.210}$$

If $\mathcal{B}$ is not upper bounded, then our volume-type bounds break without the DEC.

Assuming we work with strips R, for a single evaluation of $\mathcal{A}$ and $\mathcal{B}$, we need to numerically solve the ODEs given by (10.56) and (10.57). For simplicity we will restrict to strips, since for spheres we cannot solve for $\Phi(r)$ analytically. In this case we would need to solve a set of three coupled equations instead.

Let us now specify our family of initial data. An explicit computation gives that

$$\begin{aligned} \mathcal{E} &= \frac{1}{2}\dot{\phi}(r)^2 + \frac{1}{2}\left(\frac{r^2}{L^2} - \frac{\omega(r)}{r^{d-2}}\right)\phi'(r)^2 + V(\phi), \\ \mathcal{J} &= \dot{\phi}(r)\phi(r), \end{aligned} \tag{10.211}$$

where $\dot{\phi} = t^a\nabla_a\phi|_\Sigma$. Specifying an initial dataset now corresponds to specifying the two profiles $\phi(r)$ and $\dot{\phi}(r)$, together with the initial value of $\omega(r_0)$. Letting

$$\Delta = d/2 + \sqrt{(d/2)^2 + m^2L^2} \tag{10.212}$$

be the scaling dimension of the CFT operator dual O to ϕ, the profiles we consider are

$$\begin{aligned}\phi &= f_1 \exp\left[-\left(\frac{r-f_2}{f_3}\right)^2\right] + \frac{f_4}{r^\Delta} + \frac{f_5}{r^{\Delta+2}} \\ \dot{\phi} &= f_6 \exp\left[-\left(\frac{r-f_7}{f_8}\right)^2\right] + \frac{f_9}{r^{\Delta+1}} + \frac{f_{10}}{r^{\Delta+3}}\end{aligned} \tag{10.213}$$

which gives a ten-parameter family of initial data. The gaussians give localized lumps of matter, while the power law falloffs ensures that we can turn on a VEV of O in the CFT, with $\langle O\rangle \propto f_4$ and $\langle \partial_t O\rangle \propto f_8$. Note that the seemingly unusual $1/r^{\Delta+1}$ falloff in $\dot{\phi}$ is just caused by the fact that the time derivative is with respect to a unit normal rather than the more standard global time coordinate near the conformal boundary.

What remains is to pick $\omega(r_0)$. To minimize the CFT energy, we want $\omega(r_0)$ small. When the DEC holds, we know that AdS hyperbolicity implies that $\omega(r_0) \geq 0$, as proven in Lemma 14. However, without the DEC we can have that $\omega(r_0)$ is negative, but not arbitrarily negative. If we pick $\omega(r_0)$ too negative, it will forbid an embedding of Σ in a complete slice. The difficulty is that how negative $\omega(r_0)$ can be depends on $\phi(r_0)$, $\dot{\phi}(r_0)$, and $V(\phi)$. We will thus restrict to $\omega(r_0) = 0$ and relegate a more complete study of the future. Even with $\omega(r_0) = 0$, it is far from obvious if our results survive breaking of the DEC, as we can easily obtain large regions of $\omega < 0$ even with $\omega(r_0) = 0$.

Finally, we need to deal with invalid datasets. For a given scalar profile, it could be that $\omega(r)$ overshoots r^d/L^2. In this case, the relevant solution does not correspond to a spacelike hypersurface, and so it must be discarded. In this case we conventionally define $\mathcal{A} = \mathcal{B} = 0$. Consequently, the functions we are maximizing will have discontinuities.

We are now ready to proceed to the numerical results.

$d = 2$

We now consider a free massive scalar field with mass

$$m^2L^2 = 0.9m^2_{\text{BF}}L^2 = -0.9, \tag{10.214}$$

dual to a relevant operator with $\Delta \approx 1.32$. We do not consider saturation of the BF bound, since this requires modification of the mass formula, and additionally causes $|\partial_t S|$ to be divergent (for any d). Furthermore, we do not want to go too close to the BF bound, since then $\omega(r)$ converges slowly at large r, and so the numerical maximization procedure becomes prohibitively expensive.

Using now Mathematica's built in `NMaximize` function, trying all methods for nonconvex optimization implemented in Mathematica and picking the best result, we find that

$$\mathcal{A}_{\text{max}} \approx 0.999 \leq v_E|_{d=2} = 1. \tag{10.215}$$

Thus, in $d = 2$ we have evidence that (10.25) holds without the DEC – at least in free tachyonic scalar theories.

Next, maximizing $\mathcal{B}$, we find that

$$\mathcal{B}_{\max} \approx 3.29 \leq \kappa_{d=2} = 2\pi. \tag{10.216}$$

This provides evidence that (10.32) holds when the DEC is violated, and that $\mathcal{O}(\ell^d \langle T_{tt} \rangle c_{\text{eff}}^{-1})$ corrections are not needed, even though we could not prove their absence outside thin shell spacetimes. In fact, given the large gap between $\mathcal{B}_{\max}$ and $\kappa_{d=2}$, the numerical results suggest that our proofs might possibly be sharpened.

$d = 3$

Now we consider two potentials:

$$\begin{aligned} V_{\text{I}}(\phi) &= \frac{1}{2}\left(0.9 m_{\text{BF}}^2\right)\phi^2, \\ V_{\text{II}}(\phi) &= 1 - \cosh\sqrt{2}\phi, \end{aligned} \tag{10.217}$$

with ϕ dual to operators with scaling dimensions $\Delta_{\text{I}} \approx 1.97$ and $\Delta_{\text{II}} = 2$, respectively. The potential V_{II} comes from a consistent truncation and dimensional reduction of eleven-dimensional SUGRA on $\text{AdS}_4 \times S^7$ [118]. We find

$$\begin{aligned} \mathcal{A}_{\text{I,max}} &\approx 0.693, \\ \mathcal{A}_{\text{II,max}} &\approx 0.702. \end{aligned} \tag{10.218}$$

In both cases $\mathcal{A}_{\max} < 1$, and so we have evidence that the conjectured bound (10.28) is true – even without the DEC and outside thin-shell spacetimes.

Now, we have that

$$v_E^{(\text{SAdS})} = \frac{\sqrt{3}}{2^{4/3}} = 0.687\ldots$$

In both cases $\mathcal{A}_{\max}$ is close to $v_E^{(\text{SAdS})}$, although it is slightly larger. It seems possible that a stronger bound

$$\mathfrak{R} \leq v_E^{(\text{SAdS})} + \delta v_E \tag{10.219}$$

is true for some small δv_E that potentially depends on the scalar potential.

For $\mathcal{B}$ we find

$$\begin{aligned} \mathcal{B}_{\text{I,max}} &\approx 1.71 \leq \kappa_{d=3} \approx 2.62, \\ \mathcal{B}_{\text{II,max}} &\approx 1.72. \end{aligned} \tag{10.220}$$

Again, there is a significant gap, with the implications being the same as for $d = 2$.

$d = 4$

We now consider

$$V(\phi) = \frac{1}{2}\left(0.9 m_{\text{BF}}^2\right)\phi^2. \tag{10.221}$$

and find

$$\mathcal{A}_{\max} \approx 0.643. \tag{10.222}$$

Table 10.1: Proven bounds on entanglement, spatial Wilson loops and equal-time correlators. We suppress $O(1)$ numerical constants in the table. Dots mean corrections scaling as $\mathcal{O}(\ell^d \langle T_{tt}\rangle / c)$ where ℓ is the relevant characteristic length scale, corresponding to strip width or ball radius. We abbreviate the effective central charge and 't Hooft coupling as c and λ, respectively. For proof validity equal to quench+, we mean proofs valid for states dual to spacetimes with thin-shell matter, which includes quenches as a subset.

$\boldsymbol{\partial_t S \leq}$	$\boldsymbol{d}$	**Region** $\boldsymbol{R}$	**Proof validity**	**Eq.**
$\sqrt{\langle T_{tt}\rangle / c}$	2	n intervals	general	(10.25)
$\mathrm{Vol}[R]\,\langle T_{tt}\rangle + \ldots$	≥ 2	small strip or ball	general	(10.30)
$\mathrm{Area}[\partial R][\langle T_{tt}\rangle / c]^{(d-1)/d}$	≥ 2	n strips	quench+	(10.28)
$\mathrm{Area}[\partial R][\langle T_{tt}\rangle / c]^{(d-1)/d}$	≥ 2	ball	quench+	(10.28)
$\mathrm{Vol}[R]\,\langle T_{tt}\rangle$	≥ 2	strip or ball	quench+	(10.32)
$\boldsymbol{\partial_\ell S \geq}$				
$\partial_\ell S_{\text{vacuum}}$	≥ 2	n strips	general	(10.197)
$\boldsymbol{\partial_t \ln \langle \mathcal{W}(C)\rangle \leq}$		$\mathbf{Length}[C]$		
$\sqrt{\lambda}\mathrm{Length}[C]\,[\langle T_{tt}\rangle / c]^{1/2}$	4	any	general	(10.173)
$\sqrt{\lambda}\mathrm{Length}[C]^{d-1}\,\langle T_{tt}\rangle + \ldots$	3, 4	small	general	(10.127)
$\sqrt{\lambda}\mathrm{Length}[C]\,[\langle T_{tt}\rangle / c]^{2/3}$	3	any	quench+	(10.132)
$\boldsymbol{\partial_t \ln \langle O(x)O(0)\rangle \leq}$		$\boldsymbol{\lvert x\rvert}$		
$\Delta\sqrt{\langle T_{tt}\rangle / c}$	2	any	general	(10.171)
$\Delta\lvert x\rvert\,\langle T_{tt}\rangle / c + \ldots$	2	small	general	(10.126)
$\boldsymbol{\partial_x \ln \langle O(x)O(0)\rangle \leq}$				
$\partial_x \ln \langle O(x)O(0)\rangle_{\text{vacuum}}$	2	any	general	(10.200)

Again we find evidence that (10.8) is true without the DEC or outside thin-shell spacetimes. We have

$$v_E^{(\mathrm{SAdS})} = \frac{\sqrt{2}}{3^{3/4}} = 0.620\ldots, \tag{10.223}$$

and so that the instantaneous growth can be above $v_E^{(\mathrm{SAdS})}$, but possibly only slightly so.

We also find

$$\mathcal{B}_{\max} = 1.91 \leq \kappa_{d=4} \approx 2.43. \tag{10.224}$$

Again, there is a significant gap, with the implications being the same as for $d = 2$.

10.6 Discussion

In this work we have proven several new upper bounds on the rate of change of entanglement entropy, spacelike Wilson loops, and equal-time two-point functions of heavy operators. The proofs apply for spatially homogeneous and isotropic states in strongly coupled CFTs with a holographic dual. We summarize our bounds in table 10.1. We have also provided numerical evidence that the bounds have broader validity than our proofs. We will now discuss our findings and possible future directions.

A 2d QWEC: The bound (10.5) can also be seen as a quantum weak energy condition (QWEC). Let S be the entropy of be a single interval as a function of one of the endpoints p, so that $\partial_t S$ now refers to the change of S under the perturbation of this single interval endpoint, rather than both. Then we have

$$\langle T_{tt}\rangle \geq \langle T_{tt}\rangle_{\text{vac}} + \frac{3}{2\pi c}\left(\partial_t S\right)^2, \tag{10.225}$$

while the classical weak energy condition implies that $T_{tt} \geq 0$. Equation (10.225) closely resembles the conformal quantum null energy condition (QNEC) [19, 124–128] in two dimensions.[17] Consider $2d$ Minkowski space, where $\langle T_{tt}\rangle_{\text{vac}} = 0$. Letting $x^{\pm}$ be null coordinates, the conformal QNEC says that [124, 127]

$$\langle T_{++}\rangle\,|_p \geq \frac{1}{2\pi}\partial_+^2 S + \frac{3}{\pi c}\left(\partial_+ S\right)^2. \tag{10.226}$$

The structural similarity between (10.225) and (10.226) is obvious. While (10.225) does not contain a second derivative, it is in principle possible that (10.225) could be true also for inhomogeneous states, provided we include a term $a\partial_t^2 S$ to the right hand side for some fixed constant a. In fact, the conformal QNEC suggests that $a = (4\pi)^{-1}$, since in the special case of a half-space in a homogeneous state, where $\partial_x S = 0$, the conformal QNEC and $T_{\mu}{}^{\mu} = 0$ implies

$$\langle T_{tt}\rangle\,|_p \geq \frac{1}{4\pi}\partial_t^2 S + \frac{3}{2\pi c}\left(\partial_t S\right)^2. \tag{10.227}$$

Why do things fall? In [46] it was proposed that the process of gravitational attraction is dual to the increase of complexity in the CFT. Assuming the complexity=volume conjecture [68], this was given a precise realization in [47–49] (see also [100]), where it was shown that the rate of change of the volume of a maximal volume slice is given by the momentum integrated on the slice. However, our formula

$$\frac{\mathrm{d}S_R}{\mathrm{d}t} = \int_X G t^a n^b \mathcal{T}_{ab} \tag{10.228}$$

shows that change in entanglement can also be seen as directly responsible for the radial momentum of matter. Thus, at present, "the increase of entanglement" seems like an equally good explanation for why things fall.

Relevant scalars, compact dimensions, and DEC breaking: Our proofs rely critically on the dominant energy condition – almost all steps of the proofs break without it. This rules out having scalars with negative squared mass, which are dual to relevant operators in the CFT. Nevertheless, we found numerical evidence that the bounds hold true without the DEC, as long as the scalar theories we consider allow a positive mass theorem, so that AdS is stable and $\langle T_{tt}\rangle$ is guaranteed to be positive.

However, there are other reasons to believe that our bounds remain true for these theories beyond our numerical findings – at least when working with top-down theories. Consider working with a theory that is a dimensional reduction and consistent truncation of type IIA, IIB, or eleven-dimensional SUGRA, so that any solution can be lifted to solutions on

[17] See [129] for a study on how the QNEC constrains entanglement growth.

asymptotically $\mathrm{AdS}_{d+1} \times K$ spacetimes for some compact manifold K. These solutions will typically be warped products rather than direct products, but there exists significant evidence [130] that the entropy computed by the HRT formula in the uplifted spacetime agrees with the one computed in the dimensionally reduced spacetime – even when the product is not direct. But in the uplifted spacetime the DEC holds, since it holds for type II and eleven-dimensional SUGRA. Thus, if our methods can be generalized to work for warped compactifications over spherically symmetric AAdS_{d+1} bases, this appears to be an avenue to prove our bounds even with relevant scalars turned on. The drawback is that the proofs might have to be carried out separately for each family of compactifications.

Strengthened bounds: Our proof that

$$\left|\frac{\mathrm{d}S_R}{\mathrm{d}t}\right| \leq \frac{1}{4}\mathrm{Area}[\partial R] c_{\mathrm{eff}} \left[\frac{16\pi}{(d-1)c_{\mathrm{eff}}} \langle T_{tt}\rangle\right]^{\frac{d-1}{d}}, \tag{10.229}$$

which implies that $\mathfrak{R} \leq 1$ in neutral states, only applied to thin-shell spacetimes, which are dual to CFT states where all dynamics happen at a single energy scale (that evolves with time). However, we gave numerical evidence that this bound also holds in general planar symmetric spacetimes with extended matter profiles. A natural extension of this work is trying to generalize the proof to include this. This will likely require a better understanding of non-linearities of the Einstein constraint equations.

Next, we found that in our numerical maximization of $\mathfrak{R}$ over a ten-parameter family of initial datasets in $d = 3, 4$, that

$$\mathfrak{R} \leq v_E^{(\mathrm{SAdS})} + \delta v_E, \tag{10.230}$$

where

$$v_E^{(\mathrm{SAdS})} = \sqrt{\frac{d}{d-2}} \left(\frac{d-2}{2(d-1)}\right)^{\frac{d-1}{d}} \tag{10.231}$$

is the entanglement velocity computed in a holographic quench having a neutral final state, and δv_E a small correction that seemed to depend on the scalar potential, but which was always small for the theories we studied (less than 0.03). This hints that it might be possible to strengthen the prefactor in (10.229). Similarly, our numerics suggested that the prefactors of (10.6) could be strengthened, and furthermore that this bound is true without $\mathcal{O}\left(\ell^d \langle T_{tt}\rangle / c_{\mathrm{eff}}\right)$ corrections.

$1/N$ corrections: It seems quite likely that $\mathfrak{R} \leq 1$ remains true with perturbative $1/N$ corrections. In fact, the pure QFT proofs of $\mathfrak{R} \leq 1$ for large subregions [37, 38] made no assumption about large-N, so only intermediate and small subregions could be sources violation. But for small subregions we showed that $\mathfrak{R} \leq \mathcal{O}(\ell/\beta) \ll 1$ for β the effective inverse temperature, which means perturbative $1/N$ corrections are unlikely to pose a of danger.[18] For intermediate sized regions things are less clear, but for $d = 3, 4$ we numerically did not manage to push $\mathfrak{R}$ close to 1, hinting that $1/N$ corrections do not pose a danger in these dimensions.

Disentangling the vacuum: Under our assumptions, we have proven for strips that the vacuum-subtracted entanglement is positive. Thus it is impossible to disentangle the vacuum

[18]The discovery of quantum extremal surfaces (QES) [18] far from classical extremal surfaces [20, 21] might make this argument somewhat less convincing, but it seems unlikely that these dominate/exist for small subregions, whose QES reside close to the conformal boundary.

in a spatially uniform way without either breaking the existence of a holographic dual, or turning on operators dual to fields that violate the DEC in the bulk. In the scenario where classical DEC violation is caused by tachyonic scalars only, this implies that all uniform states with less entanglement than the vacuum must have non-zero VEV for some relevant scalar single trace primaries.

Finite coupling: Our bounds were proven at strong coupling, but it seems possible that our bounds survive for arbitrary coupling. In [38] it was found that the entanglement velocity of a free theory (for $d > 2$) is strictly smaller than the holographic strong coupling result, suggesting that dialing up the coupling increases the capability of generating entanglement.

Primaries close to the unitarity bound: In order to turn on bulk fields dual to relevant CFT operators with scaling dimensions Δ in the window

$$\frac{d-2}{2} \leq \Delta < \frac{d}{2}, \tag{10.232}$$

we must consider scalars with masses

$$m^2_{\text{BF}} \leq m^2 < m^2_{\text{BF}} + 1/L^2, \tag{10.233}$$

and turn on the slow falloffs rather than fast falloffs (see for example [131–134]). This leads to violation of the falloff assumptions (10.19), and causes the ordinary definition of the spacetime mass to be divergent. Then neither of the Hawking masses reduce to the CFT energy at conformal infinity. Consequently, significant modifications of our proofs would be required. The same holds if we turn on sources that perturb us away from a CFT. Things can get even more challenging, given that for some falloffs $\partial_t S$ itself might become divergent [40]. In this case we should only try to bound finite quantities, like the mutual information or the renormalized entanglement entropy [93, 94], where these divergences cancel.

Hartman-Maldacena surfaces and half-spaces: Our bounds are restricted to subregions that lie in a single component of the conformal boundary. Our formalism can however readily be modified to deal with HRT surfaces corresponding to a CFT region R with connected components in different CFTs, so that the corresponding HRT surface threads a wormhole, like those studied in [135].[19] It would be interesting to see if the bounds remain unchanged. We also only considered strips of finite width, but it is clear that our formalism can also handle half-spaces, where $\ell = \infty$. Since (10.5) and (10.8) do not depend on ℓ, they almost certainly remain true for half-spaces.

End-of-the-world branes: Our bounds are proven with the assumption that the Cauchy slices are complete manifolds, which rules out spacetimes with end-of-the-world (EOW) branes. However, for HRT surfaces that do not end on the EOW brane, our proofs survive as long as we assume that the Hawking mass of the brane is positive, which ensures that Lemma 14 remains true. In the case where the surfaces end on the EOW brane, we expect the analysis to look similar to the analysis for HRT surfaces threading a wormhole.

Vaidya has optimal entanglement growth: There is a sense in which Vaidya spacetimes are maximizing the growth of entanglement for some given distribution of bulk energy density $\mathcal{E}(r)$, within the class DEC respecting planar spacetimes. In (10.89), we saw that the DEC imposed that the energy density at a point was lower bounded by the

[19] See [136] for a discussion of speed limits for these surfaces in static black holes.

momentum-density the same point (up to a factor). As is clear from our proofs, excess energy density above the lower bound contributes to increasing the CFT energy without increasing $\partial_t S$, and so we see that spacetimes saturating the DEC have the greatest $\partial_t S$ compatible with their distribution of energy density. But a direct computation shows that all Vaidya spacetimes saturate the DEC. This makes sense, since we have shown that infalling matter leads to entanglement growth and outgoing matter gives entanglement decrease, and in Vaidya no matter is wasted on being outgoing. While this is no proof, it suggests that speed limits that hold in all Vaidya spacetimes also hold in other planar-symmetric spacetimes respecting the DEC.

Inhomogeneous states: It would be interesting to prove bounds for inhomogeneous states, but this appears to be a difficult task, given how crucial the bulk planar symmetry was for our proofs. Furthermore, our bounds imply a strengthening of the positive mass theorem (PMT) for planar symmetry, and if generalizations of our results exist without the uniformity assumption, it appears likely that these result will strengthen the general PMT for AAdS spacetimes. Given how challenging it was to prove the PMT [137, 138], and how the Penrose inequality, which is the most famous strengthening of the PMT, still does not have a general proof, this is a daunting task.

Bounds with charge: It is a persistent finding that $U(1)$ gauge fields tend to slow down the growth of extremal surfaces of various dimensions [33, 34, 39, 139]. It thus seems plausible that our bounds can be strengthened by taking into account nonzero charges in the CFT. It is suggestive that, in spherical symmetry, $U(1)$ gauge fields contribute energy density, but they have no pure contribution to the momentum density – that is – they only contribute to $\mathcal{J}$ through gauge covariant derivatives acting on other matter fields.

Approximate bounds for large regions: To derive our bounds scaling with the volume of the region R, we used the inequality

$$\frac{L^2}{r_0} \leq \eta_d \ell, \tag{10.234}$$

where η_d is a numerical constant. Assume now that we consider a state that has volume law scaling for the entropy when ℓ is roughly above some length scale β. For $\ell \gg \beta$, as we increase ℓ, we expect r_0 to saturate at some radius r_{barrier}, and so (10.234) becomes a very poor bound. It seems likely that we can get stronger approximate bounds in this limit, by relating r_{barrier} to an effective inverse temperature β_{eff}, so that we effectively get a bound like $|\partial_t S| \leq \#\text{Area}[\partial R]\beta_{\text{eff}} \langle T_{tt}\rangle$. This bound might or might not reduce to (10.229).

A numerical laboratory: Our formalism makes it very easy to numerically compute properties HRT surfaces in planar symmetric spacetimes and test various hypothesis about their behavior without constructing full spacetimes and having to deal with the evolution of the Einstein equations. We imagine this can be used as a laboratory to learn more about the properties of HRT surfaces.

Other boundary geometries: Except for $d = 2$, our proofs always assumed Minkowski space on the boundary. However, our bounds survive if we compactify on a torus, so the boundary geometry is $\mathbb{R} \times T^{d-1}$, provided we make a few additional assumptions. For the bounds of the type $|\partial_t S| \leq \kappa \text{Vol}[R] \langle T_{tt}\rangle + \ldots$ (and the similar bounds for Wilson loops), we should always consider regions less than half the system size – otherwise Vol$[R]$ should be replaced with the volume of the complement. For bounds with multiple strips, if we have

a torus, we need to make sure that the entangling surfaces are all parallel, which happens automatically in Minkowski due to the parallel postulate.

Next, our proofs for single regions should imply growth bounds for CFTs on the static cylinder $\mathbb{R} \times S^{d-1}$, as long as we take the regions to be very small compared to the curvature radius of the boundary sphere. The results for balls will translate to results for small caps, while the result for strips will translate to results for thin belts around the equator.

Acknowledgments

It is a pleasure to thank Sean Colin-Ellerin, Netta Engelhardt, Gary Gibbons, Matt Headrick, Hong Liu, Dan Roberts, Jon Sorce, and Brian Swingle for discussions. A.D. was supported by the University of Minnesota Doctoral Dissertation Fellowship and by the National Science Foundation Graduate Research Fellowship under Grant No. 00039202. The research of Å.F. is supported in part by the John Templeton Foundation via the Black Hole Initiative, NSF grant no. PHY-2011905, and an Aker Scholarship. The research of Å.F. was also supported in part by the Heising-Simons Foundation, the Simons Foundation, and NSF grant no. PHY-1748958.

10.7 Appendix

10.7.1 The mean curvature of X in Σ

Let now $A, B, \ldots$ be indices for tensors on X, and $\alpha, \beta, \ldots$ be indices for tensors on Σ, and consider intrinsic coordinates on Σ and X from Sec. 10.2.2. The induced metrics are

$$\begin{aligned} \mathrm{d}s^2|_\Sigma &= H_{\mu\nu}\mathrm{d}x^\mu \mathrm{d}x^\nu = B(r)\mathrm{d}r^2 + r^2\left[\mathrm{d}\phi^2 + \delta_{ij}\mathrm{d}x^i\mathrm{d}x^j\right] \\ \mathrm{d}s^2|_X &= \gamma_{AB}\mathrm{d}y^A\mathrm{d}y^B = \left[B(r) + r^2\Phi'(r)^2\right]\mathrm{d}r^2 + r^2\delta_{ij}\mathrm{d}x^i\mathrm{d}x^j. \end{aligned} \tag{10.235}$$

A basis of tangent vectors to X in Σ is $\{e^\alpha_A\}$, with coordinate expressions

$$\begin{aligned} e^\mu_r &= (1, \Phi(r), 0), \\ e^\mu_i &= (0, 0, \delta^\alpha_i). \end{aligned} \tag{10.236}$$

The normal to X inside Σ reads

$$n_\mu = \sqrt{\frac{Br^2}{B + r^2(\Phi')^2}}\,(\Phi'(r), -1, 0) \tag{10.237}$$

With this in hand, we can compute the mean curvature of X in Σ:

$$\begin{aligned} \mathcal{K} &= \gamma^{AB} e^\mu_A e^\nu_B \nabla_\mu n_\nu \\ &= \frac{1}{r\sqrt{B}\left[B + r^2(\Phi')^2\right]^{3/2}}\left[(d-1)r^3\Phi'(r)^3 + \left(dB - \frac{1}{2}rB'\right)r\Phi' + r^2 B\Phi''\right]. \end{aligned} \tag{10.238}$$

10.7.2 Explicit form of $K - n^\alpha n^\beta K_{\alpha\beta}$

Noting that

$$n^\mu = \sqrt{\frac{Br^2}{B + r^2(\Phi')^2}} \left(\frac{\Phi'(r)}{B(r)}, -\frac{1}{r^2}, 0 \right) \tag{10.239}$$

we get

$$\begin{aligned} K = H^{\alpha\beta} K_{\alpha\beta} &= \frac{1}{B} K_{rr} + \frac{d-1}{r^2} K_{\phi\phi}, \\ K_{\alpha\beta} n^\alpha n^\beta &= \frac{Br^2}{B + r^2 \Phi'(r)^2} \left[\frac{\Phi'(r)^2}{B^2} K_{rr} + \frac{1}{r^4} K_{\phi\phi} \right]. \end{aligned} \tag{10.240}$$

Inserting $K_{rr}(r) = B(r)F(r)$ and doing some algebra, $K - n^\alpha n^\beta K_{\alpha\beta} = 0$ becomes (10.54).

10.7.3 Deriving formulas for $\partial_t S$ and $\partial_\ell S$

A formula for $\partial_t S$

In this section, we show that the entropy growth is proportional to the infalling matter flux. We will first need to prove that

$$\left. \frac{\mathrm{dArea}[X_t]}{\mathrm{d}t} \right|_{t=0} = \int_{\partial X} N^a \eta_a = -\frac{\mathrm{Area}[\partial R]}{L^{d-2}} \lim_{r\to\infty} r^{d-3} K_{\phi\phi} \tag{10.241}$$

For the calculation, we will construct the vector η^a, which is tangent to the boundary $\partial\mathcal{M}$, and N^a, the outwards unit normal to $\partial\Sigma = \partial\mathcal{M} \cap \Sigma$ in Σ, in a coordinate system. To do so, introduce the ADM coordinates adapted to the extended homology hypersurface

$$\mathrm{d}s^2|_{\mathcal{M}} = -r^2 \mathrm{d}\tau^2 + H_{\mu\nu}(\tau, x) \mathrm{d}x^\nu \mathrm{d}x^\nu, \tag{10.242}$$

where we took the shift to be vanishing, and the lapse to be r. $x^\mu = (r, \boldsymbol{x})$ are the coordinates on Σ, and $H_{\mu\nu}(\tau = 0, x)$ its induced metric, given by (10.36). The extinsic curvare of Σ reads

$$K_{\alpha\beta} = \frac{1}{2r} \partial_\tau H_{\alpha\beta}|_{\tau=0}. \tag{10.243}$$

Imagine now we have the coordinates $z^i = (t, \boldsymbol{x})$ on $\partial\mathcal{M}$ and take $\partial\Sigma$ to be located at $(r = r_c, t = \tau = 0)$ with a temporary cutoff $r = r_c$. We want to find embedding coordinates $(r(t), \tau(t))$ for $\partial\mathcal{M}$ such that the induced metric reads

$$\mathrm{d}s^2|_{\partial\mathcal{M}} = h_{ij} \mathrm{d}z^i \mathrm{d}z^j = \frac{r_c^2}{L^2} \left[-\mathrm{d}t^2 + L^2 \mathrm{d}\phi^2 + L^2 \, \mathrm{d}\boldsymbol{x}^2 \right] \tag{10.244}$$

The components of the induced metric then satisfy

$$h_{tt} = g_{rr} \dot{r}^2 - r^2 \dot{\tau}^2 = -\frac{r_c^2}{L^2}, \tag{10.245}$$

$$h_{\phi\phi} = g_{\phi\phi} = r_c^2. \tag{10.246}$$

Taking the derivative of the second equation, and then setting $t = 0$ gives a system of equations that is easily solved to give (see the appendix of [100])

$$\dot{\tau}(0) = \frac{1}{L\sqrt{1 - BK_{\phi\phi}^2 r^{-2}}}\bigg|_{r=r_c}, \tag{10.247}$$

$$\dot{r}(0) = -\frac{K_{\phi\phi}}{L\sqrt{1 - BK_{\phi\phi}^2 r^{-2}}}\bigg|_{r=r_c}, \tag{10.248}$$

where we have chosen the branch $\dot{\tau} > 0$. Thus, η^a in our ADM coordinate system reads

$$\eta^a = (\partial_t)^a = \dot{\tau}(0)(\partial_t)^a + \dot{r}(0)(\partial_r)^a. \tag{10.249}$$

Now, the tangents to ∂X are tangent to ∂R, and the sole remaining tangent vector e_r^α in (10.236) is then the normal to $\partial\Sigma$. Hence, up to a normalization C,

$$N^\mu = Ce_r^\mu = C(1, \Phi', 0), \tag{10.250}$$

which can be unit normalized and pushed forward to a spacetime vector yielding (in the coordinates (10.242)),

$$N^a = \frac{1}{\sqrt{B(r) + r_c^2\Phi'(r)^2}}(0, 1, \Phi'(r), 0)|_{r=r_c}. \tag{10.251}$$

We can now compute the integral on the cutoff regulated ∂X:

$$\begin{aligned}\int_{\partial X} \eta^a N_a &= -\frac{K_{\phi\phi}}{L\sqrt{1 - B(r_c)K_{\phi\phi}^2/r_c^2}} \times \frac{B(r_c)}{\sqrt{B(r_c) + r_c^2(\partial_r\Phi)^2}} \times r_c^{d-2}\int \mathrm{d}^{d-2}\boldsymbol{x} \\ &= -\frac{K_{\phi\phi}\sqrt{B(r_c)}}{L\sqrt{1 - B(r_c)K_{\phi\phi}^2/r_c^2}} \times \sqrt{1 - (r_0/r_c)^2} \times r_c^{d-2}\frac{\mathrm{Area}[\partial R]}{L^{d-2}},\end{aligned} \tag{10.252}$$

where we have used the differential equation for the embedding function $\Phi(r)$. In the large r limit, the asymptotic behaviors are

$$B(r) \sim \mathcal{O}(r^{-2}), \qquad K_{\phi\phi} \sim \mathcal{O}(r^{-(d-3)}). \tag{10.253}$$

Taking the cutoff to the boundary, one finds the area growth to be given by (10.43).

A covariant formula for $\partial_t S$

Now let us write the intergral formula for $\partial_t S$ in a covariant way. Letting t^a be the future unit normal to Σ and n^a the outwards normal to X tangent to Σ, we have for some function

G on X that only depends on r:

$$\begin{aligned} 8\pi G_N \int_X G(r) t^a n^b \mathcal{T}_{ab} &= \int \mathrm{d}^{d-2}\boldsymbol{x} \int_{r_0}^{\infty} \mathrm{d}r r^{d-2} \sqrt{B + r^2(\Phi')^2} G(r) n^r \mathcal{J} \\ &= \frac{\mathrm{Area}[\partial R]}{L^{d-2}} \int \mathrm{d}r r^{d-2} \frac{\sqrt{B} r}{B + r^2(\Phi')^2} \Phi' G(r) \mathcal{J} \\ &= \frac{\mathrm{Area}[\partial R]}{L^{d-2}} \int \mathrm{d}r r^{d-2} \left(\frac{r_0}{r}\right)^{2d-2} G(r) \mathcal{J} \sqrt{(r/r_0)^{2d-2} - 1} \\ &= \frac{\mathrm{Area}[\partial R]}{L^{d-2}} \int \mathrm{d}r \frac{r_0^{d-1}}{r^d} G(r) \mathcal{J} \sqrt{r^{2d-2} - r_0^{2d-2}}, \end{aligned} \tag{10.254}$$

where we used (10.237) and (10.40). Letting

$$G(r) = \frac{2\pi r^d}{(d-1) r_0^{d-1}}, \tag{10.255}$$

we get a covariant formula for the entropy growth

$$\frac{\mathrm{d}S_R}{\mathrm{d}t} = \int_X G t^a n^b \mathcal{T}_{ab} \tag{10.256}$$

A formula for $\partial_\ell S$

Consider a one-parameter family of HRT surfaces X_ℓ anchored at the strip region region R_ℓ, given by (10.21), but letting now ℓ vary, holding t fixed. Taking the vector field η^a generating the flow of ∂X_ℓ to be $\eta^a|_{\partial\mathcal{M}} = \frac{1}{L}(\partial_\phi)^a$, (10.41) becomes

$$\frac{\mathrm{d}}{\mathrm{d}\ell}\mathrm{Area}[X_\ell] = \frac{\mathrm{Area}[\partial R]}{L^{d-1}} \lim_{r\to\infty} r^{d-2} g_{\phi\phi} N^\phi. \tag{10.257}$$

Using (10.251), (10.242), and (10.40), this evaluates to

$$\frac{\mathrm{d}}{\mathrm{d}\ell}\mathrm{Area}[X_\ell] = \frac{\mathrm{Area}[\partial R]}{L^{d-1}} \lim_{r\to\infty} r^d \frac{\Phi'}{\sqrt{B + r^2(\Phi')^2}} = \frac{\mathrm{Area}[\partial R]}{L^{d-1}} r_0^{d-1}. \tag{10.258}$$

10.7.4 Expression for $\partial_t \|\boldsymbol{X}_t\|$

The derivation of (10.148) is almost identical to the derivation in Sec. 10.7.3. Let us just highlight what must be changed. First, we do not have an explicit formula for $\Phi'(r)$, but this does not matter, since everything we need is its rate of falloff, which we can read off to be $\Phi'(r) \sim \mathcal{O}(1/r^{q+2})$ from (10.143). Next, in (10.252), it is sufficient to replace $r_c^{d-2} \to r_c^q$. After doing that, and taking into account that $K_{\phi\phi}$ now has falloff $\mathcal{O}(1/r^{q-1})$, the $r_c \to \infty$ limit of (10.252) with these modifications gives (10.148).

10.7.5 Geometric properties of X and Σ

Let us now prove various properties of extended homology hypersurfaces. We give the proof for HRT surfaces of strips and comment how the proofs are modified for $(q+1)$-dimension extremal surfaces anchored at spheres.

Lemma 18. *The extended homology hypersurface Σ of a strip region in a spacetime with planar symmetry cannot have a throat in its interior, i.e., a radius where $B(r)$ diverges.*

Proof. Assume for contradiction that Σ has a throat T in its interior – if there are multiple, take T to be the outermost one. Since Σ by definition terminates at the plane tangent to the tip of X, this means that X must pass beyond the throat. But if X crosses the throat, there must be a point on $X \cap T$ where X is not tangent to T. Let now $U \subset \Sigma$ be the region outside T, which we can always cover with a coordinate system

$$\mathrm{d}s^2|_U = B(r)\mathrm{d}r^2 + r^2\mathrm{d}\boldsymbol{x}^2, \quad r \in [r_T, \infty), \quad B(r_T) = \infty, \quad B(r > r_T) < \infty. \tag{10.259}$$

where the throat is at $r = r_T$. The fact that X is not tangent to T means that $|\Phi'(r_T)| < \infty$. But the solution for the extremal surface reads

$$\Phi'(r)^2 = \frac{B(r)}{r^2(r^{2d-2}c - 1)} \tag{10.260}$$

for some constant c, and so $|\Phi'(r_T)| = \infty$, which is a contradiction. Hence T cannot exist in the interior of Σ. □

This proof goes through the case of X anchored at a dimension q sphere, as discussed in Sec. 10.3, simply taking Φ' to be computed from (10.143).

Next, we have the following:

Lemma 19. *The HRT surface X of a strip region R can only have one turning point.*

Proof. Assume for contradiction that X has multiple turning points. Then there must be at least one turning point of X in the interior of Σ that is a local maximum of the embedding function $r(\phi)$. Let $\phi = \phi_t$ where this turning point occurs, and let us restrict our attention to a neighbourhood $\phi \in \mathcal{O}_\epsilon = (\phi_t, \phi_t + \epsilon)$, where we can invert $r(\phi)$ to get $\Phi(r)$, describing the embedding in $\mathcal{O}_\epsilon$. We see that $\Phi'(r) < 0$ and $\Phi''(r) < 0$ in this neighbourhood. Now, the equation for the HRT surface in the neighbourhood $\mathcal{O}_\epsilon$ is

$$(d-1)r^3\Phi'^3 + r^2B\Phi'' + r\Phi'\left(dB - \frac{r}{2}B'\right) = 0. \tag{10.261}$$

Since we are in the interior of Σ, B is bounded on $\mathcal{O}_\epsilon$ by Lemma 18. Since Φ' diverges at the turning point, the equation for Φ' near the turning point reads

$$r^2B\Phi'' = -(d-1)r^3\Phi'^3\left[1 + \mathcal{O}\left(\frac{1}{(\Phi')^2}\right)\right], \tag{10.262}$$

where the correction can be neglected to arbitrarily good precision since B is bounded. But this implies that Φ'' and Φ' must have opposite signs in $\mathcal{O}_\epsilon$ for sufficiently small ϵ, which is a contradiction. Hence X can only have one turning point. □

We did not consider the case where $r'(\phi_t) = r''(\phi_t) = 0$, but this was shown to be ruled out by [109]. Also, note that this proof survives the case of spherical boundary anchoring and a $(q+1)$-dimensional extremal surface, since the new term $(\Phi')^3\chi(r)$ in the extremality equation (10.139) is subleading at the prospective turning point, since it scales like $\chi \sim 1/\Phi'$. Thus, (10.262) remains true (up to a numerical factor and an $\mathcal{O}(1/\Phi')$ correction).

10.7.6 General extremality conditions

Let

$$H_{ab} = g_{ab} + t_a t_b \tag{10.263}$$

be the induced metric on Σ. Then we have $h^{ab} = H^{ab} - \delta^{IJ} n_I^a n_J^b$, and so

$$\begin{aligned} \mathcal{K}^0 &= (H^{ab} - \delta^{IJ} n_I^a n_J^b)\nabla_a t_b \\ &= K - \delta^{IJ} n_I^a n_J^b \nabla_{(a} t_{b)} \\ &= K - \delta^{IJ} n_I^a n_J^b K_{ab} \end{aligned} \tag{10.264}$$

where we in the last line used that $n_I^a n_J^b$ is tangent to Σ, which projects out the difference $\nabla_{(a}t_{b)} - K_{ab}$. Now, a collection of tangents e_I^μ to X are

$$\begin{aligned} e^\mu{}_r &= (1, \Phi'(r), 0, 0), \\ e^\mu{}_i &= (0, 0, \delta_i^\mu, 0), \end{aligned} \tag{10.265}$$

where i runs over sphere directions. The second to last slot here runs over the sphere directions, with the last slot runs over the $\boldsymbol{z}$-directions. The unit normals to X (in Σ) are

$$\begin{aligned} n^r{}_\mu &= (\alpha, \beta, 0, 0), \\ n^i{}_\mu &= (0, 0, 0, r\delta_\mu^{y^i}), \end{aligned} \tag{10.266}$$

for some α, β we now work out, and where the coordinates are with respect to the index μ on Σ. r, i should be view as indices in the orthonormal tangent basis labeled by I on X. Now

$$\begin{aligned} 0 &= n^r{}_\mu e^\mu{}_r = \alpha + \beta\Phi'(r), \\ 1 &= H^{\mu\nu} n^r{}_\mu n^r{}_\nu = \frac{\alpha^2}{B} + \frac{\beta^2}{r^2}. \end{aligned} \tag{10.267}$$

Solving for α, β, we get

$$n^r{}_\mu = \sqrt{\frac{Br^2}{B + r^2(\Phi')^2}}\,(\Phi', -1, 0, 0)\,, \quad n^{r\mu} = \sqrt{\frac{Br^2}{B + r^2(\Phi')^2}}\left(\frac{1}{B}\Phi', -\frac{1}{r^2}, 0, 0\right). \tag{10.268}$$

Now we want to impose

$$\mathcal{K}^0 = K - \delta^{IJ} n_I^a n_J^b K_{ab} = 0. \tag{10.269}$$

Explicitly we have

$$\begin{aligned} K = H^{\mu\nu}K_{\mu\nu} &= \frac{1}{B}K_{rr} + \frac{1}{r^2}K_{\phi\phi} + \frac{w^{ij}}{r^2\phi^2} \times \underbrace{\phi^2 w_{ij} K_{\phi\phi}}_{K_{ij}} + \frac{d-2-q}{r^2}\underbrace{K_{\phi\phi}}_{K_{zz}}, \\ &= \frac{1}{B}K_{rr} + \frac{d-1}{r^2}K_{\phi\phi}, \end{aligned} \tag{10.270}$$

and

$$\delta^{IJ} n_I^a n_J^b K_{ab} = (n^{rr})^2 K_{rr} + (n^{r\phi})^2 K_{\phi\phi} + (d-2-q)\frac{1}{r^2} K_{\phi\phi}. \tag{10.271}$$

Thus, with $K_{rr} = BF$, condtion (10.269) reads

$$\left[1 - B(n_r^r)^2\right] F + \left[\frac{q+1}{r^2} - (n_r^\phi)^2\right] K_{\phi\phi} = 0 \tag{10.272}$$

Using the explicit formula for n_r^r, n_r^ϕ in (10.268), inserting the explicit formula for $\Phi(r)$ from (10.143), and solving for $F(r)$, we find (10.151).

Note also that thanks to (10.270), (10.78) is unchanged, and so the proof that $\pm\theta_\pm[\partial\Sigma] \geq 0$ for strips survive for dimension–$q+1$ surfaces anchored at q-spheres.

References

[1] A. Kitaev and J. Preskill, *Topological entanglement entropy*, *Phys. Rev. Lett.* **96** (2006) 110404, [hep-th/0510092].

[2] M. Levin and X.-G. Wen, *Detecting Topological Order in a Ground State Wave Function*, *prl* **96** (Mar., 2006) 110405, [cond-mat/0510613].

[3] A. Hamma, R. Ionicioiu, and P. Zanardi, *Ground state entanglement and geometric entropy in the kitaev model*, *Physics Letters A* **337** (mar, 2005) 22–28.

[4] C. H. Bennett and S. J. Wiesner, *Communication via one- and two-particle operators on einstein-podolsky-rosen states*, *Phys. Rev. Lett.* **69** (Nov, 1992) 2881–2884.

[5] C. H. Bennett, G. Brassard, C. Crépeau, R. Jozsa, A. Peres, and W. K. Wootters, *Teleporting an unknown quantum state via dual classical and einstein-podolsky-rosen channels*, *Phys. Rev. Lett.* **70** (Mar, 1993) 1895–1899.

[6] C. H. Bennett and G. Brassard, *Quantum cryptography: Public key distribution and coin tossing*, *Theoretical Computer Science* **560** (dec, 2014) 7–11.

[7] L. Bombelli, R. K. Koul, J. Lee, and R. D. Sorkin, *Quantum source of entropy for black holes*, *Phys. Rev. D* **34** (Jul, 1986) 373–383.

[8] J. D. Bekenstein, *Generalized second law of thermodynamics in black-hole physics*, *Phys. Rev. D* **9** (Jun, 1974) 3292–3300.

[9] D. N. Page, *Information in black hole radiation*, *Phys. Rev. Lett.* **71** (1993) 3743–3746, [hep-th/9306083].

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

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

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

[13] A. C. Wall, *A proof of the generalized second law for rapidly changing fields and arbitrary horizon slices*, *Phys. Rev. D* **85** (2012) 104049, [arXiv:1105.3445]. [Erratum: Phys.Rev.D 87, 069904 (2013)].

[14] A. C. Wall, *Maximin Surfaces, and the Strong Subadditivity of the Covariant Holographic Entanglement Entropy*, *Class. Quant. Grav.* **31** (2014), no. 22 225007, [arXiv:1211.3494].

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

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

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

[18] N. Engelhardt and A. C. Wall, *Quantum Extremal Surfaces: Holographic Entanglement Entropy beyond the Classical Regime*, *JHEP* **01** (2015) 073, [arXiv:1408.3203].

[19] R. Bousso, Z. Fisher, S. Leichenauer, and A. C. Wall, *Quantum focusing conjecture*, *Phys. Rev. D* **93** (2016), no. 6 064044, [arXiv:1506.02669].

[20] A. Almheiri, N. Engelhardt, D. Marolf, and H. Maxfield, *The entropy of bulk quantum fields and the entanglement wedge of an evaporating black hole*, *JHEP* **12** (2019) 063, [arXiv:1905.08762].

[21] G. Penington, *Entanglement Wedge Reconstruction and the Information Paradox*, *JHEP* **09** (2020) 002, [arXiv:1905.08255].

[22] H. Casini and M. Huerta, *A Finite entanglement entropy and the c-theorem*, *Phys. Lett. B* **600** (2004) 142–150, [hep-th/0405111].

[23] P. Calabrese and J. L. Cardy, *Evolution of entanglement entropy in one-dimensional systems*, *J. Stat. Mech.* **0504** (2005) P04010, [cond-mat/0503393].

[24] H. Casini and M. Huerta, *A c-theorem for the entanglement entropy*, *J. Phys. A* **40** (2007) 7031–7036, [cond-mat/0610375].

[25] I. R. Klebanov, D. Kutasov, and A. Murugan, *Entanglement as a probe of confinement*, *Nucl. Phys. B* **796** (2008) 274–293, [arXiv:0709.2140].

[26] P. V. Buividovich and M. I. Polikarpov, *Numerical study of entanglement entropy in SU(2) lattice gauge theory*, *Nucl. Phys. B* **802** (2008) 458–474, [arXiv:0802.4247].

[27] R. C. Myers and A. Sinha, *Seeing a c-theorem with holography*, *Phys. Rev. D* **82** (2010) 046006, [arXiv:1006.1263].

[28] R. C. Myers and A. Sinha, *Holographic c-theorems in arbitrary dimensions*, *JHEP* **01** (2011) 125, [arXiv:1011.5819].

[29] H. Casini and M. Huerta, *On the RG running of the entanglement entropy of a circle*, *Phys. Rev. D* **85** (2012) 125016, [arXiv:1202.5650].

[30] S. N. Solodukhin, *The a-theorem and entanglement entropy*, arXiv:1304.4411.

[31] V. Balasubramanian, A. Bernamonti, J. de Boer, N. Copland, B. Craps, E. Keski-Vakkuri, B. Muller, A. Schafer, M. Shigemori, and W. Staessens, *Thermalization of Strongly Coupled Field Theories*, *Phys. Rev. Lett.* **106** (2011) 191601, [arXiv:1012.4753].

[32] V. Balasubramanian, A. Bernamonti, J. de Boer, N. Copland, B. Craps, E. Keski-Vakkuri, B. Muller, A. Schafer, M. Shigemori, and W. Staessens, *Holographic Thermalization*, *Phys. Rev. D* **84** (2011) 026010, [arXiv:1103.2683].

[33] H. Liu and S. J. Suh, *Entanglement Tsunami: Universal Scaling in Holographic Thermalization*, *Phys. Rev. Lett.* **112** (2014) 011601, [arXiv:1305.7244].

[34] H. Liu and S. J. Suh, *Entanglement growth during thermalization in holographic systems*, *Phys. Rev. D* **89** (2014), no. 6 066012, [arXiv:1311.1200].

[35] K. Van Acoleyen, M. Mariën, and F. Verstraete, *Entanglement rates and area laws*, *Phys. Rev. Lett.* **111** (Oct, 2013) 170501.

[36] S. Bravyi, *Upper bounds on entangling rates of bipartite hamiltonians*, *Phys. Rev. A* **76** (Nov, 2007) 052319.

[37] T. Hartman and N. Afkhami-Jeddi, *Speed Limits for Entanglement*, arXiv:1512.02695.

[38] H. Casini, H. Liu, and M. Mezei, *Spread of entanglement and causality*, *JHEP* **07** (2016) 077, [arXiv:1509.05044].

[39] M. Mezei, *On entanglement spreading from holography*, *JHEP* **05** (2017) 064, [arXiv:1612.00082].

[40] D. Marolf and A. C. Wall, *State-Dependent Divergences in the Entanglement Entropy*, *JHEP* **10** (2016) 109, [arXiv:1607.01246].

[41] S. Kundu and J. F. Pedraza, *Spread of entanglement for small subsystems in holographic CFTs*, *Phys. Rev. D* **95** (2017), no. 8 086008, [arXiv:1602.05934].

[42] V. Balasubramanian and S. F. Ross, *Holographic particle detection*, *Phys. Rev. D* **61** (2000) 044007, [hep-th/9906226].

[43] P. Calabrese and J. L. Cardy, *Time-dependence of correlation functions following a quantum quench*, *Phys. Rev. Lett.* **96** (2006) 136801, [cond-mat/0601225].

[44] P. Calabrese and J. Cardy, *Quantum Quenches in Extended Systems*, *J. Stat. Mech.* **0706** (2007) P06008, [arXiv:0704.1880].

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

[46] L. Susskind, *Why do Things Fall?*, arXiv:1802.01198.

[47] J. L. F. Barbón, J. Martín-García, and M. Sasieta, *Momentum/Complexity Duality and the Black Hole Interior*, *JHEP* **07** (2020) 169, [arXiv:1912.05996].

[48] J. L. F. Barbon, J. Martin-Garcia, and M. Sasieta, *Proof of a Momentum/Complexity Correspondence*, *Phys. Rev. D* **102** (2020), no. 10 101901, [arXiv:2006.06607].

[49] J. L. F. Barbon, J. Martin-Garcia, and M. Sasieta, *A Generalized Momentum/Complexity Correspondence*, *JHEP* **04** (2021) 250, [arXiv:2012.02603].

[50] N. Engelhardt and A. C. Wall, *Coarse Graining Holographic Black Holes*, *JHEP* **05** (2019) 160, [arXiv:1806.01281].

[51] V. Balasubramanian and P. Kraus, *A Stress tensor for Anti-de Sitter gravity*, *Commun. Math. Phys.* **208** (1999) 413–428, [hep-th/9902121].

[52] J. Abajo-Arrastia, J. Aparicio, and E. Lopez, *Holographic Evolution of Entanglement Entropy*, *JHEP* **11** (2010) 149, [arXiv:1006.4090].

[53] W. Baron, D. Galante, and M. Schvellinger, *Dynamics of holographic thermalization*, *JHEP* **03** (2013) 070, [arXiv:1212.5234].

[54] E. Caceres and A. Kundu, *Holographic Thermalization with Chemical Potential*, *JHEP* **09** (2012) 055, [arXiv:1205.2354].

[55] D. Galante and M. Schvellinger, *Thermalization with a chemical potential from AdS spaces*, *JHEP* **07** (2012) 096, [arXiv:1205.1548].

[56] S. H. Shenker and D. Stanford, *Black holes and the butterfly effect*, *JHEP* **03** (2014) 067, [arXiv:1306.0622].

[57] S. H. Shenker and D. Stanford, *Multiple Shocks*, *JHEP* **12** (2014) 046, [arXiv:1312.3296].

[58] Y.-Z. Li, S.-F. Wu, Y.-Q. Wang, and G.-H. Yang, *Linear growth of entanglement entropy in holographic thermalization captured by horizon interiors and mutual information*, *JHEP* **09** (2013) 057, [arXiv:1306.0210].

[59] I. Aref'eva, A. Bagrov, and A. S. Koshelev, *Holographic Thermalization from Kerr-AdS*, *JHEP* **07** (2013) 170, [arXiv:1305.3267].

[60] P. Caputa, G. Mandal, and R. Sinha, *Dynamical entanglement entropy with angular momentum and U(1) charge*, *JHEP* **11** (2013) 052, [arXiv:1306.4974].

[61] T. Ugajin, *Two dimensional quantum quenches and holography*, arXiv:1311.2562.

[62] M. Alishahiha, M. R. Mohammadi Mozaffar, and M. R. Tanhayi, *On the Time Evolution of Holographic n-partite Information*, *JHEP* **09** (2015) 165, [arXiv:1406.7677].

[63] P. Fonda, L. Franti, V. Keränen, E. Keski-Vakkuri, L. Thorlacius, and E. Tonni, *Holographic thermalization with Lifshitz scaling and hyperscaling violation*, *JHEP* **08** (2014) 051, [arXiv:1401.6088].

[64] V. Keranen, H. Nishimura, S. Stricker, O. Taanila, and A. Vuorinen, *Dynamics of gravitational collapse and holographic entropy production*, *Phys. Rev. D* **90** (2014), no. 6 064033, [arXiv:1405.7015].

[65] M. Alishahiha, A. Faraji Astaneh, and M. R. Mohammadi Mozaffar, *Thermalization in backgrounds with hyperscaling violating factor*, *Phys. Rev. D* **90** (2014), no. 4 046004, [arXiv:1401.2807].

[66] E. Caceres, A. Kundu, J. F. Pedraza, and D.-L. Yang, *Weak Field Collapse in AdS: Introducing a Charge Density*, *JHEP* **06** (2015) 111, [arXiv:1411.1744].

[67] A. Buchel, R. C. Myers, and A. van Niekerk, *Nonlocal probes of thermalization in holographic quenches with spectral methods*, *JHEP* **02** (2015) 017, [arXiv:1410.6201]. [Erratum: JHEP 07, 137 (2015)].

[68] D. Stanford and L. Susskind, *Complexity and Shock Wave Geometries*, *Phys. Rev. D* **90** (2014), no. 12 126007, [arXiv:1406.2678].

[69] K. Babaei Velni, M. R. Mohammadi Mozaffar, and M. H. Vahidinia, *Evolution of*

entanglement wedge cross section following a global quench, *JHEP* **08** (2020) 129, [arXiv:2005.05673].

[70] S. Leichenauer and M. Moosa, *Entanglement Tsunami in (1+1)-Dimensions*, *Phys. Rev. D* **92** (2015) 126004, [arXiv:1505.04225].

[71] C. Ecker, D. Grumiller, and S. A. Stricker, *Evolution of holographic entanglement entropy in an anisotropic system*, *JHEP* **07** (2015) 146, [arXiv:1506.02658].

[72] V. Ziogas, *Holographic mutual information in global Vaidya-BTZ spacetime*, *JHEP* **09** (2015) 114, [arXiv:1507.00306].

[73] K. A. Sohrabi, *Inhomogeneous Thermal Quenches*, *Phys. Rev. D* **96** (2017), no. 2 026012, [arXiv:1509.00245].

[74] E. Caceres, M. Sanchez, and J. Virrueta, *Holographic Entanglement Entropy in Time Dependent Gauss-Bonnet Gravity*, *JHEP* **09** (2017) 127, [arXiv:1512.05666].

[75] M. Rangamani, M. Rozali, and A. Vincart-Emard, *Dynamics of Holographic Entanglement Entropy Following a Local Quench*, *JHEP* **04** (2016) 069, [arXiv:1512.03478].

[76] D. Roychowdhury, *Holographic thermalization from nonrelativistic branes*, *Phys. Rev. D* **93** (2016), no. 10 106008, [arXiv:1601.00136].

[77] C. Ecker, D. Grumiller, P. Stanzer, S. A. Stricker, and W. van der Schee, *Exploring nonlocal observables in shock wave collisions*, *JHEP* **11** (2016) 054, [arXiv:1609.03676].

[78] A. O'Bannon, J. Probst, R. Rodgers, and C. F. Uhlemann, *First law of entanglement rates from holography*, *Phys. Rev. D* **96** (2017), no. 6 066028, [arXiv:1612.07769].

[79] S. F. Lokhande, G. W. J. Oling, and J. F. Pedraza, *Linear response of entanglement entropy from holography*, *JHEP* **10** (2017) 104, [arXiv:1705.10324].

[80] R. C. Myers, M. Rozali, and B. Way, *Holographic Quenches in a Confined Phase*, *J. Phys. A* **50** (2017), no. 49 494002, [arXiv:1706.02438].

[81] I. Y. Aref'eva, M. A. Khramtsov, and M. D. Tikhanovskaya, *Thermalization after holographic bilocal quench*, *JHEP* **09** (2017) 115, [arXiv:1706.07390].

[82] V. Jahnke, *Delocalizing entanglement of anisotropic black branes*, *JHEP* **01** (2018) 102, [arXiv:1708.07243].

[83] M. Flory, J. Erdmenger, D. Fernandez, E. Megias, A.-K. Straub, and P. Witkowski, *Time dependence of entanglement for steady state formation in AdS_3/CFT_2*, *J. Phys. Conf. Ser.* **942** (2017), no. 1 012010, [arXiv:1709.08614].

[84] D. Ávila, V. Jahnke, and L. Patiño, *Chaos, Diffusivity, and Spreading of Entanglement in Magnetic Branes, and the Strengthening of the Internal Interaction*, *JHEP* **09** (2018) 131, [arXiv:1805.05351].

[85] H. Ghaffarnejad, E. Yaraie, and M. Farsam, *Holographic thermalization in AdS-Gauss-Bonnet gravity for small entangled regions*, *Gen. Rel. Grav.* **51** (2019), no. 1 10, [arXiv:1806.05976].

[86] W. Fischler, V. Jahnke, and J. F. Pedraza, *Chaos and entanglement spreading in a*

non-commutative gauge theory, *JHEP* **11** (2018) 072, [arXiv:1808.10050]. [Erratum: JHEP 02, 149 (2021)].

[87] M. Mezei, *Membrane theory of entanglement dynamics from holography*, *Phys. Rev. D* **98** (2018), no. 10 106025, [arXiv:1803.10244].

[88] J. Kudler-Flam, I. MacCormack, and S. Ryu, *Holographic entanglement contour, bit threads, and the entanglement tsunami*, *J. Phys. A* **52** (2019), no. 32 325401, [arXiv:1902.04654].

[89] J. Couch, S. Eccles, P. Nguyen, B. Swingle, and S. Xu, *Speed of quantum information spreading in chaotic systems*, *Phys. Rev. B* **102** (2020), no. 4 045114, [arXiv:1908.06993].

[90] Y. Ling, Y. Liu, and Z.-Y. Xian, *Entanglement entropy of an annulus in holographic thermalization*, *Chin. Phys. C* **44** (2020), no. 2 023101, [arXiv:1911.03716].

[91] M. Mezei and W. van der Schee, *Black holes often saturate entanglement entropy the fastest*, *Phys. Rev. Lett.* **124** (2020), no. 20 201601, [arXiv:2001.03172].

[92] K. Goto, M. Nozaki, K. Tamaoka, M. T. Tan, and S. Ryu, *Non-Equilibrating a Black Hole with Inhomogeneous Quantum Quench*, arXiv:2112.14388.

[93] H. Liu and M. Mezei, *A Refinement of entanglement entropy and the number of degrees of freedom*, *JHEP* **04** (2013) 162, [arXiv:1202.2070].

[94] H. Liu and M. Mezei, *Probing renormalization group flows using entanglement entropy*, *JHEP* **01** (2014) 098, [arXiv:1309.6935].

[95] N. Bao, C. Cao, S. Fischetti, and C. Keeler, *Towards Bulk Metric Reconstruction from Extremal Area Variations*, *Class. Quant. Grav.* **36** (2019), no. 18 185002, [arXiv:1904.04834].

[96] S. Fischetti and T. Wiseman, *A Bound on Holographic Entanglement Entropy from Inverse Mean Curvature Flow*, *Class. Quant. Grav.* **34** (2017), no. 12 125005, [arXiv:1612.04373].

[97] C. Fefferman and C. R. Graham, *Conformal invariants*, *Elie Cartan et les Mathématiques d'aujourdhui* (Astérisque) p. 95. 1985.

[98] C. Graham and J. M. Lee, *Einstein metrics with prescribed conformal infinity on the ball*, *Advances in Mathematics* **87** (1991), no. 2 186–225.

[99] C. R. Graham and E. Witten, *Conformal anomaly of submanifold observables in AdS / CFT correspondence*, *Nucl. Phys. B* **546** (1999) 52–64, [hep-th/9901021].

[100] N. Engelhardt and Å. Folkestad, *General bounds on holographic complexity*, *JHEP* **01** (2022) 040, [arXiv:2109.06883].

[101] R. Geroch, *Energy extraction**, *Annals of the New York Academy of Sciences* **224** (1973), no. 1 108–117.

[102] P. Jang and R. Wald *J. Math. Phys.* **18** (1977) 41.

[103] X. Wang, *The Mass of Asymptotically Hyperbolic Manifolds*, *Journal of Differential Geometry* **57** (2001), no. 2 273 – 299.

[104] G. Huisken and T. Ilmanen, *The Inverse Mean Curvature Flow and the Riemannian Penrose Inequality*, *Journal of Differential Geometry* **59** (2001), no. 3 353 – 437.

[105] S. W. Hawking, *Gravitational radiation in an expanding universe*, *Journal of Mathematical Physics* **9** (1968), no. 4 598–604, [https://doi.org/10.1063/1.1664615].

[106] Å. Folkestad, *The Penrose Inequality as a Constraint on the Low Energy Limit of Quantum Gravity*, arXiv:2209.00013.

[107] S. A. Hayward, *Gravitational energy in spherical symmetry*, *Phys. Rev. D* **53** (1996) 1938–1949, [gr-qc/9408002].

[108] H. Bray, S. Hayward, M. Mars, and W. Simon, *Generalized inverse mean curvature flows in spacetime*, *Commun. Math. Phys.* **272** (2007) 119–138, [gr-qc/0603014].

[109] N. Engelhardt and S. Fischetti, *Covariant Constraints on Hole-ography*, *Class. Quant. Grav.* **32** (2015), no. 19 195021, [arXiv:1507.00354].

[110] R. Bousso, *Holography in general space-times*, *JHEP* **06** (1999) 028, [hep-th/9906022].

[111] J. D. Brown and M. Henneaux, *Central Charges in the Canonical Realization of Asymptotic Symmetries: An Example from Three-Dimensional Gravity*, *Commun. Math. Phys.* **104** (1986) 207–226.

[112] J. M. Maldacena, *Wilson loops in large N field theories*, *Phys. Rev. Lett.* **80** (1998) 4859–4862, [hep-th/9803002].

[113] S.-J. Rey and J.-T. Yee, *Macroscopic strings as heavy quarks in large N gauge theory and anti-de Sitter supergravity*, *Eur. Phys. J. C* **22** (2001) 379–394, [hep-th/9803001].

[114] N. Engelhardt and Å. Folkestad, *Negative complexity of formation: the compact dimensions strike back*, *JHEP* **07** (2022) 031, [arXiv:2111.14897].

[115] T. Hertog, G. T. Horowitz, and K. Maeda, *Negative energy density in Calabi-Yau compactifications*, *JHEP* **05** (2003) 060, [hep-th/0304199].

[116] W. Boucher, *Positive energy without supersymmetry*, *Nuclear Physics B* **242** (1984), no. 2 282–296.

[117] P. Townsend, *Positive energy and the scalar potential in higher dimensional (super) gravity theories*, *Physics Letters B* **148** (1984), no. 1 55–59.

[118] M. Cvetic, M. J. Duff, P. Hoxha, J. T. Liu, H. Lu, J. X. Lu, R. Martinez-Acosta, C. N. Pope, H. Sati, and T. A. Tran, *Embedding AdS black holes in ten-dimensions and eleven-dimensions*, *Nucl. Phys. B* **558** (1999) 96–126, [hep-th/9903214].

[119] H. Lu and C. N. Pope, *Exact embedding of N=1, D = 7 gauged supergravity in D = 11*, *Phys. Lett. B* **467** (1999) 67–72, [hep-th/9906168].

[120] H. Lu, C. N. Pope, and T. A. Tran, *Five-dimensional N=4, SU(2) x U(1) gauged supergravity from type IIB*, *Phys. Lett. B* **475** (2000) 261–268, [hep-th/9909203].

[121] M. Cvetic, H. Lu, and C. N. Pope, *Gauged six-dimensional supergravity from massive type IIA*, *Phys. Rev. Lett.* **83** (1999) 5226–5229, [hep-th/9906221].

[122] P. Breitenlohner and D. Z. Freedman, *Positive energy in anti-de Sitter backgrounds AND gauged extended supergravity*, *Phys. Lett.* **B115** (1982) 197.

[123] P. Breitenlohner and D. Z. Freedman, *Stability in Gauged Extended Supergravity*, *Annals Phys.* **144** (1982) 249.

[124] A. C. Wall, *Testing the Generalized Second Law in 1+1 dimensional Conformal Vacua: An Argument for the Causal Horizon*, *Phys. Rev. D* **85** (2012) 024015, [arXiv:1105.3520].

[125] R. Bousso, Z. Fisher, J. Koeller, S. Leichenauer, and A. C. Wall, *Proof of the Quantum Null Energy Condition*, *Phys. Rev. D* **93** (2016), no. 2 024017, [arXiv:1509.02542].

[126] C. Akers, J. Koeller, S. Leichenauer, and A. Levine, *Geometric Constraints from Subregion Duality Beyond the Classical Regime*, arXiv:1610.08968.

[127] J. Koeller and S. Leichenauer, *Holographic Proof of the Quantum Null Energy Condition*, *Phys. Rev. D* **94** (2016), no. 2 024026, [arXiv:1512.06109].

[128] S. Balakrishnan, T. Faulkner, Z. U. Khandker, and H. Wang, *A General Proof of the Quantum Null Energy Condition*, *JHEP* **09** (2019) 020, [arXiv:1706.09432].

[129] M. Mezei and J. Virrueta, *The Quantum Null Energy Condition and Entanglement Entropy in Quenches*, arXiv:1909.00919.

[130] P. A. R. Jones and M. Taylor, *Entanglement entropy in top-down models*, *JHEP* **08** (2016) 158, [arXiv:1602.04825].

[131] T. Hertog and G. T. Horowitz, *Towards a big crunch dual*, *JHEP* **07** (2004) 073, [hep-th/0406134].

[132] M. Henneaux, C. Martinez, R. Troncoso, and J. Zanelli, *Asymptotically anti-de Sitter spacetimes and scalar fields with a logarithmic branch*, *Phys. Rev. D* **70** (2004) 044034, [hep-th/0404236].

[133] M. Henneaux, C. Martinez, R. Troncoso, and J. Zanelli, *Asymptotic behavior and Hamiltonian analysis of anti-de Sitter gravity coupled to scalar fields*, *Annals Phys.* **322** (2007) 824–848, [hep-th/0603185].

[134] T. Hertog and K. Maeda, *Black holes with scalar hair and asymptotics in $N = 8$ supergravity*, *JHEP* **07** (2004) 051, [hep-th/0404261].

[135] T. Hartman and J. Maldacena, *Time Evolution of Entanglement Entropy from Black Hole Interiors*, *JHEP* **05** (2013) 014, [arXiv:1303.1080].

[136] Z. Li and R.-Q. Yang, *Upper bounds of holographic entanglement entropy growth rate for thermofield double states*, *JHEP* **10** (2022) 072, [arXiv:2205.15154].

[137] R. Schoen and S. T. Yau, *Proof of the positive mass theorem. ii*, *Comm. Math. Phys.* **79** (1981), no. 2 231–260.

[138] E. Witten, *A Simple Proof of the Positive Energy Theorem*, *Commun. Math. Phys.* **80** (1981) 381.

[139] D. Carmi, S. Chapman, H. Marrochio, R. C. Myers, and S. Sugishita, *On the Time Dependence of Holographic Complexity*, *JHEP* **11** (2017) 188, [arXiv:1709.10184].