

The Markov gap for geometric reflected entropy

Patrick Hayden, Onkar Parrikar and Jonathan Sorce

*Stanford Institute for Theoretical Physics, Stanford University,
382 Via Pueblo Mall, Stanford, CA 94305-4060, U.S.A.*

E-mail: phayden@stanford.edu, parrikar@stanford.edu,
jsorce@stanford.edu

ABSTRACT: The reflected entropy $S_R(A : B)$ of a density matrix ρ_{AB} is a bipartite correlation measure lower-bounded by the quantum mutual information $I(A : B)$. In holographic states satisfying the quantum extremal surface formula, where the reflected entropy is related to the area of the entanglement wedge cross-section, there is often an order- N^2 gap between S_R and I . We provide an information-theoretic interpretation of this gap by observing that $S_R - I$ is related to the fidelity of a particular Markov recovery problem that is impossible in any state whose entanglement wedge cross-section has a nonempty boundary; for this reason, we call the quantity $S_R - I$ the *Markov gap*. We then prove that for time-symmetric states in pure AdS₃ gravity, the Markov gap is universally lower bounded by $\log(2)\ell_{\text{AdS}}/2G_N$ times the number of endpoints of the cross-section. We provide evidence that this lower bound continues to hold in the presence of bulk matter, and comment on how it might generalize above three bulk dimensions. Finally, we explore the Markov recovery problem controlling $S_R - I$ using fixed area states. This analysis involves deriving a formula for the quantum fidelity — in fact, for all the sandwiched Rényi relative entropies — between fixed area states with one versus two fixed areas, which may be of independent interest. We discuss, throughout the paper, connections to the general theory of multipartite entanglement in holography.

KEYWORDS: AdS-CFT Correspondence, Gauge-gravity correspondence

ARXIV EPRINT: [2107.00009](https://arxiv.org/abs/2107.00009)

Contents

1	Introduction	1
2	Information-theoretic origin of the Markov gap	9
2.1	Quantum information preliminaries	10
2.2	Cross-section boundaries and bulk entanglement	12
3	Lower bounds in pure AdS₃ gravity	15
3.1	Right angled hyperbolic pentagons	15
3.2	Intervals and asymptotic regions	17
3.3	The general proof	21
4	Bulk matter and higher dimensions	31
4.1	Spherically symmetric matter in three dimensions	31
4.2	Quantum perturbations	34
4.3	Higher dimensions	36
5	Holographic Markov recoveries and fixed area states	38
5.1	“Too Good to be True” recovery	39
5.2	Fixed area Markov recovery	41
6	Discussion	49
6.1	Generalizing the bound	49
6.2	Multipartite holographic entanglement	50
A	KRT surfaces and the quantum maximin formula	51

1 Introduction

Any operator A acting on a Hilbert space \mathcal{H} can be interpreted as a vector $|A\rangle\rangle$ in the Hilbert space $\mathcal{H} \otimes \mathcal{H}^*$, with \mathcal{H}^* the dual of \mathcal{H} . The inner product on this “doubled” Hilbert space is the Hilbert-Schmidt inner product

$$\langle\langle A|B\rangle\rangle = \text{Tr}(A^\dagger B). \tag{1.1}$$

If ρ is a density matrix on \mathcal{H} , then the state $|\sqrt{\rho}\rangle\rangle$ is a purification of ρ , i.e., it satisfies

$$\rho = \text{Tr}_{\mathcal{H}^*} |\sqrt{\rho}\rangle\rangle \langle\langle \sqrt{\rho}|. \tag{1.2}$$

Because the state $|\sqrt{\rho}\rangle\rangle$ is defined without reference to any choice of basis on the Hilbert space \mathcal{H} , it is called the *canonical purification* of ρ .

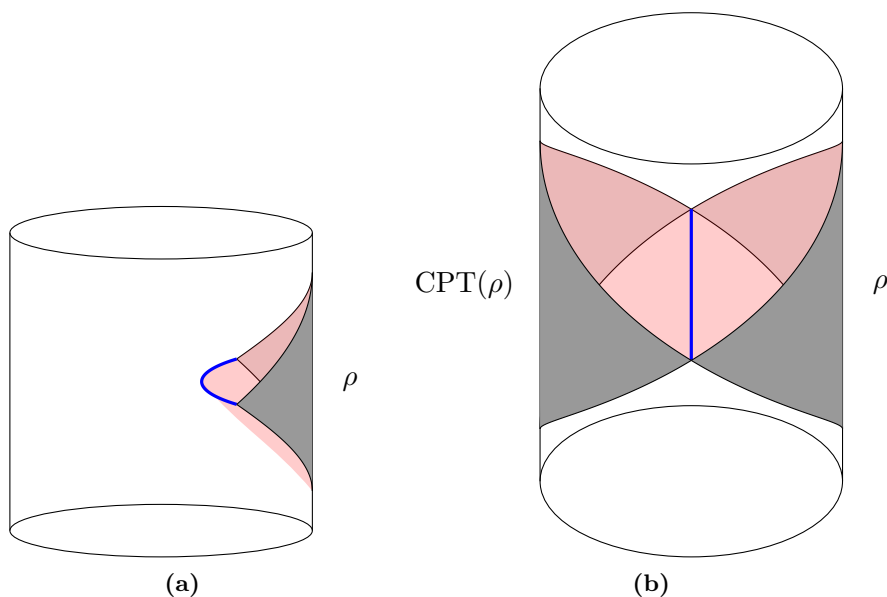


Figure 1. (a) A boundary state ρ whose bulk dual is a Rindler wedge in AdS_3 . (b) The initial data for the canonical purification $|\sqrt{\rho}\rangle\rangle$, formed by pasting $W(\rho)$ to its CPT conjugate along the quantum extremal surface that forms the spatial boundary of $W(\rho)$. The canonical purification is the AdS_3 vacuum state defined on a boundary circle whose circumference is twice the spatial extent of the domain of dependence on which ρ was initially defined.

In [1], Dutta and Faulkner provided an interesting interpretation of the canonical purification when ρ is a semiclassical state in a holographic theory of quantum gravity. Being semiclassical, ρ has a dual description in terms of an “entanglement wedge” $W(\rho)$ — a bulk domain of dependence bounded spatially by a quantum extremal surface.¹ The authors of [1] argued, using the equivalence of bulk and boundary path integrals in holography, that the canonical purification $|\sqrt{\rho}\rangle\rangle$ has a bulk description constructed by pasting $W(\rho)$ to its CPT conjugate along the quantum extremal surface, then solving the bulk equations of motion with this “doubled wedge” as initial data. An example of this pasting for the simple case when ρ describes an interval in the AdS_3 vacuum is sketched in figure 1.

In the case where ρ_{AB} is a bipartite density matrix, Dutta and Faulkner defined the *reflected entropy* as the entanglement entropy of the AA^* system in the canonical purification:

$$S_R(A : B) = S_{|\sqrt{\rho_{AB}}\rangle\rangle}(AA^*). \tag{1.3}$$

If ρ_{AB} is a semiclassical state satisfying the quantum extremal surface formula [15], then its canonical purification ought to satisfy the quantum extremal surface formula as well. It follows, then, that the reflected entropy $S_R(A : B)$ is equal — up to corrections expo-

¹We will assume the reader is familiar with the basics of the quantum extremal surface formula for boundary entropy and with the basics of entanglement wedge reconstruction. The relevant references are [2–20].

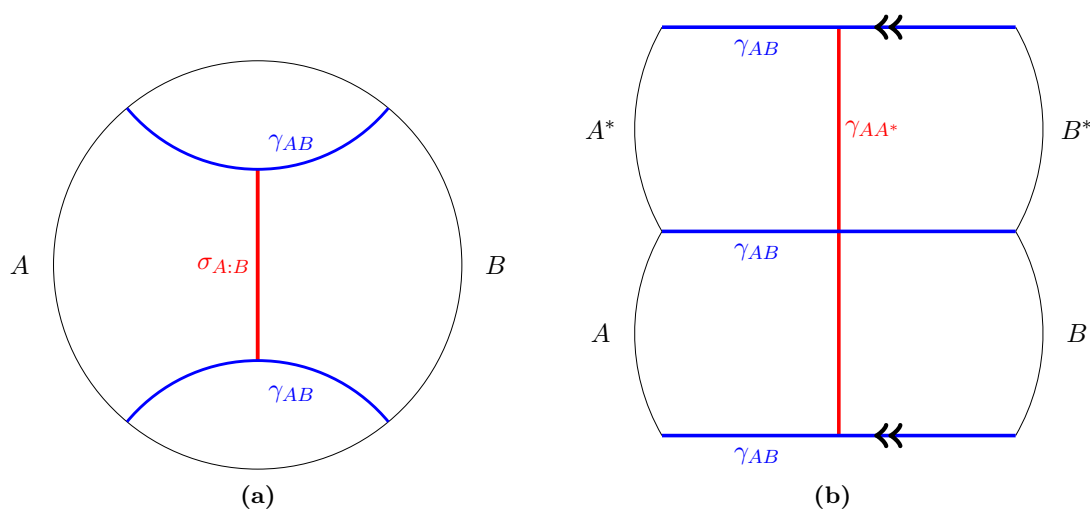


Figure 2. (a) A time slice of the AdS₃ vacuum with two boundary intervals A and B labeled. The intervals have been chosen large enough so that their minimal surface is γ_{AB} . The entanglement wedge cross-section is labeled $\sigma_{A:B}$. (b) A time slice of the canonical purification, formed by pasting $W(\rho_{AB})$ to its CPT conjugate along the surface γ_{AB} . The top and bottom edges of this figure are identified. The minimal surface of boundary region AA^* is the image of $\sigma_{A:B}$ under the symmetry $A \leftrightarrow A^*, B \leftrightarrow B^*$.

nentially small in $1/G_N$ — to the generalized entropy² of the minimal quantum extremal surface dividing the spacetime dual of $|\sqrt{\rho_{AB}}\rangle\rangle$ into one region through which the surface is homologous to AA^* and another through which it is homologous to BB^* . This is sketched in figure 2b for the special case when ρ_{AB} is the density matrix of two equal-time intervals in the AdS₃ vacuum.

Because the canonical purification is symmetric under interchange of $W(\rho_{AB})$ and $W(\rho_{A^*B^*})$, if the minimal quantum extremal surface computing $S_{|\sqrt{\rho_{AB}}\rangle\rangle}(AA^*)$ is unique then it must also obey this symmetry. The portion of this surface lying in the original spacetime $W(\rho_{AB})$ — equivalently, the image of this surface under the $A \leftrightarrow A^*, B \leftrightarrow B^*$ quotient — is called the *entanglement wedge cross-section* $\sigma_{A:B}$.³ The surface $\sigma_{A:B}$ divides

²The generalized entropy of a codimension-2 bulk surface γ with respect to a boundary domain of dependence Ω — defined only when γ is homologous to the spacelike slices of Ω — is defined by

$$S_{\text{gen}}(\gamma|\Omega) = \frac{\text{area}(\gamma)}{4G_N} + S_{\text{bulk}}(\gamma|\Omega),$$

where $S_{\text{bulk}}(\gamma|\Omega)$ is the entropy of bulk quantum fields lying in the domain of dependence spacelike between γ and Ω .

³The entanglement wedge cross-section was originally defined in [21, 22] as a proposed dual to a boundary quantity called the *entanglement of purification*. Similar claims have been made about the logarithmic negativity [23, 24] and the balanced partial entanglement [25]. These conjectures are not incompatible with the proposed duality between entanglement wedge cross-sections and reflected entropy, but they are more speculative. As such, we will focus entirely on the reflected entropy within this paper; however, under the conjectures mentioned above, all of our results can equivalently be made into statements about the other proposed duals of the entanglement wedge cross-section.

the entanglement wedge $W(\rho_{AB})$ into a portion through which $\sigma_{A:B}$ is homologous to A and a portion through which $\sigma_{A:B}$ is homologous to B . See figure 2a for a sketch.

The area contribution to the entropy $S_{|\sqrt{\rho_{AB}}\rangle}(AA^*)$ is twice the area of the entanglement wedge cross-section $\sigma_{A:B}$. The bulk entropy contribution is easily seen to be the reflected entropy of the bulk fields in $W(\rho_{AB})$ with respect to the bipartition induced by the dividing surface $\sigma_{A:B}$. From this, we find

$$S_{\text{gen}}(\gamma_{AA^*}|AA^*) = \frac{2 \text{area}(\sigma_{A:B})}{4G_N} + S_{R,\text{bulk}}(\sigma_{A:B}). \tag{1.4}$$

The quantum extremal surface formula tells us that $S_{|\sqrt{\rho_{AB}}\rangle}(AA^*)$ is the minimum of the above quantity over all quantum extremal surfaces γ_{AA^*} , which we can rewrite without reference to the canonical purification as⁴

$$S_R(A : B) = \min_{\sigma_{A:B}} \left[\frac{2 \text{area}(\sigma_{A:B})}{4G_N} + S_{R,\text{bulk}}(\sigma_{A:B}) \right], \tag{1.5}$$

where the minimum is taken over all candidates for the entanglement wedge cross-section that locally extremize the quantity in brackets. We will henceforth use the notation $\sigma_{A:B}$ to refer to the surface that achieves this minimum.

Many interesting properties of the reflected entropy were explored in [1]. One of particular interest is that in any quantum state, the reflected entropy always exceeds the mutual information:

$$S_R(A : B) \geq I(A : B). \tag{1.6}$$

While this inequality holds for *all* quantum states, it has a particularly suggestive geometric interpretation for holographic states satisfying the quantum extremal surface formula. For any bipartite boundary state ρ_{AB} with a semiclassical dual, one can define a special surface within the homology class of A by taking the union of the entanglement wedge cross-section $\sigma_{A:B}$ together with the portions of the AB quantum extremal surface that lie “between $\sigma_{A:B}$ and A ”. This is sketched for a simple case in figure 3. The surface thus constructed is homologous to A .⁵ We will call this surface the *KRT surface* $KRT(A)$ — for “kinked Ryu-Takayanagi,” because the surface will have right-angled kinks where $\sigma_{A:B}$ meets the quantum extremal surface of AB , as in figure 3 — and we will call the true minimal quantum extremal surface within that homology class $RT(A)$.⁶

⁴A generalization of this formula, which includes explicit contributions from entanglement islands, has appeared in [26].

⁵In fact, the requirement that there is a bipartition of the quantum extremal surface of AB into pieces $\gamma^{(A)} \cup \gamma^{(B)}$ such that $\gamma^{(A)} \cup \sigma_{A:B}$ is homologous to A and $\gamma^{(B)} \cup \sigma_{A:B}$ is homologous to B could be taken as the definition of an entanglement wedge cross-section — see section 3.3 for more discussion on this point.

⁶Veterans of the field might object: we’ve taken such great pains to ensure the present discussion holds for all quantum extremal surfaces, why should we suddenly start calling them RT surfaces when the term “RT” is usually used to refer to classically minimal surfaces in time-symmetric states? Unfortunately, we’ve found the acronyms QES(A) and KQES(A) just slightly too long to fit in our figures; we have elected to simplify matters by calling all minimal quantum extremal surfaces “RT surfaces.” Our apologies to colleagues Dong, Engelhardt, Faulkner, Headrick, Hubeny, Lewkowycz, Maldacena, Rangamani, and Wall, all of whom have made important contributions to the general quantum extremal surface formula but who will hopefully understand our avoiding the acronym DEFHLMRRTW.

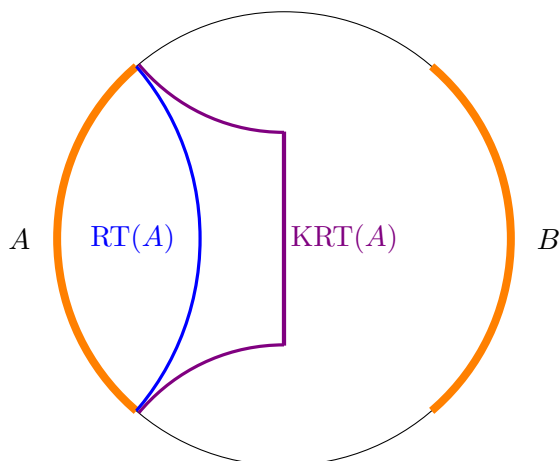


Figure 3. A sketch of the surfaces $\text{RT}(A)$ and $\text{KRT}(A)$ for the case where A and B are two equal-time intervals in the AdS_3 vacuum. $\text{KRT}(A)$ is formed by taking the union of the entanglement wedge cross-section $\sigma_{A:B}$ with the portion of $\text{RT}(AB)$ lying “towards A ” from the cross-section. Both $\text{KRT}(A)$ and $\text{RT}(A)$ are homologous to A .

The area term in $S_R(A : B) - I(A : B)$ is given by

$$[S_R - I]_{\text{area}} = \frac{2 \text{area}(\sigma_{A:B}) - \text{area}(\text{RT}(A)) - \text{area}(\text{RT}(B)) + \text{area}(\text{RT}(AB))}{4G_N}, \quad (1.7)$$

which we may rewrite in terms of KRT surfaces as

$$[S_R - I]_{\text{area}} = \frac{\text{area}(\text{KRT}(A)) - \text{area}(\text{RT}(A))}{4G_N} + \frac{\text{area}(\text{KRT}(B)) - \text{area}(\text{RT}(B))}{4G_N}. \quad (1.8)$$

To put the bulk entropy term of $S_R - I$ in a suggestive form, we apply the general inequality $S_R \geq I$ for the bulk state in ρ_{AB} subject to the bipartition induced by the entanglement wedge cross-section $\sigma_{A:B}$. Using the definition of the mutual information $I(P : Q) = S(P) + S(Q) - S(PQ)$, we obtain

$$S_{R,\text{bulk}}(\sigma_{A:B}) \geq S_{\text{bulk}}(\text{KRT}(A)|A) + S_{\text{bulk}}(\text{KRT}(B)|B) - S_{\text{bulk}}(\text{RT}(AB)|AB), \quad (1.9)$$

where S_{bulk} was defined in footnote 2.

Using equations (1.8) and (1.9), together with the boundary reflected entropy formula (1.5) and the boundary mutual information formula

$$I(A : B) = S_{\text{gen}}(\text{RT}(A)|A) + S_{\text{gen}}(\text{RT}(B)|B) - S_{\text{gen}}(\text{RT}(AB)|AB), \quad (1.10)$$

we obtain the inequality

$$S_R(A : B) - I(A : B) \geq [S_{\text{gen}}(\text{KRT}(A)|A) - S_{\text{gen}}(\text{RT}(A)|A)] + [S_{\text{gen}}(\text{KRT}(B)|B) - S_{\text{gen}}(\text{RT}(B)|B)]. \quad (1.11)$$

Because $\text{KRT}(A)$ (respectively $\text{KRT}(B)$) is in the same homology class as $\text{RT}(A)$ (respectively $\text{RT}(B)$), we might expect that its generalized entropy is superminimal, and thus that

each bracketed term on the right-hand side of the above expression is individually nonnegative. One has to be a little careful with this statement, though, because $\text{RT}(A)$ does not have minimal generalized entropy within its homology class, but rather minimal generalized entropy *among the quantum extremal surfaces in its homology class*. This caveat is necessary to allow for the fact that highly boosted surfaces in the $\text{RT}(A)$ homology class can have arbitrarily small areas. This is an obstacle to understanding inequality (1.11), because KRT surfaces are not generally quantum extremal. For states with no quantum matter, however, nonnegativity of the right-hand side of (1.11) can be established using the maximin formula [8]. If one accepts the quantum focusing conjecture [27] and the quantum maximin formula [28], then nonnegativity of the right-hand side of inequality (1.11) can be established in complete generality. The details of these arguments are given in appendix A.

While we did have to apply $S_R \geq I$ to the bulk quantum state in order to obtain inequality (1.11), the final expression is quite suggestive. It tells us that the quantity $S_R - I$ is related to the difference in generalized entropy between two preferred surfaces in the A and B homology classes — the surfaces $\text{KRT}(A)$ and $\text{RT}(A)$ (respectively $\text{KRT}(B)$ and $\text{RT}(B)$). The difference in generalized entropy between certain special candidate surfaces in a homology class is an interesting physical quantity; one might imagine that by studying it more carefully, we could improve our understanding of the inequality $S_R \geq I$ in holographic states. The goal of the present work is to undertake that analysis, and in doing so to prove a stronger inequality than $S_R \geq I$ in certain holographic states. Indeed, holographic states generally satisfy some special constraints on their entanglement structure; the monogamy of mutual information [29] is an example of an inequality that is satisfied by classical holographic states, but not by general quantum states. Finding such constraints which are special to holographic states can in turn provide insight into the entanglement structure of these states.⁷

The main technical contribution of this paper is a proof that in time-symmetric states of pure AdS_3 gravity satisfying the Ryu-Takayanagi formula, the inequality $S_R \geq I$ can be enhanced to

$$S_R(A:B) - I(A:B) \geq \frac{\log(2)\ell_{\text{AdS}}}{2G_N} \times (\# \text{ of cross-section boundaries}) + o\left(\frac{1}{G_N}\right). \quad (1.12)$$

In AdS_3 gravity, the entanglement wedge cross-section is a one-dimensional curve — the “# of cross-section boundaries” appearing in the above inequality is the number of endpoints of that curve. The main conceptual contribution of this paper will be an interpretation of the quantity $S_R - I$ — which we call the *Markov gap* for reasons made clear in section 2 — in terms of the optimal fidelity of a particular recovery process on the canonical purification $|\sqrt{\rho_{AB}}\rangle\rangle$. The interpretation of the Markov gap in terms of a recovery process will suggest a close link between the quantity $S_R - I$ and the geometry of the boundaries of the entanglement wedge cross-section; this link will be bolstered by the proof of inequality (1.12). We will also comment on potential generalizations of this inequality for non-time-symmetric states, states in higher-dimensional theories of gravity, and states with bulk matter.

⁷There is an interesting related research program that attempts to classify all von Neumann entropy inequalities implied by the Ryu-Takayanagi formula; see for example [30, 31].

Before proceeding to the plan of the paper, we detour to highlight a link between our work and a very interesting article by Akers and Rath [32]. In that paper, the authors used the Dutta-Faulkner formula for holographic reflected entropy to disprove a conjecture made in [33] about the entanglement structure of holographic states. Motivated by structural theorems about bit threads calculating holographic entropies, the original conjecture was that entanglement between three spatial regions in semiclassical boundary states is mostly *bipartite* in nature, in the sense that any multipartite contribution would be subleading in $1/G_N$. (The analogous statement for four boundary regions, by contrast, was already known to be false [34].) In [32], Akers and Rath showed that states with mostly bipartite entanglement patterns satisfy $S_R - I \approx 0$. They wrote down some particular boundary states for which $S_R - I$ is order $1/G_N$, proved some useful statements about the continuity of the reflected entropy, and concluded that the states in question must have significant amounts of tripartite entanglement. Combining their observations with our inequality (1.12) makes for a tantalizing observation: the mere *presence* of boundaries in an entanglement wedge cross-section requires significant quantities of tripartite entanglement in the corresponding boundary state. The fact that our inequality (1.12) scales with the *number* of cross-section boundaries seems to suggest that *every* boundary in the entanglement wedge cross-section must be supported by its own pattern of tripartite entanglement in the boundary state — this idea is taken up in considerably more detail in section 2.2.

The nature of that tripartite entanglement remains to be understood but at least one familiar form can be ruled out. The monogamy of mutual information [8, 29] eliminates a four- or higher-party GHZ entanglement structure in classical states in holography, because GHZ states explicitly violate the inequality.⁸ Even the three-party GHZ state can be ruled out as follows: consider the two party state obtained by tracing out one of the factors, and construct the bulk geometry dual to the canonical purification of that two party state using the Dutta-Faulkner trick described above. Had the original state been GHZ, the canonical purification would necessarily also be GHZ and violate the monogamy of mutual information, which is then a contradiction. It is our hope that inequality (1.12), and related inequalities that may be provable in more general holographic theories of gravity, will contribute to our understanding of the thorny but fascinating problem of multipartite holographic entanglement.⁹

One other related paper worth highlighting is [39]. In that paper, the authors started from the observation that in the vacuum state of pure AdS₃ gravity, the quantity $S_R - I$ for two neighboring, equal-time intervals is independent of the size of those intervals and equal to $\log(2)\ell_{\text{AdS}}/2G_N$. This is a saturation of inequality (1.12), because in that case (sketched in figure 4), the entanglement wedge cross-section has only one boundary. The authors computed $S_R - I$ — and the related quantity $2E_P - I$ with E_P the entanglement of purification — in various critical spin chains and found interesting universal features.

⁸More precisely, states that are *entirely* GHZ are excluded by monogamy. It is conceivable that GHZ-type entanglement could co-exist with other types of entanglement in such a way that the monogamy property remained satisfied overall even if it were violated by a GHZ-entangled factor in the boundary Hilbert space.

⁹While not directly related to the present paper, we would also like to draw attention to the series of papers [35–38], which undertake an orthogonal approach to understanding multipartite entanglement by defining multi-party generalizations of entanglement wedge cross-sections.

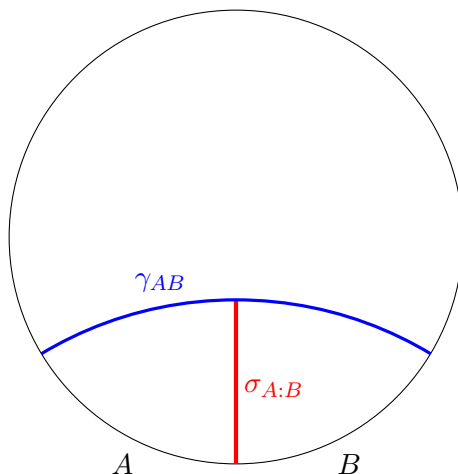


Figure 4. The minimal surface γ_{AB} and entanglement wedge cross-section $\sigma_{A:B}$ for the case where A and B are neighboring intervals on a time slice of the AdS_3 boundary. In this case, the entanglement wedge cross-section has only one boundary. It has no boundary at infinity, since that point is at infinite distance.

The plan of the present paper is as follows.

In section 2, we introduce the idea of a *Markov recovery process* from quantum information theory, and show that $S_R(A : B) - I(A : B)$ is lower bounded by a function of the fidelity of a particular Markov recovery process on $|\sqrt{\rho_{AB}}\rangle\rangle$. We give a preliminary holographic interpretation of this recovery process, and argue that boundaries of the entanglement wedge cross-section present obstructions to this recovery process that are reflected in the Markov gap. The geometric argument of this section is largely heuristic, and relies on the general principle from [40] that geometric connections in a bulk state are supported by entanglement patterns in the boundary state.

In section 3, we prove the claimed inequality (1.12) for time-symmetric states in pure AdS_3 quantum gravity. The proof is an exercise in two-dimensional hyperbolic geometry; we first prove the desired lower bound on a quantity analogous to $\text{area}(\text{KRT}(A)) - \text{area}(\text{RT}(A))$ defined purely in the hyperbolic disc, then use the fact that every hyperbolic 2-manifold is covered by the hyperbolic disc to prove the general bound. One charming feature of the proof is that it constructs a partial tiling of the homology region between $\text{KRT}(A)$ and $\text{RT}(A)$ by right-angled hyperbolic pentagons, with each kink in $\text{KRT}(A)$ being a vertex of its own pentagon; the desired inequality follows from standard trigonometric relations among the side-lengths of these pentagons.

In section 4, we discuss potential generalizations of inequality (1.12) beyond the regime of time-symmetric, pure, AdS_3 gravity. We provide evidence that the bound continues to hold in AdS_3 gravity upon the addition of bulk matter, and discuss a possible generalization of (1.12) to bulk dimensions greater than 3.

In section 5, we return to the understanding developed in section 2 relating the Markov gap to recovery processes on $|\sqrt{\rho_{AB}}\rangle\rangle$. We create a tractable model of the relevant recovery process using the fixed area states introduced in [41, 42], and compute a bound on the

fidelity of the process using the gravitational path integral. This computation reproduces some features of the inequality (1.11) directly from the perspective of Markov recovery processes. The techniques developed in this section may be of independent interest — the story we tell about recovery processes can be reduced to a more general question that is physically interesting in its own right:

- Given a boundary region R and a state ρ_R whose bulk entanglement wedge is bounded by the bulk surface γ , and which contains another covariantly defined surface $\tilde{\gamma}$ with greater generalized entropy, define $\tilde{\rho}_R$ to be a state whose entanglement wedge ends at $\tilde{\gamma}$. What is the fidelity $F(\rho_R, \tilde{\rho}_R)$?¹⁰

In a fixed-area-state model of this question, when the state $\tilde{\gamma}$ is extremal, we compute not only the fidelity but all of the sandwiched Rényi relative entropies defined in [43, 44]; this result is given in equations (5.28) and (5.33). Readers primarily interested in understanding this result can skip directly to the matching bullet of section 5.2 without losing any essential context.

Finally, in section 6, we conclude with a summary of our results and some possible directions for future work. We pay particular attention to what the present explorations might teach us about multipartite holographic entanglement, and to what techniques might be used to generalize inequality (1.12).

In appendix A, we show that the inequality $S_{\text{gen}}(\text{KRT}(A)|A) \geq S_{\text{gen}}(\text{RT}(A)|A)$ follows from the quantum maximin formula.

Throughout the paper we set $c = \hbar = 1$ while leaving G_N explicit. We frequently work in units where the AdS radius is set to 1. The term “area” is used to refer to the volume of any codimension-2 bulk hypersurface, regardless of the total bulk dimension. Once or twice we use the term “area” to refer to the volume of a codimension-3 hypersurface, but we signpost this explicitly in the relevant sections. When distinguishing between the regimes of validity of the “minimal” versus “extremal” surface formulas for holographic entanglement entropy, we use the term “time-symmetric” to refer to any spacetime that has a moment of time symmetry — i.e., a complete, achronal slice whose extrinsic curvature tensor vanishes. Time-symmetric states do not necessarily have a global $t \mapsto -t$ symmetry.

2 Information-theoretic origin of the Markov gap

As mentioned in the introduction, the quantity $S_R(A : B) - I(A : B)$ can be related to the fidelity of a particular Markov recovery process on the canonical purification of ρ_{AB} . In subsection 2.1, we review and explain the results from quantum information theory needed to understand this claim. We define Markov recovery maps and the quantum fidelity of states, and give a refinement of the inequality $S_R - I \geq 0$ in terms of the fidelity of a Markov recovery map. In subsection 2.2, we give a holographic interpretation of the refined inequality, and explain why the Markov gap must be nonzero at order $1/G_N$ whenever the entanglement wedge cross-section of ρ_{AB} has a nontrivial boundary.

¹⁰The curious reader is encouraged to look ahead to figure 20 for an illustration of this setup.

2.1 Quantum information preliminaries

Given a three-party quantum state ρ_{ABC} , a *Markov recovery map* is a quantum channel from a one-party subsystem into a two-party subsystem. For example, we might have a map $\mathcal{R}_{B \rightarrow BC}$ that takes system B into system BC . By acting on the reduced state ρ_{AB} with this channel, we produce a tripartite state on the whole system:

$$\tilde{\rho}_{ABC} = \mathcal{R}_{B \rightarrow BC}(\rho_{AB}). \quad (2.1)$$

By a *Markov recovery process*, we will mean the problem of trying to reproduce the tripartite state ρ_{ABC} from one of its bipartite reduced states (for example ρ_{AB}) using a Markov recovery map (for example $\mathcal{R}_{B \rightarrow BC}$.)

The name “Markov” was first associated with this kind of problem in [45] — based on earlier work in [46] on an analogous problem involving countably many systems — where a state ρ_{ABC} that can be *perfectly* recovered from ρ_{AB} via a quantum channel on B was called a *short quantum Markov chain* for the ordering $A \rightarrow B \rightarrow C$. I.e., ρ_{ABC} is a quantum Markov chain for $A \rightarrow B \rightarrow C$ if there exists a quantum channel $\mathcal{R}_{B \rightarrow BC}$ satisfying

$$\rho_{ABC} = \mathcal{R}_{B \rightarrow BC}(\rho_{AB}). \quad (2.2)$$

It follows from the results of [47] that ρ_{ABC} is a quantum Markov chain for $A \rightarrow B \rightarrow C$ if and only if the conditional mutual information $I(A : C|B)$ vanishes.¹¹

The conditional mutual information is defined as the following linear combination of entanglement entropies:

$$I(A : C|B) = S(AB) + S(BC) - S(ABC) - S(B) = I(A : BC) - I(A : B). \quad (2.3)$$

The famous strong subadditivity inequality is exactly the statement that conditional mutual information is always nonnegative. Roughly speaking, $I(A : C|B)$ measures the amount of correlation between the A and C subsystems that does not “pass through” the B subsystem. When $I(A : C|B)$ vanishes, all correlations between A and C are visible to the B subsystem, and can be reproduced by a quantum channel acting only on B ; this is the intuition behind the statement that $A \rightarrow B \rightarrow C$ is a quantum Markov chain iff $I(A : C|B)$ vanishes.

In [48], the authors proved a refinement of the relationship between Markov chains and conditional mutual information.¹² They showed that states with small (but nonzero) conditional mutual information are *approximate* Markov chains — that is, they admit Markov recovery processes that do a good job reproducing ρ_{ABC} as measured by the quantum fidelity. The formal statement is as follows:

$$\max_{\mathcal{R}_{B \rightarrow BC}} F(\rho_{ABC}, \mathcal{R}_{B \rightarrow BC}(\rho_{AB})) \geq e^{-I(A:C|B)}. \quad (2.4)$$

¹¹Classically, the random variables X, Y, Z form a Markov chain if and only if their joint distribution factorizes as $p(x, y, z) = p(x)p(y|x)p(z|y)$. When $I(A : C|B)$ vanishes in the quantum case, this factorization generalizes to a canonical form for ρ_{ABC} [45].

¹²A series of follow-up papers culminating in [49] actually proved a much more general result — a refinement of the monotonicity of relative entropy under quantum channels — of which strong subadditivity is a special case. A good review of the general theorem is presented in chapter 12 of [50], with the application to conditional mutual information given in section 12.6.1.

The quantity F appearing on the left-hand side of this inequality is the quantum fidelity, defined by

$$F(\rho, \sigma) = \left[\text{Tr} \sqrt{\sqrt{\rho} \sigma \sqrt{\rho}} \right]^2. \quad (2.5)$$

It is symmetric in its arguments, lies in the range $0 \leq F(\rho, \sigma) \leq 1$, equals 1 if and only if ρ and σ are equal, and equals 0 if and only if ρ and σ have orthogonal support. If two states have fidelity close to 1, then for any bounded observable \mathcal{O} the expectation values $\text{Tr}(\rho \mathcal{O})$ and $\text{Tr}(\sigma \mathcal{O})$ are similar.¹³ Elementary properties of the fidelity are nicely reviewed in chapter 9.2.2 of [51].

Let us now take a closer look at inequality (2.4). The left-hand side of the inequality involves a maximum over all Markov recovery channels $\mathcal{R}_{B \rightarrow BC}$. If $I(A : C|B)$ is close to zero, then the right-hand side of the inequality is close to 1; this means that there must exist *some* Markov recovery map $\mathcal{R}_{B \rightarrow BC}$ for which the fidelity $F(\rho_{ABC}, \mathcal{R}_{B \rightarrow BC}(\rho_{AB}))$ is close to one. So when $I(A : C|B)$ is small, there exists some Markov recovery map on B that approximately recovers ρ_{ABC} from ρ_{AB} . The way this inequality was proved in [48, 49] was by constructing a particular Markov recovery map — called the *rotated Petz map* or *twirled Petz map* — that beats $e^{-I(A:C|B)}$ in terms of fidelity.¹⁴

In its current form, inequality (2.4) looks like a bound on the fidelity of an optimal Markov recovery process in terms of the conditional mutual information. But we are free to rewrite it as a bound on the conditional mutual information in terms of the fidelity of the optimal Markov recovery process:

$$I(A : C|B) \geq - \max_{\mathcal{R}_{B \rightarrow BC}} \log F(\rho_{ABC}, \mathcal{R}_{B \rightarrow BC}(\rho_{AB})). \quad (2.6)$$

We can view this as a state-dependent enhancement of the strong subadditivity inequality. The ordinary strong subadditivity inequality is $I(A : C|B) \geq 0$. If we happen to know, however, that the optimal Markov recovery process for the chain $A \rightarrow B \rightarrow C$ is imperfect — i.e., its fidelity is less than one — then inequality (2.6) tells us that $I(A : C|B)$ must be bounded away from zero.

¹³The precise version of this “similarity” statement goes as follows. The fidelity is related to the one-norm distance by (see eqs. (9.100–9.101) of [51]), though note their definition of fidelity differs by a power of 2)

$$\|\rho - \sigma\|_1 \leq 2\sqrt{1 - F(\rho, \sigma)},$$

so when F is close to one, the one-norm distance is close to zero. The Schatten norms satisfy a form of Hölder’s inequality, giving

$$|\text{Tr}((\rho - \sigma)\mathcal{O})| \leq \text{Tr}(|(\rho - \sigma)\mathcal{O}|) = \|(\rho - \sigma)\mathcal{O}\|_1 \leq \|\rho - \sigma\|_1 \|\mathcal{O}\|_\infty,$$

where $\|\cdot\|_\infty$ is the *operator norm*, equal to the largest absolute value among the eigenvalues of \mathcal{O} .

¹⁴A curious feature of the twirled Petz map is that the reconstructed state

$$\tilde{\rho}_{ABC} = \mathcal{R}_{\text{tPetz}, B \rightarrow BC}(\rho_{AB})$$

is not only very close to ρ_{ABC} in terms of the fidelity, but is *exactly* equal to ρ_{ABC} on the BC subsystem; i.e., we have $\rho_{BC} = \tilde{\rho}_{BC}$. Again, we refer the reader to section 12.6.1 of [50] for a review of this fact.

For the reflected entropy of a bipartite state ρ_{AB} , the quantity $S_R(A : B) - I(A : B)$ can be written as a conditional mutual information of the canonical purification:

$$S_R(A : B) - I(A : B) = I(A : B^*|B) = I(B : A^*|A). \tag{2.7}$$

This observation was made in [1], and it is in fact how they proved the inequality $S_R \geq I$. Using inequality (2.6), however, we can prove two stronger inequalities:

$$S_R(A : B) - I(A : B) \geq - \max_{\mathcal{R}_{B \rightarrow BB^*}} \log F(\rho_{ABB^*}, \mathcal{R}_{B \rightarrow BB^*}(\rho_{AB})). \tag{2.8}$$

$$S_R(A : B) - I(A : B) \geq - \max_{\mathcal{R}_{A \rightarrow AA^*}} \log F(\rho_{AA^*B}, \mathcal{R}_{A \rightarrow AA^*}(\rho_{AB})). \tag{2.9}$$

These inequalities are what lead us to call the quantity $S_R - I$ the *Markov gap*.

2.2 Cross-section boundaries and bulk entanglement

We will now argue on quite general grounds — albeit rather heuristically — that in holographic states, the fidelities of Markov recovery processes on the canonical purification are controlled by the number of boundaries in the entanglement wedge cross-section. Put simply: the more boundaries there are in the entanglement wedge cross-section, the harder it is for a Markov recovery process to accurately reproduce the chain $A \rightarrow B \rightarrow B^*$ (or $B \rightarrow A \rightarrow A^*$). This discussion serves as a prelude to the concrete analysis of section 3, where we will prove our claimed inequality (1.12) lower-bounding $S_R - I$ in terms of the number of cross-section boundaries for time-symmetric states in AdS₃ gravity.

Let us examine in some detail a particular example of a canonical purification. Let ρ_{AB} be the density matrix of two equal-time intervals in the AdS₃ vacuum, with the intervals being large enough that the entanglement wedge of ρ_{AB} is connected. In figure 5, we have sketched a static time-slice of the canonical purification of ρ_{AB} . As explained in the introduction, the bulk dual of a canonical purification has, as initial data, the geometry formed by gluing together two copies of the entanglement wedge along their spacelike boundaries. For the specific case of two equal-time intervals in the AdS₃ vacuum, the canonical purification is a two-boundary wormhole.

We will focus on the Markov chain $A \rightarrow B \rightarrow B^*$, though we could equally well study the chain $B \rightarrow A \rightarrow A^*$ with similar conclusions. We would like to know, abstractly, how hard it is to recover the 3-party state ρ_{ABB^*} from the two-party state ρ_{AB} with a Markov map $\mathcal{R}_{B \rightarrow BB^*}$. The bulk regions dual to the density matrices ρ_{AB} , ρ_{BB^*} , and ρ_{ABB^*} are indicated with shading and crosshatching in figure 5. In particular, any sufficiently small tubular neighborhood of the “jagged” surfaces indicated in figure 5 with thick, wavy lines is contained in the entanglement wedge of ρ_{ABB^*} , but no tubular neighborhood is contained entirely within the entanglement wedges of either ρ_{AB} or ρ_{BB^*} .

There is by now very good reason to believe that geometric connections in bulk states are sourced, in a meaningful sense, by large amounts of boundary entanglement. This heuristic, originally advocated by Van Raamsdonk in [40] using the Ryu-Takayanagi formula, is the fundamental principle underlying many important developments such as

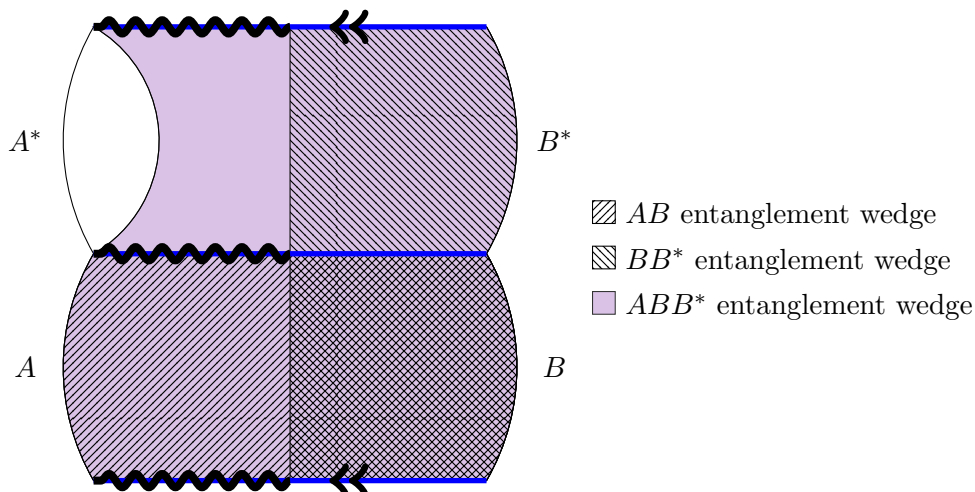


Figure 5. A sketch of the canonical purification for the density matrix of two intervals in the AdS_3 vacuum; the top and bottom lines are identified, making the entire geometry a two-boundary wormhole. The ABB^* entanglement wedge is shaded in violet, and the AB and BB^* entanglement wedges are indicated with crosshatching. The blue lines are minimal surfaces for AB and A^*B^* . Jagged curves have been imposed over two minimal surfaces whose tubular neighborhoods are visible to ABB^* but not to AB or BB^* ; we argue in the text of the paper that these surfaces must be supported by boundary entanglement contributing to $I(A : B^*|B)$.

“ER = EPR” [52] and entanglement wedge reconstruction behind black hole horizons [53–55].¹⁵ If we apply this heuristic to the situation elaborated in the preceding paragraph, it seems natural to conclude that the boundary entanglement sourcing the geometric connections across the jagged surfaces of figure 5 is visible to ρ_{ABB^*} , but not to ρ_{AB} or ρ_{BB^*} . This suggests that we could never hope to reproduce that entanglement by acting on ρ_{AB} with a Markov channel $\mathcal{R}_{B \rightarrow BB^*}$ — the entanglement sourcing the smooth geometries of the jagged surfaces isn’t already present in ρ_{AB} , and it can’t be added to the state with a channel that doesn’t access the A subsystem. *The very existence of the jagged surfaces in figure 5 precludes a perfect Markov recovery* $\rho_{ABB^*} = \mathcal{R}_{B \rightarrow BB^*}(\rho_{AB})$. This, in turn, guarantees a nontrivial Markov gap by inequality (2.8).

The preceding discussion suggests that each boundary in the entanglement wedge cross-section necessarily makes some contribution to the Markov gap, because every boundary in the entanglement wedge cross-section is attached to a “jagged surface” supported by entanglement contributing to $I(A : B^*|B)$. In our detailed analysis below, we will focus our attention on “corners,” geometric structures formed by the intersection of a jagged surface and an entanglement wedge cross-section. As we will see, each corner is supported by entanglement making an irreducible contribution to the Markov gap, even in the limit that the length of the jagged surface itself vanishes. In that sense, the contribution of these corners to the Markov gap can be viewed as more fundamental than the entanglement supporting the jagged surface itself.

¹⁵The “geometric connection” appearing in the behind-the-horizon reconstruction problem is most apparent in the doubly holographic model of entanglement islands introduced in [56]. See also [57] for a singly holographic realization.

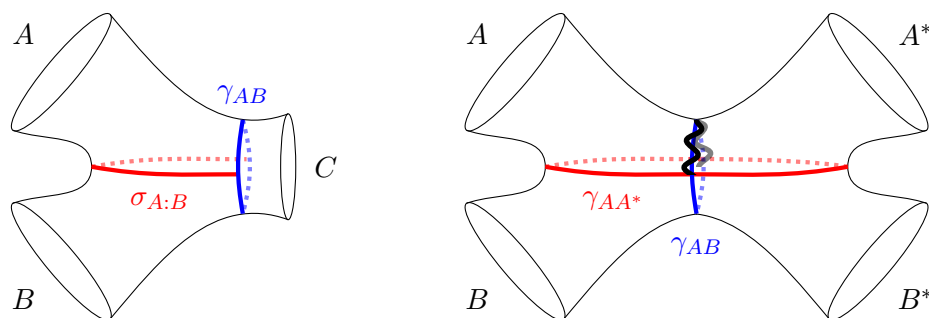


Figure 6. The left side of this figure shows a three boundary wormhole with ρ_{AB} having a connected entanglement wedge; the AB extremal surface is marked γ_{AB} , and the entanglement wedge cross-section is labeled $\sigma_{A:B}$. The right side shows a time-slice of the spacetime corresponding to the canonical purification. On the canonical purification, we have marked with a jagged curve the portion of γ_{AB} whose tubular neighborhoods are visible to ρ_{ABB^*} but not to ρ_{AB} or ρ_{BB^*} .

To argue for this intuitive principle, we will examine a different canonical purification than the one already considered. Let ρ_{AB} be the density matrix of two asymptotic boundaries of a three-boundary wormhole geometry, with the third asymptotic boundary being sufficiently small that the ρ_{AB} entanglement wedge is connected. A static time-slice of this geometry and a static time-slice of its canonical purification are sketched in figure 6.

There are two important features of this canonical purification that differ from the one sketched in figure 5. First, note that there is only *one* jagged surface, as opposed to the two that appeared in the canonical purification of two intervals. So we observe, immediately, that boundaries of the entanglement wedge cross-section are not necessarily in one-to-one correspondence with jagged surfaces. Second, in this case, the jagged surface is compact — and, indeed, one can tweak the moduli of the three-boundary wormhole so that the area of the jagged surface is arbitrarily small.

Still, the very existence of this surface, no matter how small it might be, requires the presence of some boundary entanglement contributing to $I(A : B^*|B)$. We can therefore interpret inequality (1.12), which asserts that at least in certain AdS₃ states every boundary of the entanglement wedge cross-section gives a finite contribution to the Markov gap, as suggesting that the geometric connection through a *corner* where a jagged surface intersects an entanglement wedge cross-section must be supported by a chunk of boundary entanglement that is, at least in some coarse sense, quantified by the lower bound on the Markov gap.

We emphasize that thus far, everything we have discussed in this subsection is entirely heuristic. The rest of the paper is dedicated to presenting concrete, technical results that elaborate and reinforce the ideas presented in this subsection. The first of these, presented in the following section, is a proof of the inequality (1.12) whose significance was explained in the introduction. We will also take up the question of Markov recovery maps again explicitly in section 5, where we construct some models of Markov recovery processes using fixed area states and compute the relevant fidelities using the gravitational path integral.

3 Lower bounds in pure AdS₃ gravity

In holographic theories of quantum gravity, the entanglement entropies of time-symmetric states with no bulk matter can be computed using the Ryu-Takayanagi formula.¹⁶ The reflected entropies of such states can be computed using the Dutta-Faulkner formula (1.5). The Markov gap $S_R - I$ is then computable as an area difference between locally minimal surfaces that satisfy certain global homology constraints, with all of the surfaces lying in a single spacelike geometry.¹⁷ Obtaining universal lower bounds for the Markov gap, like the one we claimed in the introduction (inequality (1.12)), is then reduced to an exercise in Riemannian geometry.

The problem we have set for ourselves becomes even simpler if we restrict our attention to AdS gravity in three bulk dimensions. Time-symmetric slices of asymptotically AdS₃ spacetimes are hyperbolic 2-manifolds, about which there is an extensive mathematical literature. The minimal surfaces whose areas come into the calculation of $S_R - I$ are all geodesics — or geodesic segments — that lie on these hyperbolic 2-manifolds. It is this considerable simplification that will allow us to prove the universal lower bound claimed in the introduction.¹⁸

$$S_R(A : B) - I(A : B) \geq \frac{\log(2)\ell_{\text{AdS}}}{2G_N} \times (\# \text{ of cross-section boundaries}) + o\left(\frac{1}{G_N}\right). \quad (3.1)$$

In subsection 3.1, we introduce a fundamental trigonometric identity relating the lengths of the sides of a right-angled hyperbolic pentagon. In subsection 3.2, we show how the pentagon identity can be used to prove inequality (3.1) in two special cases: (i) A and B are equal-time intervals in the AdS₃ vacuum, and (ii) A and B are asymptotic boundaries of a genus-zero, three-boundary wormhole. In subsection 3.3, we give the proof of inequality (3.1) in complete generality for all time-symmetric states in pure AdS₃ gravity.

Generalizations for non-time-symmetric states, theories with matter, and higher dimensions are addressed in section 4.

3.1 Right angled hyperbolic pentagons

We will take as our model of hyperbolic space the Poincaré disk, with metric

$$ds^2 = \frac{4}{(1-r^2)^2} (dr^2 + r^2 d\theta^2) \quad (3.2)$$

and coordinate ranges $r \in [0, 1), \theta \in [0, 2\pi)$. In choosing this model, we have set the radius of curvature to one; factors of the radius of curvature in all of our expressions can

¹⁶This was argued in great generality by Lewkowycz and Maldacena in [9]. The main weakness of the argument is that it assumes the gravitational path integrals that compute the integer Rényi entropies are dominated by a single family of replica-symmetric saddles. Luckily, this is only expected to fail at “phase transitions” where multiple families of saddles (and therefore multiple minimal surfaces) vie for dominance. See [58, 59] for some calculations in this regime.

¹⁷In the introduction, we referred to these surfaces as “KRT surfaces” and “RT surfaces”, a terminology that we will take up again in subsection 3.2.

¹⁸We remind the reader that the “# of cross-section boundaries” appearing in this inequality is the number of endpoints of the cross section.

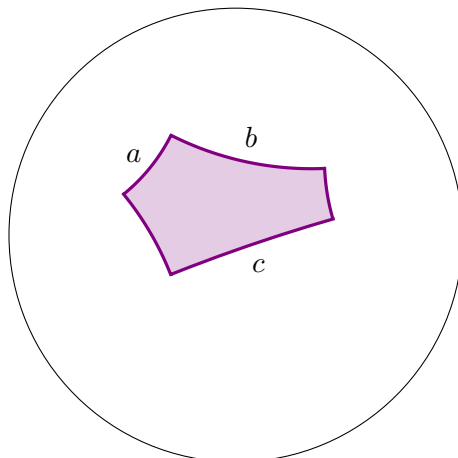


Figure 7. An example of a right-angled pentagon in the hyperbolic disk. Two adjacent sides have been labeled a and b , and the unique non-adjacent side has been labeled c .

be restored by dimensional analysis. A *right-angled hyperbolic pentagon* — see figure 7 for an example — is a compact, convex set in the Poincaré disk bounded by five geodesic segments that meet sequentially at right angles. We will also use the term “right-angled hyperbolic pentagon” to refer to a pentagon on any Riemannian manifold that is isometric to a right-angled pentagon in the Poincaré disk.

Right-angled hyperbolic pentagons satisfy a law of cosines relating their side lengths. If a and b are two adjacent sides, and c is the unique non-adjacent side (again see figure 7), then the side lengths satisfy the identity¹⁹

$$\sinh(a) \sinh(b) = \cosh(c). \tag{3.3}$$

Using this identity, we will compute a universal lower bound on the quantity $a + b - c$ for a right-angled hyperbolic pentagon; in the following subsections, this lower bound will be used to bound the Markov gap.

We may write $a + b - c$ using equation (3.3) as

$$a + b - c = a + b - \operatorname{arccosh}[\sinh(a) \sinh(b)]. \tag{3.4}$$

We’d like to minimize this quantity over all permitted values of a and b . It will actually be conceptually easier to write the identity in terms of $u \equiv \sinh(a)$ and $v \equiv \sinh(b)$,

$$a + b - c = \operatorname{arcsinh}(u) + \operatorname{arcsinh}(v) - \operatorname{arccosh}(uv), \tag{3.5}$$

and minimize over all permitted values of u and v . Equation (3.3) requires that the product uv is in the image of \cosh , so we must have $uv \geq 1$, and the condition $a, b \geq 0$ restricts

¹⁹There are many ways of proving this and other trigonometric identities on hyperbolic manifolds. Our favorite methods come from the field of *hyperbolic line geometry*, in which geodesics in three-dimensional hyperbolic space are identified with the $\operatorname{SL}_2(\mathbb{C})$ matrices that describe π -rotations about those geodesics, and geometric relations among geodesics are expressed as algebraic relations among the corresponding matrices. A very nice elaboration of this theory can be found in [60]. Our identity (3.3) is given there as equation (2) on page 82, with a sign difference due to an orientation convention adopted therein.

u and v to be individually nonnegative; this is the parameter space over which we will minimize equation (3.5).

Computing the gradient of the right-hand side of (3.5), it is straightforward to show that it vanishes only at $u, v = \pm i$, which is not in the allowed parameter space. So the minimum of $a + b - c$ must be realized on the boundary of the parameter space, either on the curve $v = 1/u$ or in a limit as u or v becomes large. Setting $v = 1/u$, we have

$$a + b - c = \operatorname{arcsinh}\left(\frac{1}{u}\right) + \operatorname{arcsinh}(u), \quad (3.6)$$

which is minimized at $u = 1$ with minimum value $2 \log(1 + \sqrt{2})$. In the limit as one of the parameters becomes large — say u without loss of generality, since the expression is symmetric in u and v — we have

$$a + b - c = \operatorname{arcsinh}(v) - \log(v), \quad (3.7)$$

which is minimized in the limit $v \rightarrow \infty$ with minimum $\log(2)$.

As the two candidate minima are $2 \log(1 + \sqrt{2})$ and $\log(2)$, and $\log(2)$ is the smaller of the two, we conclude with the universal lower bound

$$a + b - c \geq \log(2). \quad (3.8)$$

3.2 Intervals and asymptotic regions

We will now use the pentagon inequality (3.8) to derive the Markov gap bound (3.1) for two special cases, as a warmup for the general proof in subsection 3.3.

Consider first the case where A and B are two equal-time intervals in the AdS_3 vacuum. We will assume that A and B are sufficiently large that their entanglement wedge is in the connected phase; otherwise, $S_R(A : B)$ and $I(A : B)$ both vanish at order $1/G_N$, there is no entanglement wedge cross-section and thus no cross-section boundary, and inequality (3.1) is trivial. The static time slice is isometric to the Poincaré disk, and the minimal surfaces are straightforward to compute. We have sketched them in figure 8a, where γ_A and γ_B are the Ryu-Takayanagi surfaces of A and B , γ_{AB} is the Ryu-Takayanagi surface of AB , and $\sigma_{A:B}$ is the entanglement wedge cross-section.

As discussed in the introduction, portions of γ_{AB} and $\sigma_{A:B}$ can be combined to form candidate Ryu-Takayanagi surfaces for A and B that contain right-angled “kinks.” We call these surfaces KRT surfaces — for “kinked Ryu-Takayanagi surface” — and denote them by $\text{KRT}(A)$ or by ξ_A (respectively $\text{KRT}(B)$ and ξ_B). These are sketched in figure 8b.

Using the Ryu-Takayanagi formula and the Dutta-Faulkner formula, the Markov gap can be written as

$$S_R(A : B) - I(A : B) = \frac{\text{area}(\xi_A) - \text{area}(\gamma_A)}{4G_N} + \frac{\text{area}(\xi_B) - \text{area}(\gamma_B)}{4G_N} + o\left(\frac{1}{G_N}\right). \quad (3.9)$$

Since ξ_A is homologous to γ_A , and γ_A is minimal within its homology class by the assumptions of the Ryu-Takayanagi formula, each term on the right-hand side of (3.9) is individually nonnegative. We can obtain a tighter lower bound on each term by tiling the homology regions between the KRT surfaces and the RT surfaces with right-angled pentagons.

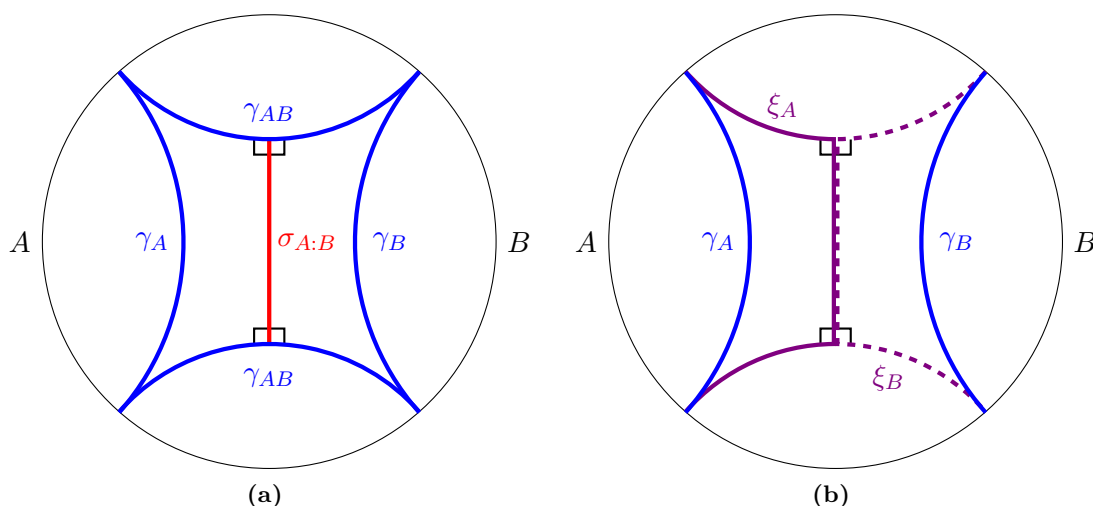


Figure 8. (a) The minimal surfaces relevant for computing $S_R - I$ for two equal-time intervals in the connected phase of the AdS_3 vacuum. (b) The same configuration, but where the disjoint union $\gamma_{AB} \vee \sigma_{A:B} \vee \sigma_{A:B}$ has been redrawn in terms of two “KRT” surfaces ξ_A and ξ_B . The surface ξ_A has been drawn with a solid curve, while ξ_B has been drawn with a dashed curve.

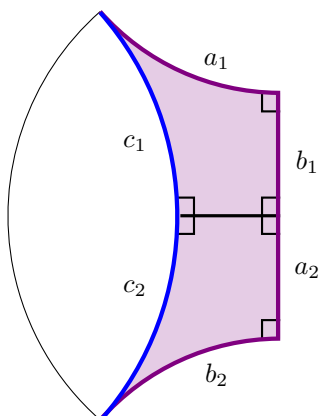


Figure 9. The homology region between the RT surface γ_A and corresponding KRT surface ξ_A , previously depicted in figure 8b, has been tiled by two degenerate right-angled pentagons by drawing the minimal surface connecting γ_A to the entanglement wedge cross-section. We have $a_j + b_j - c_j \geq \log(2)$ for $j = 1, 2$.

In figure 9, we have sketched the homology region between ξ_A and γ_A , and decomposed it into the union of two right-angled pentagons. This decomposition is constructed by drawing the minimal curve between γ_A and $\sigma_{A:B}$, which meets those curves at right angles and subdivides the homology region into two tiles. It may not be obvious that the tiles in this decomposition are right-angled pentagons, as they appear to have only four sides. However, each tile has a *degenerate* side “at infinity”; the tiles are degenerate right-angled pentagons that can be obtained in a limit as one side of an ordinary right-angled pentagon

shrinks and goes off to infinity to become a single vertex. The bound (3.8) still holds in this limit, where a and c (or b and c) are the two sides that meet at infinity.²⁰

The difference in area between ξ_A and γ_A can be written as the sum over one $(a+b-c)$ term for each pentagon in the tiling. Using the labels of figure 9, the identity is

$$\text{area}(\xi_A) - \text{area}(\gamma_A) = (a_1 + b_1 - c_1) + (a_2 + b_2 - c_2). \quad (3.10)$$

Applying inequality (3.8) gives

$$\text{area}(\xi_A) - \text{area}(\gamma_A) \geq 2 \log(2). \quad (3.11)$$

Performing the same calculation for $(\text{area}(\xi_B) - \text{area}(\gamma_B))$ and plugging the lower bounds back into equation (3.9), we obtain the inequality

$$S_R(A : B) - I(A : B) \geq \frac{\log(2)}{G_N}. \quad (3.12)$$

This matches our claimed general inequality (3.1), because for the case of two equal-time intervals in the AdS₃ vacuum, the entanglement wedge cross-section has two boundaries — one where it meets each component of γ_{AB} . Note that this minimum is saturated only in the limit where the two intervals A and B become large; this follows from the discussion of subsection 3.1.

As a second example, consider the case where A and B are two asymptotic boundaries of a genus-zero, three-boundary wormhole. We will assume that the horizon of the third boundary is sufficiently small that the entanglement wedge of AB is in the connected phase; otherwise, as in the case of two intervals, there is no entanglement wedge cross-section and inequality (3.1) is trivial. The minimal surfaces for this setup are sketched in figure 10 using the same notation as for the two-interval case in figure 8a.

In this case, the surfaces γ_A, γ_B , and γ_{AB} bound a “pair of pants” geometry, which is subdivided into two pieces by the cross-section $\sigma_{A:B}$. These two pieces are the homology regions bounded by the KRT and RT surfaces for A and B ; they are sketched in figure 11. We can further subdivide each of those pieces into two right-angled hyperbolic pentagons by drawing the minimal geodesics between γ_{AB} and γ_A (or γ_B) and between $\sigma_{A:B}$ and γ_A (or γ_B). This pentagonal tiling of the homology regions is also sketched in figure 11.²¹ As in the two-interval case discussed previously, the quantity $(\text{area}(\xi_A) - \text{area}(\gamma_A))$ can be written as a sum of two $(a+b-c)$ terms for the right-angled pentagons tiling the homology region between KRT(A) and RT(A); similarly for the quantity $(\text{area}(\xi_B) - \text{area}(\gamma_B))$. Even

²⁰Even though the two sides of the pentagon that meet at infinity each have infinite length, the divergences in the lengths cancel and the difference in their areas is well-defined and independent of cutoff procedure. This is a special case of the general result proved in [61]. In particular, choosing a cutoff procedure where we regulate the degenerate pentagon by taking it to be a limit of non-degenerate pentagons, one can show that degenerate right-angled pentagons satisfy inequality (3.8).

²¹The fact that these minimal geodesics exist is an elementary feature of the pair of pants geometry. In the general proof of subsection 3.3, we will prove the existence of pentagonal tilings for a much more general class of homology regions; the reader wondering about the existence of pentagonal tilings even for the three-boundary wormhole is encouraged to look ahead to the general proof.

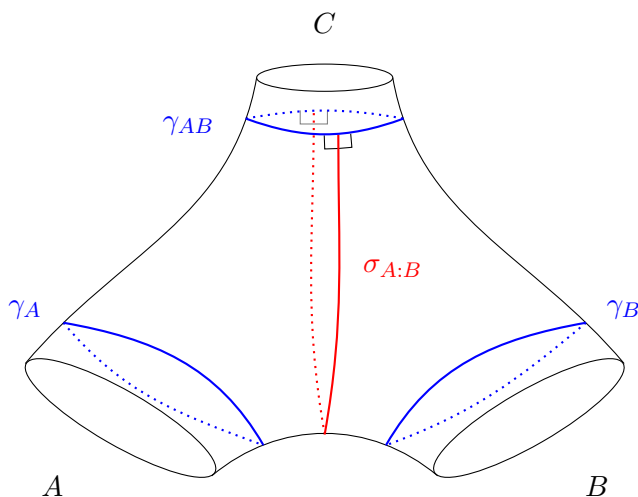


Figure 10. A three-boundary wormhole geometry with boundaries A , B , and C such that the state on AB is in its connected phase. The A and B minimal surfaces are labeled γ_A and γ_B , the AB minimal surface is labeled γ_{AB} , and the entanglement wedge cross-section is labeled $\sigma_{A:B}$.

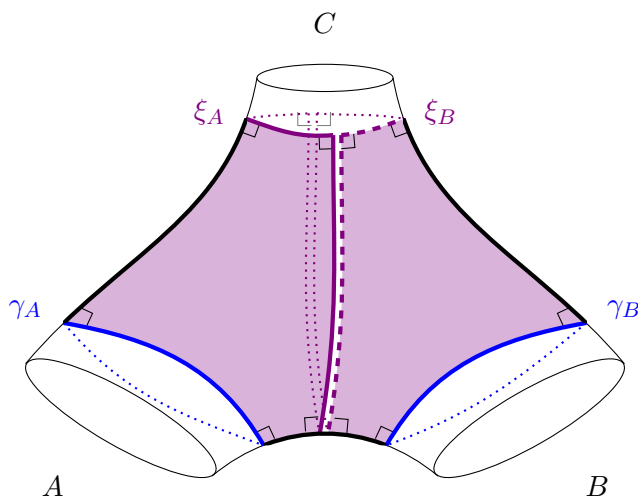


Figure 11. The KRT and RT surfaces for the three-boundary wormhole originally sketched in figure 10. We have also sketched the tiling of each KRT-RT homology region by right-angled pentagons, though we have only shaded the pentagons appearing on the front-facing side of the wormhole for the sake of visual clarity.

though the three-boundary wormhole is not globally isometric to the Poincaré disk, it is *locally* isometric to the Poincaré disk, and so each pentagon in the tiling of figure 11 is a right-angled hyperbolic pentagon. Once again applying the pentagon inequality (3.8), we obtain the bound

$$S_R(A : B) - I(A : B) \geq \frac{\log(2)}{G_N}, \quad (3.13)$$

which matches the general inequality (3.1), since the entanglement wedge cross-section has two boundaries (see again figure 10).

The general proof, for arbitrary global states and arbitrary subregions A and B within those states, is presented in the following subsection. The general idea is to show that each component of a KRT surface can be homotopically deformed to a geodesic such that the homotopy region can be tiled with right-angled hyperbolic pentagons, with each pentagon in the tiling corresponding to a single kink in the KRT surface. The homotopy tilings will be constructed so that the pentagon inequality (3.8) guarantees that the area difference between a KRT component and its homotopic geodesic is lower-bounded by $\log(2)$ times the number of kinks in the KRT component. After every component of a KRT surface has been homotopically deformed to a geodesic, the resulting “deformed” surface is smooth and is still in the same homology class as the “true” RT surface, and thus has area no greater than the true RT surface. Combining these observations will give us

$$\text{area(KRT)} - \text{area(RT)} \geq \log(2) \times (\# \text{ of right-angled kinks}), \quad (3.14)$$

which, upon relating the total number of KRT kinks to the total number of cross-section boundaries, will reproduce inequality (3.1).

This sketch will be explained in much more detail over the course of the proof.

3.3 The general proof

The entanglement wedge cross-section was originally defined in [21, 22] as a codimension-2 surface that splits the entanglement wedge into two homology regions, such that one homology region belongs to each of the two boundary regions A and B . It will be useful, for our proof, to reframe this as a statement about minimal surfaces. Formally, we say that the entanglement wedge cross-section $\sigma_{A:B}$ induces a bipartition of the minimal surface γ_{AB} into two pieces — call them $\gamma_{AB}^{(A)}$ and $\gamma_{AB}^{(B)}$ — such that the following homology equivalences hold:

$$\gamma_A \sim_{\text{homology}} \gamma_{AB}^{(A)} \cup \sigma_{A:B}, \quad (3.15)$$

$$\gamma_B \sim_{\text{homology}} \gamma_{AB}^{(B)} \cup \sigma_{A:B}. \quad (3.16)$$

The surfaces on the right-hand sides of these equations are what we have been calling the “KRT surfaces” ξ_A and ξ_B .

On a general hyperbolic 2-manifold, for general boundary subregions A and B , what can we say about the KRT surfaces for A and B ? First, trivially, each KRT surface is homologous to the corresponding RT surface. Second, each KRT surface is piecewise smooth, with each smooth portion of the surface being a geodesic segment. Third, the only non-smoothness in a KRT surface comes in the form of right-angled kinks. This last claim follows from the fact that $\sigma_{A:B}$ and γ_{AB} are individually smooth, so the only non-smoothness in a KRT surface comes from an intersection between $\sigma_{A:B}$ and γ_{AB} ; such intersections must be orthogonal, by the assumption that $\sigma_{A:B}$ has minimal length.²² Crucially, the right angles of these kinks always face “inward” toward the homology region; i.e., the homology region between a KRT surface and its corresponding RT surface is

²²If the entanglement wedge cross-section did not meet γ_{AB} at a right angle, then it would be possible to deform the cross-section to be “more orthogonal,” making its area smaller.

convex. Finally, the KRT surface is simple, i.e., it does not contain any self-intersections aside from the right-angle intersections of its geodesic segments.

Our goal is to use the properties listed in the preceding paragraph to prove that the area difference between a KRT surface and the corresponding RT surface is lower-bounded by $\log(2)$ times the number of kinks in the KRT surface. Because the KRT surface is non-self-intersecting, each of its components must be either (a) a kinked geodesic with two endpoints at infinity (as in figure 9), (b) a kinked closed geodesic (as in figure 11), (c) a smooth geodesic with two endpoints at infinity, or (d) a smooth closed geodesic. We label these four types of surfaces “KA” (for “kinked asymptotic”), “KL” (for “kinked loop”), “A” (for “asymptotic”), and “L” (for “loop”). The component decomposition of a KRT surface can be written schematically as

$$\text{KRT} = \cup_{\alpha} \text{KA}_{\alpha} \cup_{\beta} \text{KL}_{\beta} \cup_j \text{A}_j \cup_k \text{L}_k. \tag{3.17}$$

We will soon show that each kinked asymptotic geodesic KA_{α} is homotopic to a smooth asymptotic geodesic A_{α} with the same endpoints whose area is smaller by at least $\log(2)$ times the number of kinks in KA_{α} ; this area difference will follow from a pentagonal tiling of the homotopy region, where each kink in KA_{α} becomes a vertex of a right-angled pentagon, and the number of pentagons is equal to the number of kinks. We will show, similarly, that each kinked geodesic loop KL_{β} is homotopic to a smooth geodesic loop L_{β} with area smaller by at least $\log(2)$ times the number of kinks in KL_{β} . This will give us

$$\text{KRT} \sim_{\text{homotopy}} \cup_{\alpha} \text{A}_{\alpha} \cup_{\beta} \text{L}_{\beta} \cup_j \text{A}_j \cup_k \text{L}_k. \tag{3.18}$$

But because homotopy is stronger than homology, this will imply

$$\text{KRT} \sim_{\text{homology}} \cup_{\alpha} \text{A}_{\alpha} \cup_{\beta} \text{L}_{\beta} \cup_j \text{A}_j \cup_k \text{L}_k. \tag{3.19}$$

Because homology is an equivalence relation, and because we have $\text{KRT} \sim_{\text{homology}} \text{RT}$, we obtain

$$\text{RT} \sim_{\text{homology}} \cup_{\alpha} \text{A}_{\alpha} \cup_{\beta} \text{L}_{\beta} \cup_j \text{A}_j \cup_k \text{L}_k. \tag{3.20}$$

The fact that the RT surface is minimal in its homology class gives

$$\text{area}(\text{RT}) \leq \text{area}[\cup_{\alpha} \text{A}_{\alpha} \cup_{\beta} \text{L}_{\beta} \cup_j \text{A}_j \cup_k \text{L}_k] \leq \text{area}(\text{KRT}) - \log(2) \times (\text{total } \# \text{ of kinks}). \tag{3.21}$$

From this formula immediately follows the universal bound (3.1), because the total number of kinks in each KRT surface ($\text{KRT}(A)$ or $\text{KRT}(B)$) is equal to the number of boundaries of the entanglement wedge cross-section. We therefore have

$$\begin{aligned} S_R(A : B) - I(A : B) &= \frac{\text{KRT}(A) - \text{RT}(A)}{4G_N} + \frac{\text{KRT}(B) - \text{RT}(B)}{4G_N} + o\left(\frac{1}{G_N}\right) \\ &\geq \frac{\log(2) \times (\# \text{ of } A \text{ kinks})}{4G_N} + \frac{\log(2) \times (\# \text{ of } B \text{ kinks})}{4G_N} + o\left(\frac{1}{G_N}\right) \\ &= \frac{2 \log(2) \times (\# \text{ of cross-section boundaries})}{4G_N} + o\left(\frac{1}{G_N}\right), \end{aligned} \tag{3.22}$$

which is exactly the bound claimed in (3.1).

We now complete the proof of inequality (3.1) by proving three lemmas. In the first lemma, we prove that any kinked asymptotic geodesic in the Poincaré disk has the aforementioned “homotopy shrinking property,” i.e., that it is homotopic to a smooth asymptotic geodesic whose area is smaller by at least $\log(2)$ times the number of kinks. In the second and third lemmas, we use the fact that every hyperbolic 2-manifold is universally covered by the Poincaré disk to prove the analogous statements for kinked asymptotic geodesics and kinked geodesic loops in arbitrary hyperbolic 2-manifolds.

Lemma 1. *Let \widetilde{KA} be a non-self-intersecting, piecewise-geodesic curve in the Poincaré disk with the following properties: (see figure 12a for a sketch)*

- (i) *Each geodesic segment meets the next segment at a right angle.*
- (ii) *All right angles “point in the same direction” in the sense that an ant crawling along \widetilde{KA} will have to turn in the same direction each time it encounters a right angle.*
- (iii) *\widetilde{KA} has well-defined endpoints at infinity in the sense that there is a smooth asymptotic geodesic $\widetilde{A}(t)$ and a parametrization t of \widetilde{KA} such that the hyperbolic distance $|\widetilde{A}(t) - \widetilde{KA}(t)|$ remains bounded in the limits $t \rightarrow \pm\infty$.*

Then:

- (1) *The geodesic \widetilde{A} does not intersect \widetilde{KA} , and the homotopy region between the two is convex.*
- (2) *The homotopy region between \widetilde{KA} and \widetilde{A} can be tiled by hyperbolic right-angled pentagons, with one pentagon for each kink in \widetilde{KA} .*
- (3) *The area difference between \widetilde{KA} and \widetilde{A} is lower-bounded by $\log(2)$ times the number of kinks in \widetilde{KA} .*

Proof. We will break the proof up in terms of claims (1)-(3) of the lemma statement.

- (1) First, let’s assign \widetilde{KA} an orientation so that an ant crawling along \widetilde{KA} will have to turn right each time it encounters a kink. This orientation is sketched in figure 12b. We define a map ϕ from geodesic segments of \widetilde{KA} to oriented segments of the boundary of the Poincaré disk as follows. Each geodesic segment can be extended to a full, asymptotic geodesic that has a “future” and “past” endpoint with respect to the chosen orientation of \widetilde{KA} . If η labels a geodesic segment on \widetilde{KA} , then $\phi(\eta)$ will be the clockwise-oriented boundary segment that starts at the “past” endpoint of the extension of η and ends at the “future” endpoint of the extension of η . This map is also sketched in figure 12b.

The geodesic segments of \widetilde{KA} are naturally ordered with respect to its orientation; let’s call the ordered, possibly finite sequence of segments $\{\eta_j\}$. The map ϕ sends this sequence to an ordered sequence of clockwise-oriented boundary segments $\{\phi(\eta_j)\}$. Any two consecutive segments $\phi(\eta_j)$ and $\phi(\eta_{j+1})$ will have an intersection bounded by

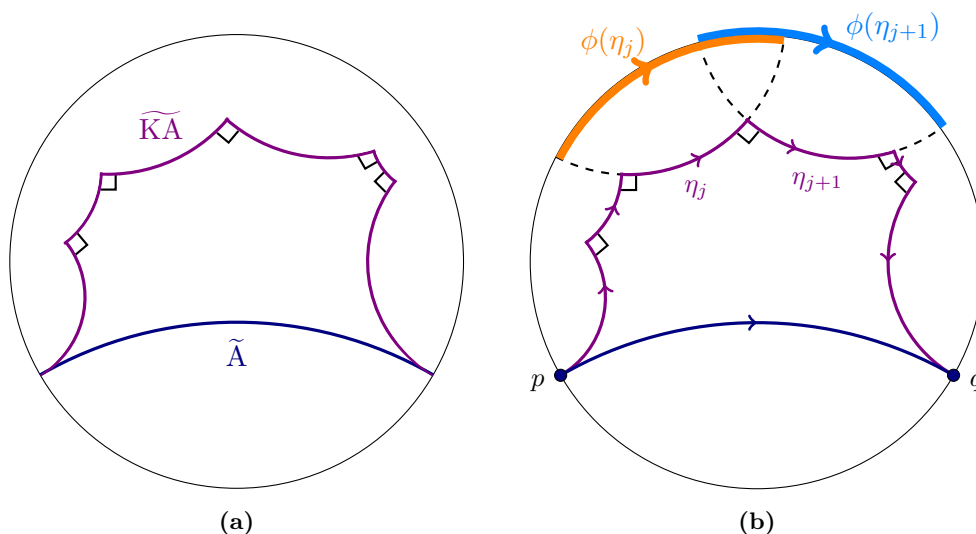


Figure 12. (a) \widetilde{KA} is a piecewise-geodesic curve in the Poincaré disk satisfying properties (i)-(iii) of lemma 1. \widetilde{A} is the unique smooth geodesic that has the same boundary endpoints as \widetilde{KA} . The curve \widetilde{KA} we have drawn here has finitely many geodesic segments, but the number of geodesic segments could in principle be infinite. (b) A choice of orientation on \widetilde{KA} , together with a visual representation of the map ϕ from geodesic segments to boundary segments that is defined in the proof of lemma 1. The boundary points of \widetilde{A} have been labeled p and q .

the clockwise endpoint of $\phi(\eta_j)$ and the counterclockwise endpoint of $\phi(\eta_{j+1})$ — again consult figure 12b.

Because \widetilde{KA} has a well-defined future endpoint q — the future endpoint of the smooth geodesic $\widetilde{A}(t)$ — the clockwise endpoints of the segments $\phi(\eta_j)$ must tend to q in the limit $j \rightarrow \infty$.²³ Similarly, because \widetilde{KA} has a well-defined past endpoint p , the counterclockwise endpoints of the segments $\phi(\eta_j)$ must tend to p in the limit $j \rightarrow -\infty$. From this fact, and the fact that \widetilde{KA} is non-self-intersecting, we may conclude that each boundary segment $\phi(\eta_j)$ lies entirely in the clockwise-oriented boundary segment that starts at p and ends at q . Because the boundary segments $\phi(\eta_j)$ translate clockwise as j increases, if the clockwise endpoint of $\phi(\eta_j)$ ever *passes* the endpoint q , then the segments will need to do a full loop around the boundary to be consistent with the $j \rightarrow \infty$ limit; such a loop would necessarily cause \widetilde{KA} to self-intersect in the bulk. An analogous line of reasoning holds for the counterclockwise endpoints.

The fact that the boundary segments $\phi(\eta_j)$ all lie within the clockwise boundary segment $p \rightarrow q$ is equivalent to claim (1), that \widetilde{KA} does not intersect \widetilde{A} — since two geodesics in the Poincaré disk only intersect if their boundary points “alternate” — and that the homotopy region is convex — because each right angle must be oriented “inwards” toward the geodesic \widetilde{A} , as in figure 12a.

²³If \widetilde{KA} only has finitely many segments, then the clockwise endpoint of the last $\phi(\eta_j)$ must be q .

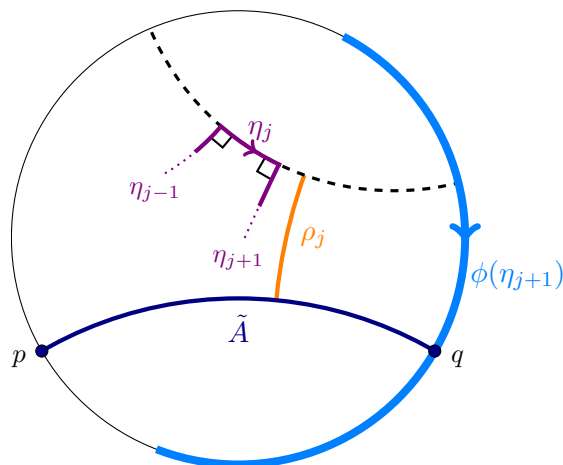


Figure 13. Here \tilde{A} is the smooth geodesic connecting the endpoints of \widetilde{KA} , and $\eta_{j-1}, \eta_j, \eta_{j+1}$ are segments of \widetilde{KA} . The curve ρ_j is the unique geodesic segment orthogonally intersecting both \tilde{A} and the geodesic extension of η_j . If we assume that ρ_j is in the “future” of the segment η_j , as sketched here, then $\phi(\eta_{j+1})$ is not contained within the clockwise boundary segment connecting p and q ; this contradicts part (1) of the proof of lemma 1. An analogous contradiction arises if ρ_j is in the “past” of η_j , which proves that ρ_j intersects the segment η_j .

- (2) For any two non-intersecting asymptotic geodesics in the Poincaré disk, there is a unique third geodesic that intersects each of the first two orthogonally. If the original two geodesics have a common endpoint at the boundary, then the mutually intersecting geodesic is just the degenerate point at infinity.

Let η_j be one of the geodesic segments of \widetilde{KA} . Let ρ_j be the unique geodesic segment that connects the full geodesic extension of η_j to the full geodesic \tilde{A} orthogonally. We will show that the intersection of ρ_j with the geodesic extension of η_j must lie within the segment η_j itself.

Suppose, toward contradiction, that ρ_j intersects the geodesic extension of η_j in a part of the geodesic extension that does not lie in the segment η_j . Suppose further, without loss of generality, that the point where ρ_j intersects the geodesic extension lies in the “future” of the segment η_j with respect to the orientation chosen in part (1) of this proof. This is sketched in figure 13.

The segment η_{j+1} , since it stems off of η_j before reaching ρ_j , must have the property that the boundary segment $\phi(\eta_{j+1})$ extends beyond the boundary segment lying clockwise between p and q . See figure 13 for a sketch. This contradicts part (1) of this proof; we conclude that ρ_j must intersect η_j .

Tiling the homotopy region between \widetilde{KA} and \tilde{A} is now simple. We draw all of the geodesic segments $\{\rho_j\}$ that lie orthogonally between \tilde{A} and the \widetilde{KA} segments $\{\eta_j\}$; the tiles divided by these geodesics, sketched in figure 14, are all hyperbolic right-angled pentagons, possibly with the last tiles in the sequence being degenerate right-angled pentagons if $\{\eta_j\}$ is a finite or half-finite sequence.

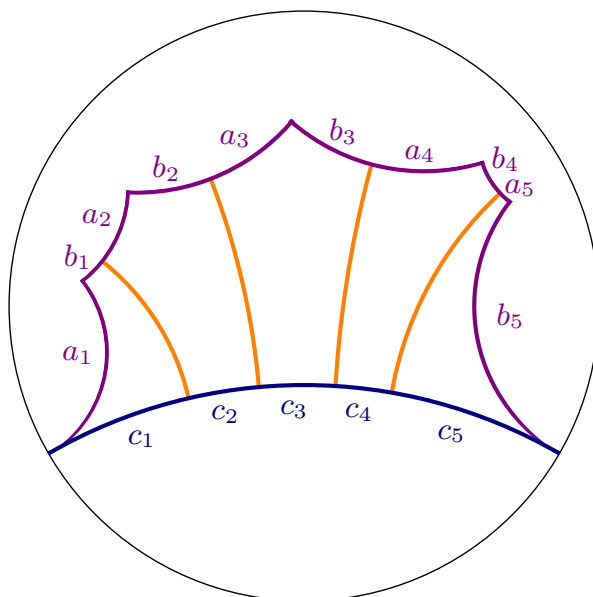


Figure 14. A pentagonal tiling of the homotopy region between the two curves originally sketched in figure 12a, formed by drawing the orthogonal geodesic segments joining each segment of \widetilde{KA} to \widetilde{A} . Every intersection in this figure is right-angled, except for the two intersections at infinity.

- (3) Each pentagon in the tiling we have just constructed has, as one of its vertices, a kink of \widetilde{KA} . The side of the pentagon opposite that vertex is a segment of the geodesic \widetilde{A} . If we label the two sides of the pentagon adjacent to the kink a_j and b_j , and label the opposite side c_j — see again figure 14 — then we have

$$\text{area}(\widetilde{KA}) - \text{area}(\widetilde{A}) = \sum_j (a_j + b_j - c_j). \quad (3.23)$$

Applying inequality (3.8), we obtain the desired bound

$$\text{area}(\widetilde{KA}) - \text{area}(\widetilde{A}) \geq \log(2) \times (\# \text{ of kinks}). \quad (3.24)$$

□

Lemma 2. *Let Σ be a complete hyperbolic 2-manifold without cusps, and let KA be a non-self-intersecting, piecewise-geodesic curve on Σ satisfying properties (i)-(iii) of lemma 1. To condition (iii) we add the requirement that the geodesic defining the endpoints of KA does not have a limit cycle, i.e., both ends of KA genuinely go off “to infinity.”*

Then:

- (1) KA is homotopic to a geodesic A with the same asymptotic endpoints as KA .
- (2) The homotopy region can be tiled by right-angled hyperbolic pentagons, with each kink in KA being a vertex of its own pentagon.
- (3) The area of KA exceeds the area of A by at least $\log(2)$ times the number of kinks in KA .

Proof. It is a fundamental theorem in hyperbolic geometry that every complete hyperbolic 2-manifold is universally covered by the Poincaré disk. Let $\Pi : \mathbb{H}_2 \rightarrow \Sigma$ be a covering map, fix a point p on KA , and choose a point \tilde{p} in the fiber $\Pi^{-1}(p)$. It is a basic fact in the theory of covering spaces that there is a unique lift of KA to the Poincaré disk that passes through \tilde{p} . This lift can be constructed by pulling a small neighborhood of p in KA back to a small neighborhood of \tilde{p} using Π^{-1} , after which the lift extends uniquely away from the pulled-back segment. The fact that no other lifts of KA can pass through \tilde{p} follows from the fact that KA is non-self-intersecting, together with the fact that Π is a local homeomorphism.

Let us denote this lift by $\widetilde{\text{KA}}$. By assumption, KA satisfies conditions (i)-(iii) of lemma 1, and all three of these conditions are preserved under the lift.²⁴ So $\widetilde{\text{KA}}$ satisfies conditions (i)-(iii) of lemma 1, and we may conclude that there is a geodesic $\tilde{\text{A}}$ in \mathbb{H}_2 with the same endpoints as $\widetilde{\text{KA}}$, that is homotopic to $\widetilde{\text{KA}}$ with the homotopy region tiled by right-angled hyperbolic pentagons, such that the area of $\widetilde{\text{KA}}$ exceeds the area of $\tilde{\text{A}}$ by $\log(2)$ times the number of kinks in $\widetilde{\text{KA}}$. Homotopies in the universal cover \mathbb{H}_2 are preserved under projection down to Σ , so KA is homotopic to $\text{A} \equiv \Pi(\tilde{\text{A}})$.

We now show that the length of KA on Σ is the same as the length of $\widetilde{\text{KA}}$ on \mathbb{H}_2 . The covering map Π is a local isometry, so this could only fail to be true if Π were non-injective on $\widetilde{\text{KA}}$. But if Π were non-injective on $\widetilde{\text{KA}}$, then the segment of $\widetilde{\text{KA}}$ lying between any two points with the same Π -image would be mapped to a loop in KA ; since KA was assumed to be non-self-intersecting, this cannot be the case and thus we have $\text{area}(\text{KA}) = \text{area}(\widetilde{\text{KA}})$.

One can also show that Π must be injective on $\tilde{\text{A}}$. Every non-compact geodesic on a cusplless hyperbolic 2-manifold is non-self-intersecting.²⁵ If A were compact, then it could not be homotopic to the asymptotic geodesic KA . So A must be non-self-intersecting, which implies that Π is injective on $\tilde{\text{A}}$, which implies $\text{area}(\text{A}) = \text{area}(\tilde{\text{A}})$. Putting this together gives

$$\text{area}(\text{KA}) - \text{area}(\text{A}) = \text{area}(\widetilde{\text{KA}}) - \text{area}(\tilde{\text{A}}) \geq \log(2) \times \# \text{ kinks}, \quad (3.25)$$

so we may conclude that claims (1)-(3) in the lemma statement hold with $\text{A} \equiv \Pi(\tilde{\text{A}})$.²⁶ \square

²⁴There is one caveat here worth emphasizing: while condition (iii) being satisfied for KA guarantees that condition (iii) is satisfied for $\widetilde{\text{KA}}$ — one can show this by defining the endpoints of $\widetilde{\text{KA}}$ using local lifts of the KA “endpoint-defining geodesic” in a neighborhood of the asymptotic boundary — it will generally not be the case that the geodesic defining the endpoints of $\widetilde{\text{KA}}$ is a lift of the geodesic we have chosen arbitrarily to define the endpoints of KA . There is only one geodesic in \mathbb{H}_2 with the same endpoints as $\widetilde{\text{KA}}$, while there are many geodesics in \mathbb{H}_2 whose projections to Σ have the same endpoints as KA . This is because geodesics ending at points that are *images* of the KA endpoints under the quotient will map to geodesics with the same endpoints as $\widetilde{\text{KA}}$.

²⁵If a non-compact geodesic had a nontrivial loop, then the element of the fundamental group corresponding to that loop would induce an infinite-order isometry of the covering space preserving the lift of the geodesic; the quotient of the lift by the group of covering transformations, which ought to be isomorphic to the original geodesic, would then have to be compact. More details on this “fundamental group \leftrightarrow covering transformation” correspondence are given in the proof of lemma 3.

²⁶In the statement of the lemma, we assumed the existence of a geodesic on Σ with the same asymptotic endpoints as KA (condition iii). This was just a formal way of saying that KA has well-defined endpoints at infinity. As we emphasized in footnote 24, that geodesic has no relation to A , which is special in that it not only has the same boundary endpoints as KA , but is homotopic to KA as well.

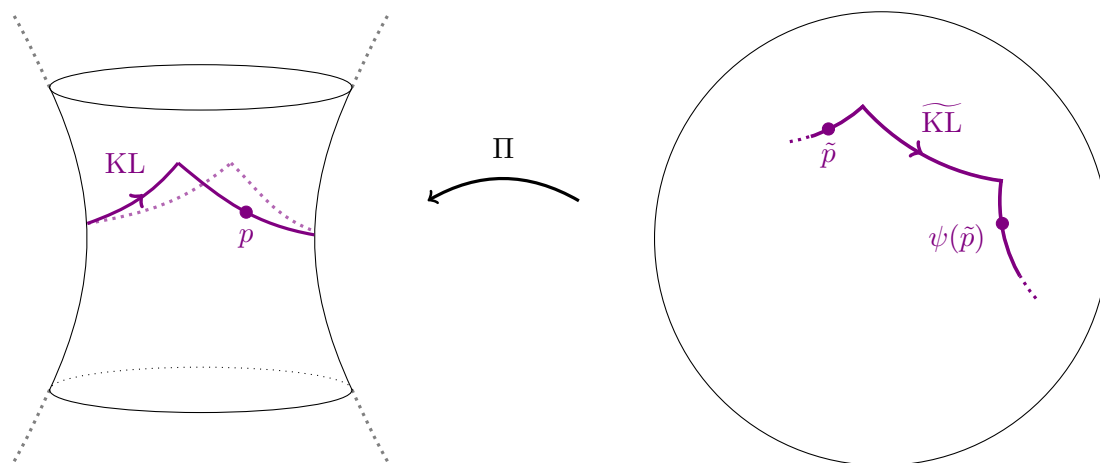


Figure 15. The left-hand side of this figure shows a piece of a hyperbolic 2-manifold containing a piecewise-geodesic loop KL satisfying the conditions of lemma 3. The right-hand side of the figure shows a portion of the unique lift \widetilde{KL} passing through the point \tilde{p} in the universal cover. We have also indicated the action of the covering isometry ψ induced by the loop KL , which maps \widetilde{KL} to itself.

Lemma 3. *Let Σ be a complete hyperbolic 2-manifold without cusps, and let KL be a non-self-intersecting, piecewise-geodesic loop on Σ satisfying conditions (i) and (ii) of lemma 1. Then:*

- (1) KL is homotopic to a unique geodesic loop L .
- (2) The homotopy region can be tiled by right-angled hyperbolic pentagons, with each kink in KL being a vertex of its own pentagon.
- (3) The area of KL exceeds the area of L by at least $\log(2)$ times the number of kinks in KL .

Proof. Claim (1) in this lemma is a special case of a classic theorem in hyperbolic geometry, which states that every homotopy class of non-self-intersecting loops on a hyperbolic manifold has a unique geodesic representative. Our proof of the present lemma works by following the proof of that theorem — specifically the proof presented in section 1.2 of [62] — and adding details as needed to prove claims (2) and (3).

As in the previous lemma, let $\Pi : \mathbb{H}_2 \rightarrow \Sigma$ be a covering of Σ by the Poincaré disk. Fix a point p in the loop KL , and let \tilde{p} be a point in the fiber $\Pi^{-1}(p)$. There is a unique lift of KL that passes through \tilde{p} , which we call \widetilde{KL} . Because conditions (i) and (ii) of lemma 1 are preserved under lifting, \widetilde{KL} is a kinked geodesic in the hyperbolic disk. Because KL is non-self-intersecting, so is \widetilde{KL} . The lifting procedure is sketched in figure 15.

There is a classic correspondence in algebraic topology between elements of the fundamental group of a space and automorphisms of its universal cover. An automorphism of a universal cover $\Pi : \widetilde{M} \rightarrow M$ is a map ψ from \widetilde{M} to itself that satisfies $\Pi \circ \psi = \Pi$. These automorphisms are often called *deck transformations*. It is a theorem in algebraic

topology that any deck transformation of a universal cover is completely determined by its action on a single point.²⁷ So after choosing a point p in the base space and a point \tilde{p} in its fiber $\Pi^{-1}(p)$, a loop γ based at p determines the deck transformation that sends \tilde{p} to the endpoint of the unique path-lift²⁸ $\tilde{\gamma}$ that starts at \tilde{p} .

If we assign KL an orientation, then we can think of it as an element of the fundamental group $\pi_1(\Sigma; p)$. Following the algorithm of the preceding paragraph, the loop KL passing through p , together with our choice of point $\tilde{p} \in \Pi^{-1}(p)$ and our choice of orientation, determines a deck transformation $\psi : \mathbb{H}_2 \rightarrow \mathbb{H}_2$. This deck transformation necessarily maps the lift \widetilde{KL} to itself. The action of ψ on a sample lift is sketched in figure 15.

Because the covering map $\Pi : \mathbb{H}_2 \rightarrow \Sigma$ is not only a local homeomorphism but also a local isometry, the deck transformation ψ must be an isometry of \mathbb{H}_2 to itself. The isometries of \mathbb{H}_2 are categorized into three types: parabolic, elliptic, and hyperbolic.²⁹ See chapter 1 of [64] for an introduction to this classification. The assumption that Σ is smooth rules out the existence of elliptic deck transformations, because these have fixed points in the universal cover that lead to conical defects in the quotient. The assumption that Σ has no cusps rules out the existence of parabolic deck transformations, because parabolic deck transformations have no lower bound on the distance between a point and its image; taking the quotient of \mathbb{H}_2 by a parabolic deck transformation always creates a cusp. We conclude that ψ must be a hyperbolic map from \mathbb{H}_2 to itself.

Hyperbolic isometries of the Poincaré disk have a geodesic axis that is mapped to itself, with one boundary endpoint acting as a source and the other as a sink. The action of a particular hyperbolic isometry is sketched in figure 16. We denote the geodesic axis of ψ by \tilde{L} .

The lift \widetilde{KL} must have well-defined boundary endpoints given by the endpoints to \tilde{L} . This follows from the fact that both are preserved by the isometry; we can pick a fundamental domain for the action of ψ on \tilde{L} , and parametrize \tilde{L} with a parameter t that gives equal time to each image of the fundamental domain. If we parametrize \widetilde{KL} similarly, then the distance $|\widetilde{KL}(t) - \tilde{L}(t)|$ is bounded in the limits $t \rightarrow \pm\infty$. This is because the fundamental domains of \widetilde{KL} and \tilde{L} , being compact, are a bounded distance apart from one another; since ψ is an isometry, that bounded distance is preserved in the limits $t \rightarrow \pm\infty$.

So \widetilde{KL} is a non-self-intersecting, kinked geodesic satisfying conditions (i)–(iii) of lemma 1 — we assumed conditions (i) and (ii), and proved condition (iii) of lemma 1 in the preceding paragraph. We can then apply lemma 1 to show that \widetilde{KL} is homotopic to \tilde{L} and that the homotopy region is tiled by right-angled pentagons. If ρ is any of the geodesic segments connecting \widetilde{KL} to \tilde{L} , then the portion of the tiling lying between ρ and its image $\psi(\rho)$ is a fundamental domain for the tiling with respect to ψ ; this tiling projects down to a tiling of the homotopy region between KL and $L \equiv \Pi(L)$, which proves the lemma. This “fundamental domain tiling” argument is sketched in figure 17. \square

²⁷This follows from the fact that deck transformations are by definition lifts of the covering map, together with the theorem that any two lifts of a map whose domain is connected must agree everywhere if they agree at a single point. This is proved as proposition 1.34 in [63].

²⁸By path-lift we mean we lift only a single copy of γ starting and ending at p , so that $\tilde{\gamma}$ is a compact curve starting at \tilde{p} and ending at some other point in the fiber $\Pi^{-1}(p)$.

²⁹The hyperbolic isometries are sometimes called “loxodromic,” especially in the analogous classification of isometries of \mathbb{H}_3 .

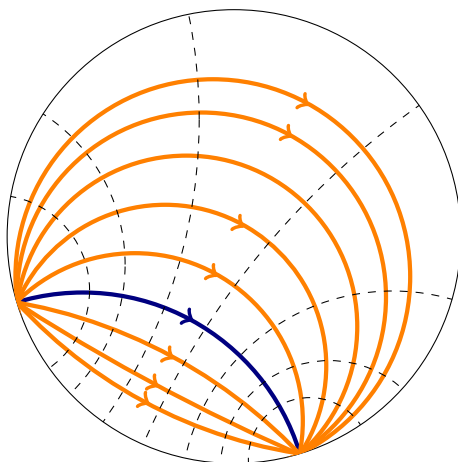


Figure 16. An example of a hyperbolic isometry acting on the Poincaré disk. The blue curve is a geodesic and is called the “axis” of the isometry. The orange curves are invariant curves of the isometry; they are mapped into themselves under a “push” whose direction is indicated by the arrows. The dashed lines are mapped one into the next by the isometry. The invariant curves are spaced apart by about half an AdS radius, as are the dashed lines.

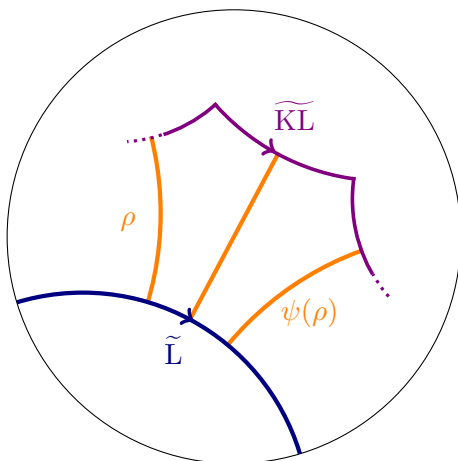


Figure 17. As explained in the proof of lemma 3, the lift \widetilde{KL} gives rise to an isometry ψ with a geodesic axis \widetilde{L} . Lemma 1 tells us the homotopy region between \widetilde{KL} and \widetilde{L} can be tiled by right-angled pentagons by drawing the orthogonal geodesics between \widetilde{L} and each segment of \widetilde{KL} . Here, we have marked one of the geodesics ρ and its isometry image $\psi(\rho)$. The portion of the tiling between these two curves is a fundamental domain for the tiling under the action of ψ .

4 Bulk matter and higher dimensions

It is natural to ask whether inequality (3.1), derived in the previous section for time-symmetric states in pure AdS₃ gravity, has an analogue in more general theories of gravity. In this section, we take up that question on a few fronts.

In subsection 4.1, we give a simple calculation showing that in *any* asymptotically AdS₃ spacetime with a spherically symmetric moment of time symmetry, the area contribution to the Markov gap for two antipodal, equal-size intervals in the moment of time symmetry approaches the universal value $[S_R - I]_{\text{area}} = \log(2)/G_N$ in the limit as the intervals become large. Because the Markov gap is saturated for two vacuum intervals only in this limit (cf. section 3.2), this implies that inequality (3.1) holds at the classical level for arbitrary spherically symmetric perturbations to the state of two intervals in the vacuum, to all orders in perturbation theory. In subsection 4.2, we show that the inequality holds even if the perturbation contains quantum matter, up to some mild assumptions about limits of bulk entanglement entropies; we also comment on some general features of the classical and quantum contributions to perturbations of the Markov gap. In subsection 4.3, we discuss a possible generalization of (3.1) to higher dimensions, where the number of cross-section boundaries is replaced by the codimension-3 area of the cross-section boundary.

We do not address the issue of non-time-symmetric states here. However, we suspect that the techniques that must be developed to generalize inequality (3.1) to states with classical matter and higher dimensions will naturally lend themselves to bounding the Markov gap in general, non-time-symmetric states; we comment on this research direction further in the discussion (section 6.1).

4.1 Spherically symmetric matter in three dimensions

The setting for this subsection will be an arbitrary asymptotically AdS₃ spacetime with a spherically symmetric moment of time symmetry.³⁰ The metric on the moment of time symmetry can be written

$$ds^2 = f(r)dr^2 + r^2d\theta^2 \quad (4.1)$$

with $r \in [0, \infty)$. Vacuum AdS₃ is the spacetime with $f(r) = 1/(1 + r^2)$. The requirement that the spacetime be asymptotically AdS₃ imposes that in a large- r expansion, $f(r)$ differs from the vacuum value $1/(1 + r^2)$ only at order $O(1/r^4)$.

In any such metric, let us consider the Ryu-Takayanagi surface corresponding to a boundary region of angular extent $\Delta\theta$. Because of the spherical symmetry, the geometry of the surface depends only on the angular extent of the boundary region and not on its position; it will also have a reflection symmetry in θ . The coordinate position of such a Ryu-Takayanagi surface is determined by an equation $\theta = h(r)$; the induced metric on this surface is

$$ds^2 = (f(r) + r^2h'(r)^2)dr^2, \quad (4.2)$$

³⁰When we say “spherically symmetric,” we will mean that the spacetime possesses not only a rotational symmetry along an angle θ but also a $\theta \mapsto -\theta$ reflection symmetry.

and the induced volume form is

$$\sqrt{g_{\text{RT}}} = \sqrt{f(r) + r^2 h'(r)^2}. \quad (4.3)$$

If $\theta = h(r)$ is a minimal surface, then it satisfies the Euler-Lagrange equations of motion for the induced volume form, which are

$$\frac{r^2 h'(r)}{\sqrt{f(r) + r^2 h'(r)^2}} = \text{const.} \quad (4.4)$$

Let r_* be the minimal value of r attained by the minimal surface $\theta = h(r)$. At $r = r_*$, we have $h'(r) = \infty$; from this relation, we can fix the constant in (4.4) and solve for $h'(r)^2$ to obtain the equation of motion

$$h'(r)^2 = \frac{r_*^2 f(r)}{r^2 (r^2 - r_*^2)}. \quad (4.5)$$

The total length of a minimal surface contained between a turning point $r = r_*$ and a radial cutoff $r = 1/\epsilon$ can be obtained by plugging (4.5) into the volume form (4.3), multiplying by two (because θ is two-valued as a function of r), and integrating. We obtain the expression

$$L_\epsilon(r_*) = 2 \int_{r_*}^{1/\epsilon} dr r \sqrt{\frac{f(r)}{r^2 - r_*^2}}. \quad (4.6)$$

To compute the mutual information of two antipodal, equal-size intervals A and B , we will need to consider two types of minimal surfaces. The two “disconnected” surfaces individually homologous to A and B , whose turning point we will label r_* , and the “connected” surfaces that are homologous only to the union $A \cup B$, whose turning point we will label r_*^c .³¹ Assuming that the intervals are large enough that the connected surface is globally minimal, the classical contribution to the mutual information is given via the Ryu-Takayanagi formula by

$$I(A : B)_{\text{area}} = \lim_{\epsilon \rightarrow 0} \frac{2L_\epsilon(r_*) - 2L_\epsilon(r_*^c)}{4G_N}. \quad (4.7)$$

Plugging in formula (4.6) for the lengths, we may write this as

$$I(A : B)_{\text{area}} = \frac{1}{G_N} \lim_{\epsilon \rightarrow 0} \left[\int_{r_*}^{1/\epsilon} dr r \sqrt{\frac{f(r)}{r^2 - r_*^2}} - \int_{r_*^c}^{1/\epsilon} dr r \sqrt{\frac{f(r)}{r^2 - (r_*^c)^2}} \right]. \quad (4.8)$$

Partially combining the integrals gives the expression

$$I(A : B)_{\text{area}} = \frac{1}{G_N} \left[\int_{r_*}^{r_*^c} dr r \sqrt{\frac{f(r)}{r^2 - r_*^2}} + \int_{r_*^c}^{\infty} dr r \sqrt{f(r)} \left(\frac{1}{\sqrt{r^2 - r_*^2}} - \frac{1}{\sqrt{r^2 - (r_*^c)^2}} \right) \right], \quad (4.9)$$

where we have removed all ϵ dependence because both integrals in this expression converge.

³¹Explicit formulas for these turning points in terms of the angular extents of the intervals can be obtained, but we will not need them for the present calculation.

Thanks to the spherical symmetry of the metric, the entanglement wedge cross-section is guaranteed to be a line of constant θ with endpoints at $r = r_*^c$. The length of this surface is given by

$$L(\sigma_{A:B}) = 2 \int_0^{r_*^c} \sqrt{f(r)}. \quad (4.10)$$

Using the Dutta-Faulkner formula $S_{R,\text{area}} = L(\sigma_{A:B})/2G_N$, we may now write the classical contribution to the Markov gap as

$$[S_R(A : B) - I(A : B)]_{\text{area}} = \frac{1}{G_N} \left[\int_0^{r_*^c} \sqrt{f(r)} - \int_{r_*}^{r_*^c} dr r \sqrt{\frac{f(r)}{r^2 - r_*^2}} - \int_{r_*^c}^{\infty} dr r \sqrt{f(r)} \left(\frac{1}{\sqrt{r^2 - r_*^2}} - \frac{1}{\sqrt{r^2 - (r_*^c)^2}} \right) \right]. \quad (4.11)$$

In the limit where the two intervals become large and collectively take up the entire boundary, spherical symmetry guarantees the limits $r_* \rightarrow 0$ and $r_*^c \rightarrow \infty$. It is straightforward to take the limit $r_* \rightarrow 0$ in the above expression; the first two integrals cancel in this limit, and the third integral simplifies, giving

$$\lim_{r_* \rightarrow 0} [S_R(A : B) - I(A : B)]_{\text{area}} = \frac{1}{G_N} \int_{r_*^c}^{\infty} dr r \sqrt{f(r)} \left(\frac{1}{\sqrt{r^2 - (r_*^c)^2}} - \frac{1}{r} \right). \quad (4.12)$$

To study the $r_*^c \rightarrow \infty$ limit of this integral, it will be convenient to make the substitution $r \mapsto r_*^c \sec \alpha$, which removes r_*^c from the limits of integration. Under this substitution, we have

$$\lim_{r_* \rightarrow 0} [S_R(A : B) - I(A : B)]_{\text{area}} = \frac{r_*^c}{G_N} \int_0^{\pi/2} d\alpha \sqrt{f(r_*^c \sec \alpha)} \sec \alpha (\sec \alpha - \tan \alpha) \quad (4.13)$$

In the $r_*^c \mapsto \infty$ limit, we may approximate $f(r_*^c \sec \alpha)$ by its large-argument expansion $f(r_*^c \sec \alpha) \sim 1/(r_*^c \sec \alpha)^2$. Making this approximation gives

$$\lim_{r_*^c \rightarrow \infty} \lim_{r_* \rightarrow 0} [S_R(A : B) - I(A : B)]_{\text{area}} = \frac{1}{G_N} \int_0^{\pi/2} d\alpha (\sec \alpha - \tan \alpha) = \frac{\log(2)}{G_N}. \quad (4.14)$$

The result of this calculation seems quite suggestive. For any asymptotically AdS₃ spacetime with a spherically symmetric moment of time symmetry, the classical contribution to the Markov gap of two large intervals approaches the lower bound of inequality (3.1). This seems to hint at some underlying universality in the Markov gap for two intervals, at least in holographic conformal field theories.

As indicated in the introduction to this section, this calculation also implies that the bound (3.1) is respected at the classical level for arbitrary spherically symmetric perturbations to the state of two vacuum intervals, to all orders in perturbation theory. This follows from the fact that (3.1) is saturated only in the limit $r_* \rightarrow 0, r_*^c \rightarrow \infty$; but in that limit, we have just shown that the Markov gap for any spherically symmetric metric approaches that of the vacuum.

4.2 Quantum perturbations

In the presence of quantum matter, the Dutta-Faulkner formula is given by equation (1.5). In small- G_N perturbation theory, the minimum over quantum extremal surfaces appearing in the holographic entanglement entropy formula can be replaced by the generalized entropy of the classically minimal surface — in fact, this calculation by Faulkner, Lewkowycz, and Maldacena in [10] served as the prelude to the general quantum extremal surface formula [15]. While it is important to remember that this perturbative approach can miss important nonperturbative contributions coming from entanglement islands [53, 54] or large breakdowns of entanglement wedge reconstruction [65], it is still useful in regimes far from any phase transition where multiple quantum extremal surfaces vie for dominance. Analogously, in small- G_N perturbation theory, the Dutta-Faulkner formula can be replaced by

$$S_R(A : B) \approx \frac{2 \text{area}(\sigma_{A:B})}{4G_N} + S_{R,\text{bulk}}(\sigma_{A:B}), \quad (4.15)$$

where $\sigma_{A:B}$ is the *classical* entanglement wedge cross-section.

In this regime, the Markov gap can be written

$$S_R(A : B) - I(A : B) \approx [S_R(A : B) - I(A : B)]_{\text{area}} + S_{R,\text{bulk}}(\sigma_{A:B}) - I_{\text{bulk}}(A : B). \quad (4.16)$$

We need to be a little careful about what each term in this equation means. The area term is just the sum of area differences for the appropriate classical KRT and RT surfaces. The reflected entropy term is also fairly straightforward: it is the reflected entropy of the quantum fields in the bulk entanglement wedge $W(AB)$ subject to the bipartition induced by the cross-section $\sigma_{A:B}$. The bulk term $I_{\text{bulk}}(A : B)$, however, is *not* the mutual information of $W(AB)$ subject to that bipartition. It is the sum of the entropies of the regions bounded by $\text{RT}(A)$ and $\text{RT}(B)$, minus the entropy of the region bounded by $\text{RT}(AB)$; it is not actually the mutual information of a state of the bulk quantum fields.

See figure 18 for a representation of the various bulk regions that need to be considered in this calculation. $S_{R,\text{bulk}}(\sigma_{A:B})$ is the reflected entropy of the bulk state on $aa' \cup bb'$. $I_{\text{bulk}}(A : B)$ is the entropy of the bulk state on a plus that of the state on b , minus the entropy of the bulk state on $aa'bb'$. Using nonnegativity of the Markov gap for general quantum states gives us

$$S_{R,\text{bulk}}(\sigma_{A:B}) = S_R(aa' : bb') \geq I(aa' : bb') = S(aa') + S(bb') - S(aa'bb'). \quad (4.17)$$

Applying this inequality to the bulk term appearing in equation (4.16) gives the inequality

$$\begin{aligned} S_{R,\text{bulk}}(\sigma_{A:B}) - I_{\text{bulk}}(A : B) &= S_{R,\text{bulk}}(\sigma_{A:B}) - S(a) - S(b) + S(aa'bb') \\ &\geq S(aa') + S(bb') - S(a) - S(b). \end{aligned} \quad (4.18)$$

For general quantum states, the right-hand side of this inequality is not necessarily non-negative. This may seem puzzling — after all, as we noted in the introduction in equation (1.11), combining the area and bulk terms in equation (4.16) with inequality (4.18) gives a difference in generalized entropies that is guaranteed to be nonnegative when one

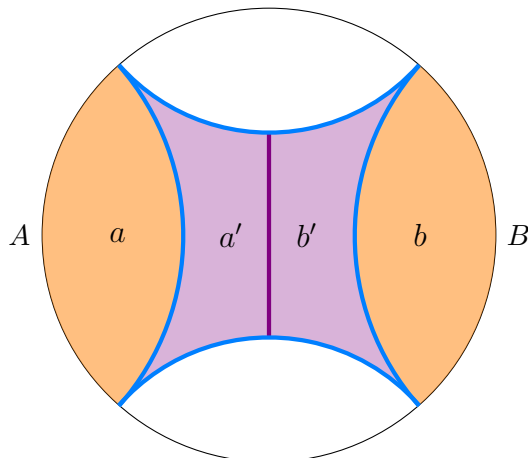


Figure 18. The bulk regions one needs to consider to compute the bulk term in equation (4.16). The region a lies between A and $\text{RT}(A)$, a' lies between $\text{RT}(A)$ and $\text{KRT}(A)$, and analogously for b and b' .

uses the full, nonperturbative quantum extremal surface formula. At the level of perturbation theory, though, the potential negativity of the bulk contribution to (4.16) seems like it could be a problem. However, this problem only arises when the area term in (4.16) vanishes — the perturbative bulk term cannot compete with a nonzero area term, since the area term contributes to the Markov gap at order $1/G_N$ while the perturbative bulk term contributes at order 1. The potential issue in equation (4.16) arising in perturbation theory when the area term vanishes is a manifestation of the principle explained in the first paragraph of this subsection, that perturbative holographic entropy formulas cannot be trusted near entanglement phase transitions.

For the case discussed in subsection 4.1, however, where the boundary regions are equal-size antipodal intervals in a spherically symmetric time-slice, it seems reasonable to assume that the right-hand side of inequality (4.18) will vanish in the limit as the intervals become large. In this limit regions aa' and a both approach a “half-spacetime” region bounded in the bulk by a line of constant θ . While they approach this limit in different ways — for example, the boundary of aa' always has corners while the boundary of a does not — the bulk entropies appearing in (4.18) are supposed to be renormalized entropies that capture universal contributions to the entanglement entropy without counting boundary divergences.³² Barring subtleties in this universality, we suspect that the quantities $S(aa') - S(a)$ and $S(bb') - S(b)$ for antipodal, equal-size intervals in a spherically symmetric state will vanish

³²There are some subtleties to keep in mind here, because quantum field theory states on spatial regions with non-smooth boundaries — such as aa' — tend to have divergences associated with those corners. Furthermore, higher-derivative corrections to the classical piece of the generalized entropy involve extrinsic curvature terms that are ill-defined at corners. Under a suitable regularization procedure, where the corners are smoothed out enough to regulate the corner divergences and to define the higher-derivative classical entropy, but not enough to appreciably change the area contribution to the entropy, it seems reasonable to hope that the generalized entropy of regions with boundary corners is renormalizable and UV-finite. See section 4.2 of [28] for some discussion of this point.

in the limit as those intervals become large. This establishes the claim made in the introduction to this section, that the bound $S_R - I \geq \log(2)/G_N$ holds in perturbation theory for such “symmetric two-large-interval states” even in the presence of quantum matter.

Proving an inequality like (3.1) for general holographic states with quantum matter seems quite delicate. For states without quantum matter, one could imagine that a clever application of the weak or null energy condition could be used to guarantee inequality (3.1) in arbitrary states. States with quantum matter are known to violate all the energy conditions satisfied by classical matter, however, and it seems possible that states with quantum matter could cause the area-term contribution to the Markov gap to violate inequality (3.1). For the bound to hold in such states, the bulk contribution $S_{R,\text{bulk}} - I_{\text{bulk}}$ would need to be large enough to make up for the deficit in the area term. If this is true, then it implies a rather delicate energy condition in bulk quantum field theories: whenever the energy configuration of the quantum matter is such that $[S_R - I]_{\text{area}}$ dips below the lower bound imposed by equation (3.1), the bulk contribution $S_{R,\text{bulk}} - I_{\text{bulk}}$ will need to be larger than the magnitude of that dip.

Should we expect this to be true? Maybe! It certainly seems like a rather intricate condition to impose on bulk quantum fields, especially since, as explained in the discussion surrounding equation (4.18), the bulk contribution $S_{R,\text{bulk}} - I_{\text{bulk}}$ is not the Markov gap of any bulk state and does not a priori need to be nonnegative. Determining whether $S_{R,\text{bulk}} - I_{\text{bulk}}$ must be nonnegative due to some other general principle would be an interesting avenue of investigation. Despite the intricacy of these conditions, the classical universality of (3.1) suggested by the analysis of subsection 4.1 indicates to us that this inequality really is capturing some interesting universal feature of holographic entanglement. It is our *hope*, then, that the inequality is truly universal or has some universal generalization. However, it is important to note that another classical holographic entropy inequality, the monogamy of mutual information [8, 29], does not generally hold in states with quantum matter [28].

We comment further on these points in the discussion (section 6).

4.3 Higher dimensions

In inequality (3.1), the Markov gap is lower bounded by a universal constant times the number of boundaries in the entanglement wedge cross-section. Because the entanglement wedge cross-section is one-dimensional, its boundary is zero-dimensional, and the only geometric quantity we can associate to it is the number of components. In higher dimensions, the boundary of the entanglement wedge cross-section can be more complicated; it is a codimension-3 surface, and one natural generalization of the number of connected components is its codimension-3 volume. So we might expect that in $D + 1$ bulk dimensions, the generalization of (3.1) will be something like

$$S_R(A : B) - I(A : B) \geq \frac{C_D}{4G_N} \times \text{area}(\text{cross-section boundary}) \quad (4.19)$$

with C a universal, dimension-dependent constant. If we define the area of a zero-dimensional surface to be its number of components, this bound reproduces (3.1) in $D = 2$ with $C_2 = 2 \log(2)$.

We have attempted to check (4.19) in the case $D > 2$ by computing $S_R - I$ numerically for some simple families of states and minimizing the quantity $4G_N(S_R - I)/\text{area}$. We hoped that we would find some universal minimum appearing in multiple, a priori unrelated families of states — this is what happens in the case $D = 2$, where both the “two-interval” states and the “two-asymptotic boundaries of a wormhole” states achieve, as a limit within their families, the minimum value $C_2 = 2\log(2)$. Unfortunately, this is not the case for the families we have considered in the case $D > 2$. This is not evidence *against* the conjectured inequality (4.19), but it does indicate that the situation above three bulk dimensions is subtle.

The two families we considered in four bulk dimensions were: (i) A and B are two equal-time strips on the boundary of AdS_{D+1} in Poincaré coordinates; (ii) A and B are two equal-time antipodal caps on the boundary of AdS_{D+1} in global coordinates. While these cases are equivalent in three bulk dimensions, they are not in higher dimensions — for example, the mutual information between two strips is infinite because they “touch” at the point on the boundary sphere that isn’t covered by the Poincaré coordinates (see [66] for details).

We will not reproduce the calculations here explicitly, because we do not find them particularly instructive. The goal is to compute the quantity

$$\chi = 4G_N \frac{S_R(A : B) - I(A : B)}{\text{area}(\text{cross-section boundary})}. \quad (4.20)$$

The mutual information between two strips in arbitrary dimension was computed analytically in [66]; the mutual information between two caps was computed numerically in [67]. The methods of those two papers can be adapted quite straightforwardly to compute $S_R(A : B)$ and the codimension-3 area of the cross-section boundary. The final results, in arbitrary dimension $D > 2$, are:

1. When A and B are vacuum strips in boundary Poincaré coordinates, χ monotonically decreases with the size of the strips and monotonically increases with the separation between the strips. Its minimum value, achieved when the separation goes to zero and the strip size goes to infinity, is

$$4G_N \frac{S_R(A : B) - I(A : B)}{\text{area}(\text{cross-section boundary})} \geq \frac{2}{D-2} \left[1 - \sqrt{\pi} \frac{\Gamma\left(\frac{D}{2(D-1)}\right)}{\Gamma\left(\frac{1}{2(D-1)}\right)} \right]. \quad (4.21)$$

One can check that the right-hand side goes to $2\log(2)$ for $D \rightarrow 2$.

2. When A and B are vacuum caps in boundary global coordinates, χ approaches the right-hand side of (4.21) in the limit as the caps become large and their separation becomes small. This isn’t so surprising, because the strips become “cap-like” in global coordinates as they increase in size and decrease in separation. However, unlike in the case of strips, in the case of caps χ can be shown numerically to *decrease* with the separation between the caps — i.e., the minimal value of χ for caps is achieved when the caps are as small as possible while still having a connected entanglement wedge,

a limit in which there is generally no analytical control of the calculation. (See section 2.2 of [67].) There are, therefore, cap configurations that violate inequality (4.21), and the minimum value of χ for antipodal caps has no nice analytic formula.

From these considerations, it is clear that establishing an inequality like (4.19) in arbitrary dimensions is a nuanced undertaking. We mention some techniques that might be helpful for that problem in the discussion (section 6.1).

5 Holographic Markov recoveries and fixed area states

In this section, we revisit the Markov recovery processes introduced in section 2. The essential observation of that section was that the Markov gap $S_R - I$ for a state ρ_{AB} is lower-bounded by a function of the fidelity of Markov recovery processes on the canonical purification. We reproduce the relevant inequalities here:

$$S_R(A : B) - I(A : B) \geq - \max_{\mathcal{R}_{B \rightarrow BB^*}} \log F(\rho_{ABB^*}, \mathcal{R}_{B \rightarrow BB^*}(\rho_{AB})). \quad (5.1)$$

$$S_R(A : B) - I(A : B) \geq - \max_{\mathcal{R}_{A \rightarrow AA^*}} \log F(\rho_{AA^*B}, \mathcal{R}_{A \rightarrow AA^*}(\rho_{AB})). \quad (5.2)$$

Focusing on the first inequality, we argued in section 2.2 that even the best Markov recovery channel $\mathcal{R}_{B \rightarrow BB^*}$ will be unable to reproduce the boundary entanglement supporting the portion of the quantum extremal surface γ_{AB} that lies outside the BB^* entanglement wedge.³³ We will call this surface $J(A)$, referring to section 2.2, where we called it a “jagged surface” based on the way it was depicted in figures 5 and 6. The failure of the optimal Markov recovery map to produce the entanglement required to support a geometric connection across $J(A)$ implies an imperfect fidelity of recovery, and therefore a nonzero contribution to the right-hand side of inequality (5.1).

The intuition given in the preceding paragraph is entirely qualitative. It does not give us any quantitative estimate for how the failure of $\mathcal{R}_{B \rightarrow BB^*}$ to create the entanglement supporting $J(A)$ constrains the fidelity appearing in inequality (5.1). The primary achievement of this section will be to provide such a quantitative estimate. We will argue, in a fixed-area-state toy model of the Markov recovery process, that any Markov recovery map $\mathcal{R}_{B \rightarrow BB^*}$ must have fidelity satisfying the upper bound

$$F(\rho_{ABB^*}, \mathcal{R}_{B \rightarrow BB^*}(\rho_{AB})) \leq e^{-\Delta} \left[{}_2F_1 \left(\frac{1}{2}, -\frac{1}{2}; 2; e^{-\Delta} \right) \right]^2, \quad (5.3)$$

with

$$\Delta = \frac{\text{area}(\text{KRT}(A)) - \text{area}(\text{RT}(A))}{4G_N} \quad (5.4)$$

being the entropy difference between the KRT and RT surfaces defined in the introduction. Plugging inequality (5.3) in to inequality (5.1) gives the lower bound

$$S_R(A : B) - I(A : B) \geq \Delta - 2 \log \left[{}_2F_1 \left(\frac{1}{2}, -\frac{1}{2}; 2; e^{-\Delta} \right) \right]. \quad (5.5)$$

³³In the language of section 3, this is the non-cross-section portion of the “kinked Ryu-Takayanagi” surface $\text{KRT}(A)$.

Estimating the right-hand side of (5.1) for states with quantum matter — whose boundary entropies include bulk entropy contributions and thus cannot be modeled as straightforwardly using fixed area states³⁴ — is a more intricate task that we do not undertake in this paper; however, we roughly expect that in suitable toy models, the right-hand side of equation (5.4) would be replaced by a difference of generalized entropies.

We pause to note that one can check, using $\Delta \geq 0$, that the right-hand side of (5.5) is greater than or equal to Δ . If we take this toy model seriously, then we learn that $S_R - I$ is lower-bounded by the entropy difference $S_{\text{gen}}(\text{KRT}(A)) - S_{\text{gen}}(\text{RT}(A))$. At this point, we could declare success; applying the same analysis for the Markov chain $B \rightarrow A \rightarrow A^*$ would give us the two inequalities

$$S_R(A : B) - I(A : B) \geq S_{\text{gen}}(\text{KRT}(A)) - S_{\text{gen}}(\text{RT}(A)). \tag{5.6}$$

$$S_R(A : B) - I(A : B) \geq S_{\text{gen}}(\text{KRT}(B)) - S_{\text{gen}}(\text{RT}(B)). \tag{5.7}$$

If we look back to inequality (1.11) from the introduction, we could interpret this analysis as teaching us that Markov recovery arguments can reproduce each term of inequality (1.11) individually. This might, however, seem like a bit of a letdown; after all, since (1.11) tells us that $S_R - I$ is lower-bounded by the *sum* of these entropy differences, it is stronger than the two inequalities (5.6) and (5.7). However, we shouldn't expect either of the inequalities (5.6) or (5.7) to be anywhere close to saturation. For one thing, the information inequalities (5.1) & (5.2) are not guaranteed to be tight; it is entirely possible, in fact even generic, for the optimal recovery map $\mathcal{R}_{B \rightarrow BB^*}$ to have fidelity greater than $e^{-(S_R - I)}$. There is also, as we explain in subsection 5.1, generically a significant gap in inequality (5.3). Finally, the second term on the right-hand side of (5.5) is never exactly zero. These three sources of error will lead to significant gaps between the left- and right-hand sides of (5.6) and (5.7), consistent with inequality (1.11).

In the following subsections, we give the details of the argument sketched above. In subsection 5.1 we describe a particular state $\tilde{\rho}_{ABB^*}$, which we call the “Too Good to be True” or TGT state, and give physical arguments for the inequality

$$F(\rho_{ABB^*}, \mathcal{R}_{B \rightarrow BB^*}(\rho_{AB})) \leq F(\rho_{ABB^*}, \tilde{\rho}_{ABB^*}). \tag{5.8}$$

In subsection 5.2, we compute the fidelity $F(\rho_{ABB^*}, \tilde{\rho}_{ABB^*})$ for analogous states in fixed-area models, and reproduce inequality (5.3).

5.1 “Too Good to be True” recovery

The three party reduced state ρ_{ABB^*} has, as its entanglement wedge, the spacelike portion of the canonical purification spacetime bounded in the bulk by $\text{RT}(A^*)$ and on the boundary by ABB^* . We sketched a spacelike slice of this entanglement wedge in figure 5 for the case where ρ_{AB} is the density matrix of two equal-time intervals in the AdS_3 vacuum; we have reproduced a version of that figure with labeling more appropriate for our present purposes as figure 19.

³⁴However, see [20] for some entropy calculations in fixed area states with bulk matter.

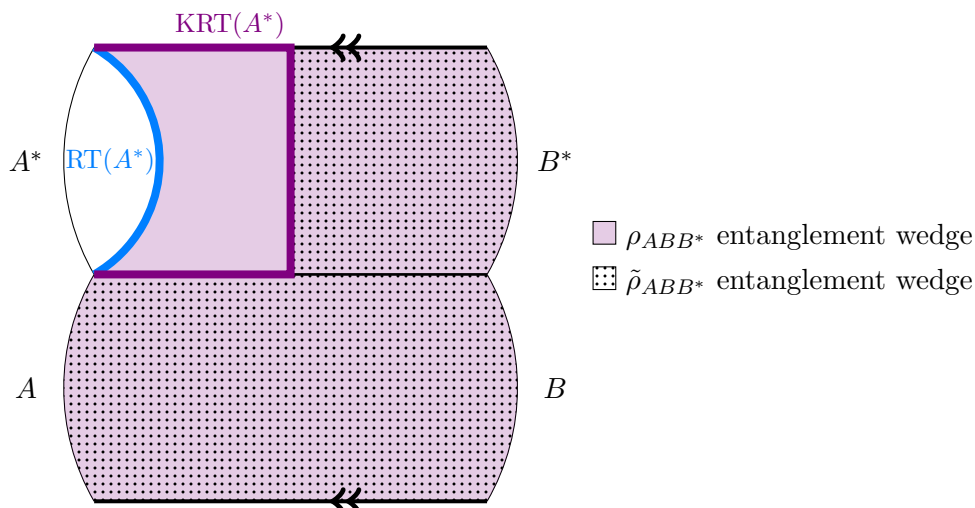


Figure 19. A sketch of the canonical purification for two equal-time intervals in the AdS_3 vacuum, including a sketch of two possible states of the system ABB^* . The state ρ_{ABB^*} , obtained as a partial trace of the canonical purification, has $\text{RT}(A^*)$ as its entanglement wedge boundary. The TGT state $\tilde{\rho}_{ABB^*}$ has $\text{KRT}(A^*)$ as its entanglement wedge boundary.

By the “Too Good to be True” state, which we will denote $\tilde{\rho}_{ABB^*}$ and sometimes call the TGT state, we will mean a state whose bulk entanglement wedge is the union of the AB and BB^* entanglement wedges. In figure 19, this is the bulk region filled with dots. In the language of section 3, this is a state whose entanglement wedge is bounded in the bulk by the KRT surface $\text{KRT}(A^*)$.

We pause, for a moment, to reflect on the definition of this state. Within the semiclassical theory defined on the spacetime dual to the canonical purification, the state $\tilde{\rho}_{ABB^*}$ is perfectly well defined — one simply traces out the bulk QFT degrees of freedom lying between $\text{RT}(A^*)$ and $\text{KRT}(A^*)$. Boundary states, however, are not exactly dual to a single bulk spacetime; the path integrals preparing them admit multiple saddles that contribute to their physics, and tracing out a portion of the dominant saddle is not an operation that has an obvious boundary interpretation. When we talk about the TGT state, we will really mean *any* state for which low-energy correlators and entanglement entropies are accurately reproduced, using the usual AdS/CFT dictionary, by the portion of the canonical purification spacetime lying between $\text{KRT}(A^*)$ and ABB^* . In particular, the entropy $S(\tilde{\rho}_{ABB^*})$ will be the generalized entropy of the KRT surface $\text{KRT}(A^*)$. There may be a whole family of states with this property, differing in the details of their subleading saddles, but their physical properties should only differ nonperturbatively in $1/G_N$. Or perhaps it is not actually possible to construct a TGT state in full AdS/CFT, since the kinks in $\text{KRT}(A^*)$ prevent it from being truly quantum extremal. It *will* be possible to give a precise construction of the TGT state in the fixed-area toy model of the following subsection, which is good enough for our purposes.

Ultimately, the details of the TGT state $\tilde{\rho}_{ABB^*}$ will not be so important. All we will really want is a prescription for computing the fidelity $F(\rho_{ABB^*}, \tilde{\rho}_{ABB^*})$. The reason for this is that we will assume, on physical grounds, that any Markov recovery map $\mathcal{R}_{B \rightarrow BB^*}$ acting on ρ_{AB} satisfies the inequality

$$F(\rho_{ABB^*}, \mathcal{R}_{B \rightarrow BB^*}(\rho_{AB})) \leq F(\rho_{ABB^*}, \tilde{\rho}_{ABB^*}). \quad (5.9)$$

This follows, essentially, from the physical arguments made in section 2.2. Any $B \rightarrow BB^*$ channel that touches only the B subsystem of ρ_{AB} can *at best* reproduce the AB entanglement wedge, the BB^* entanglement wedge, and whatever shared entanglement is needed to attach those wedges smoothly. In fact, even this is usually too much to ask, which is why we called $\tilde{\rho}_{ABB^*}$ the “too good to be true” state.

Indeed, suppose it were possible to obtain the TGT state via a Markov recovery map, i.e., assume we have

$$\tilde{\rho}_{ABB^*} = \mathcal{R}_{B \rightarrow BB^*}(\rho_{AB}) \quad (5.10)$$

for some quantum channel $\mathcal{R}_{B \rightarrow BB^*}$. It is straightforward to verify the equality

$$S_R(A : B) - I(A : B) = -S(\rho_{ABB^*}) + S(\rho_{BB^*}) - S(\rho_{AB} || I_A \otimes \rho_B), \quad (5.11)$$

where $S(\cdot || \cdot)$ is the relative entropy. Monotonicity of relative entropy under application a quantum channel gives

$$S(\rho_{AB} || I_A \otimes \rho_B) \leq S(\mathcal{R}_{B \rightarrow BB^*}(\rho_{AB}) || I_A \otimes \mathcal{R}_{B \rightarrow BB^*}(\rho_B)) = S(\tilde{\rho}_{ABB^*} || I_A \otimes \rho_{BB^*}), \quad (5.12)$$

where we have used the fact that the BB^* reduced state of $\tilde{\rho}_{ABB^*}$ agrees with that of ρ_{ABB^*} . Plugging this inequality back into equation (5.11) and writing the relative entropy in terms of entanglement entropies gives

$$S_R(A : B) - I(A : B) \leq S(\tilde{\rho}_{ABB^*}) - S(\rho_{ABB^*}) = S_{\text{gen}}(\text{KRT}(A^*)) - S_{\text{gen}}(\text{RT}(A^*)). \quad (5.13)$$

Exploiting the $A \leftrightarrow A^*, B \leftrightarrow B^*$ symmetry of the canonical purification, we may rewrite this as

$$S_R(A : B) - I(A : B) \leq S_{\text{gen}}(\text{KRT}(A)) - S_{\text{gen}}(\text{RT}(A)). \quad (5.14)$$

But, referring back to equation (1.11) from the introduction, we see that this is a contradiction unless $S_{\text{gen}}(\text{KRT}(B))$ equals $S_{\text{gen}}(\text{RT}(B))$.

We take this to suggest that, except in very special cases, inequality (5.9) really ought to be strict. The gap in that inequality should contribute to a gap in the inequality (5.5); which, as we discussed in the paragraphs following that inequality, will generically be fairly far from equality.

5.2 Fixed area Markov recovery

We now proceed to computing the fidelity between the canonical purification reduced state ρ_{ABB^*} and the TGT state $\tilde{\rho}_{ABB^*}$ in a toy model where the fidelity can be computed exactly. Our toy model will be made out of fixed area states, introduced in [41, 42], whose

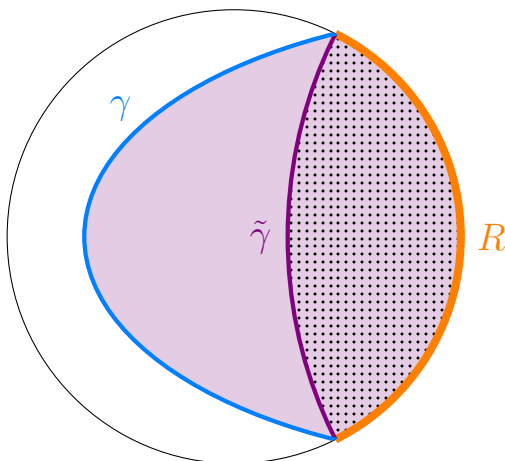


Figure 20. A sketch of a boundary state ρ_R whose bulk entanglement wedge contains two covariantly defined surfaces: the minimal quantum extremal surface γ and another surface $\tilde{\gamma}$ with greater generalized entropy. The goal of this section is to compute, in a toy model, the quantum fidelity between ρ_R and a “coarser” state $\tilde{\rho}_R$ whose entanglement wedge is only the dotted region.

gravitational path integral rules were shown in [42] to be sufficiently simple that many entropy calculations can be done exactly with fairly mild assumptions.

As we emphasized in the introduction, the surfaces $\text{RT}(A^*)$ and $\text{KRT}(A^*)$ are best thought of as two preferred members of the class of codimension-2 spacelike surfaces homologous to A^* . The surface $\text{RT}(A^*)$ is singled out because it has minimal generalized entropy among the surfaces that are quantum extremal; the surface $\text{KRT}(A^*)$ is singled out because it contains the entanglement wedge cross-section. So we might rephrase the question of computing $F(\rho_{ABB^*}, \tilde{\rho}_{ABB^*})$ as follows:

- Given a boundary region R and a state ρ_R whose bulk entanglement wedge is bounded by the bulk surface γ , and which contains another covariantly defined surface $\tilde{\gamma}$ with greater generalized entropy, define $\tilde{\rho}_R$ to be a state whose entanglement wedge ends at $\tilde{\gamma}$. What is the fidelity $F(\rho_R, \tilde{\rho}_R)$?

If we substitute $R \rightarrow ABB^*$, $\gamma \rightarrow \text{RT}(A^*)$, and $\tilde{\gamma} \rightarrow \text{KRT}(A^*)$, this is exactly the question we posed in the previous subsection. A general state of this kind is sketched in figure 20.

This is an interesting question in general — presumably the answer tells us something about the boundary degrees of freedom encoding the bulk region between γ and $\tilde{\gamma}$. The physical question we are asking is something like, “how much information do you lose when you trace out that region?” While we know of no way to compute this quantity for general bulk states, we can compute it exactly in a fixed area state toy model. To apply the fixed-area path integral formalism of [42], we will have to assume that both γ and $\tilde{\gamma}$ are classically extremal. This is not exactly the case for the TGT state defined in the previous subsection, even in states with no quantum matter — while $\tilde{\gamma} = \text{KRT}(A^*)$ is locally extremal, the isolated kinks prevent it from being a truly extremal surface. Verifying that the fixed-area calculation goes through without caveats for extremal surfaces with isolated kinks would

be interesting, but would require a careful analysis of boundary conditions at those kinks that is beyond the scope of the present calculation. We will therefore treat $\text{KRT}(A^*)$ as a truly extremal surface. This is not such an extreme assumption, as inequality (5.5) is already only supposed to hold in a toy model; the extra assumption that we can treat $\text{KRT}(A^*)$ as extremal should be thought of as an extra assumption on the model.

We now review the bare essentials of fixed area states and their path integrals, originally described in [41, 42]; we refer the reader to section 4.2 of [20] for a very nice, thorough introduction to these techniques. Starting with a static, geometric state in AdS/CFT, a fixed area state is constructed by picking a gauge-invariant surface Γ ,³⁵ and projecting the state onto a narrow band of eigenspaces³⁶ of the operator that gives the area of Γ within a code space of excitations. It was argued in [42] that fixed area states can be constructed by taking the gravitational path integral preparing the original state and restricting the integral to be taken over only those metrics for which the area of the designated surface — again, defined in some gauge-invariant way — lies within the designated band.³⁷ At the level of solutions to the classical equations of motion, fixing the area of an *extremal* surface Γ amounts to allowing arbitrary conical defects around that surface; if one approximates a fixed area path integral by summing over saddles, the relevant saddles are spacetimes that are solutions to Einstein’s equations everywhere but at the designated surface, where there may be a conical defect.

For our toy model, we will construct a fixed area analogue of ρ_R by first taking the state to be classical — i.e., it contains no quantum matter and therefore boundary entropies are computed by bulk areas, with $\text{area}(\gamma) < \text{area}(\tilde{\gamma})$ — and then fixing the areas of *both* surfaces γ and $\tilde{\gamma}$. We will henceforth abuse notation and call this fixed area state ρ_R , since from here on out we will work entirely in the toy model. We will define $\tilde{\rho}_R$ to be a state where *only* the area of $\tilde{\gamma}$ is fixed. At the level of the saddles contributing to the fixed area path integral we will see shortly that it is sensible to say that ρ_R knows about both surfaces γ and $\tilde{\gamma}$, while $\tilde{\rho}_R$ knows only about the surface $\tilde{\gamma}$; this is the justification for the toy model.

Now, suppose $|\psi\rangle$ is a state in AdS/CFT that can be prepared by a Euclidean path integral. We will assume that $|\psi\rangle$ is geometric; what this means at a technical level is that the path integral computing $\langle\psi|\psi\rangle$ is dominated by a single gravitational saddle point that “fills in” the bulk; this is sketched in figure 21. If $|\psi^{\text{fix}}\rangle$ is a fixed area state constructed from $|\psi\rangle$, and σ_R^{fix} is the corresponding density matrix, then saddles contributing to $\text{Tr}((\sigma_R^{\text{fix}})^n)$ can be constructed by replicating the saddle dominating $\langle\psi|\psi\rangle$ n times around the fixed surface. This is sketched in figure 22. The main assumption that goes into the fixed area

³⁵In fact, it isn’t enough to require Γ to be gauge invariant; we must also require that it can be sensibly defined in *any* sufficiently smooth spacetime with the right boundary conditions, in order to be able to identify an analogue of Γ in spacetimes contributing to the gravitational path integral. The surfaces $\text{RT}(A)$ and $\text{KRT}(A)$ satisfy this property, since they are defined based on certain minimality statements.

³⁶The reason we must project onto a “narrow band” of eigenspaces rather than an exact eigenspace is that the eigenvalues of an area operator are continuous.

³⁷This discussion is all a bit imprecise, as many of the areas one would like to fix are infinite, and must be somehow regulated. We will not address those subtleties here.

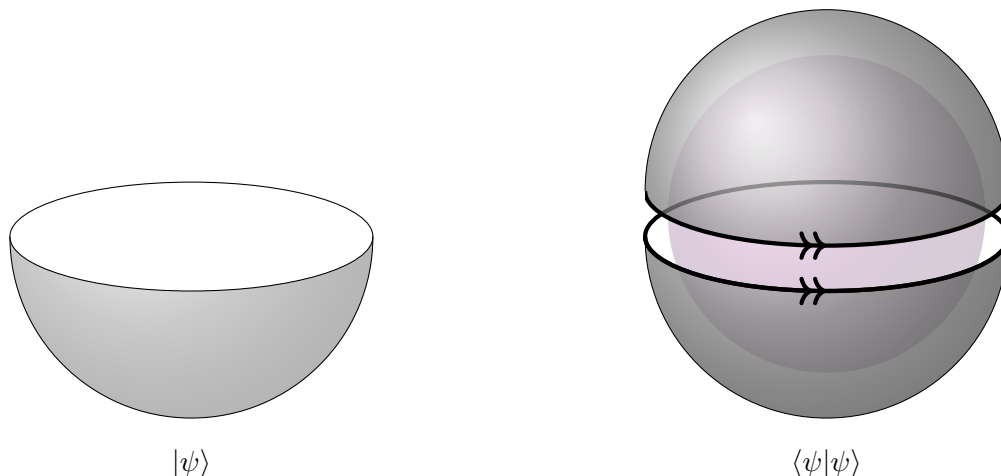


Figure 21. Left-hand side: a Euclidean path integral in the boundary theory preparing a state $|\psi\rangle$. Right-hand side: the Euclidean path integral computing $\langle\psi|\psi\rangle$, whose gravitational dual is assumed to be dominated by a single bulk saddle.

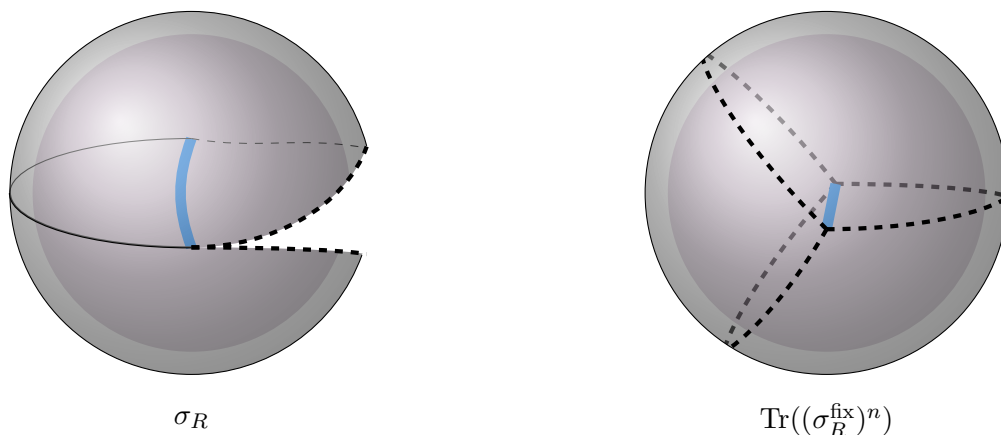


Figure 22. Left-hand side: the path integral preparing σ_R , whose trace is dominated by a single gravitational saddle containing a surface γ . Right-hand side: a saddle for $\text{Tr}((\sigma_R^{\text{fix}})^3)$, constructed by replicating the single-trace saddle three times around γ and thereby giving γ a conical defect angle 6π .

path integral is that saddles constructed in this way are the only ones that need to be considered to get an accurate answer for quantities like $\text{Tr}((\sigma_R^{\text{fix}})^n)$.

We will now describe how to compute the fidelity $F(\rho_R, \tilde{\rho}_R)$. This is accomplished by computing the quantity

$$G_{m,n}(\rho_R, \tilde{\rho}_R) = \text{Tr}((\tilde{\rho}_R^m \rho_R \tilde{\rho}_R^m)^n) \tag{5.15}$$

for integer values of m and n using the gravitational path integral, analytically continuing to non-integer values of m and n , and taking

$$F(\rho_R, \tilde{\rho}_R) = G_{\frac{1}{2}, \frac{1}{2}}(\rho_R, \tilde{\rho}_R)^2. \tag{5.16}$$

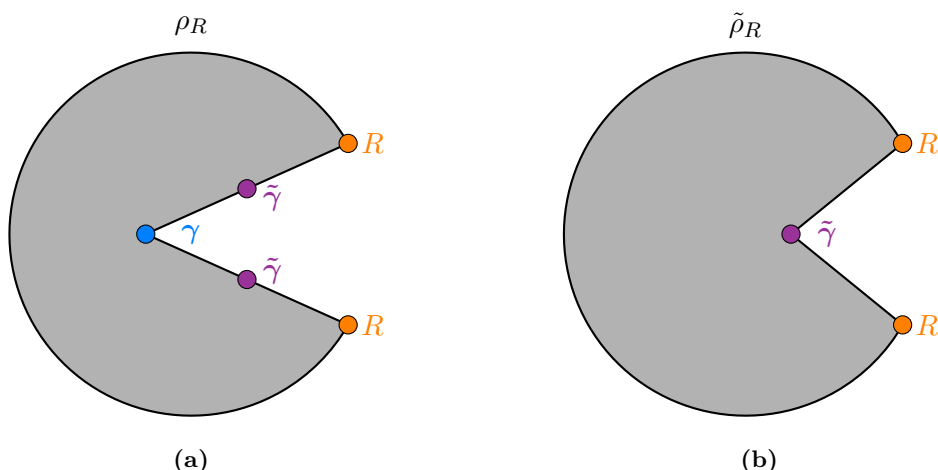


Figure 23. (a) A visual representation of the path integral preparing ρ_R . We can think of this as a cross-section of the path integral preparing the state shown in figure 20, with the two fixed-area surfaces shown in the dominant bulk saddle. (b) A representation of the path integral preparing $\tilde{\rho}_R$, which has only one fixed area surface.

To compute $G_{m,n}$, we will introduce a useful visual calculus developed in [59] and [20]. Every time a factor of ρ_R appears in an expression, we insert a bulk “pie” that looks like the one drawn in figure 23a. This is the bulk saddle that dominates the computation of $\text{Tr}(\rho_R)$, with a cut emanating from the fixed-area surface γ and ending at R . Every time a factor of $\tilde{\rho}_R$ appears, we insert a pie like the one shown in figure 23b. To aid in drawing figures, we will “unwrap” each pie and draw it as a “wedge” — sketched in figure 24a. The path integral for $\text{Tr}(\rho_R^n)$, in this language, is sketched in figure 24b; the segments of cuts neighboring the boundary R are all glued together cyclically, but there is one saddle for each possible gluing of the “intermediate” segments of cuts, which are in one-to-one correspondence with permutations in the group S_n . Readers who remain confused about this technique are encouraged to consult the more detailed explanation of [20].

In this language, the operator $\tilde{\rho}_R^m \rho_R \tilde{\rho}_R^m$ is prepared by a path integral sketched in figure 25. The function $G_{m,n}(\rho_R, \tilde{\rho}_R)$ is computed by taking n copies of this path integral and gluing the boundaries together cyclically. There is one saddle for each permutation $\sigma \in S_n$, with σ determining the gluing pattern for the interior segments of the sides of the wedges. In the saddle corresponding to permutation σ , there are $C(\sigma)$ copies of surface γ , where $C(\sigma)$ is the number of cycles in the cycle decomposition of σ . Before gluing, the total angle subtending γ -type surfaces is $2\pi n$, making the total defect

$$\text{defect}(\gamma) = 2\pi n - 2\pi C(\sigma). \tag{5.17}$$

The total number of copies of surface $\tilde{\gamma}$ after gluing is $C(\tau^{-1} \circ \sigma)$, where τ is the cyclic permutation. The total angle around $\tilde{\gamma}$ -type surfaces before gluing is $2\pi n(2m + 1)$. This makes the total defect around $\tilde{\gamma}$ surfaces

$$\text{defect}(\tilde{\gamma}) = 2\pi n(2m + 1) - 2\pi C(\tau^{-1} \circ \sigma). \tag{5.18}$$

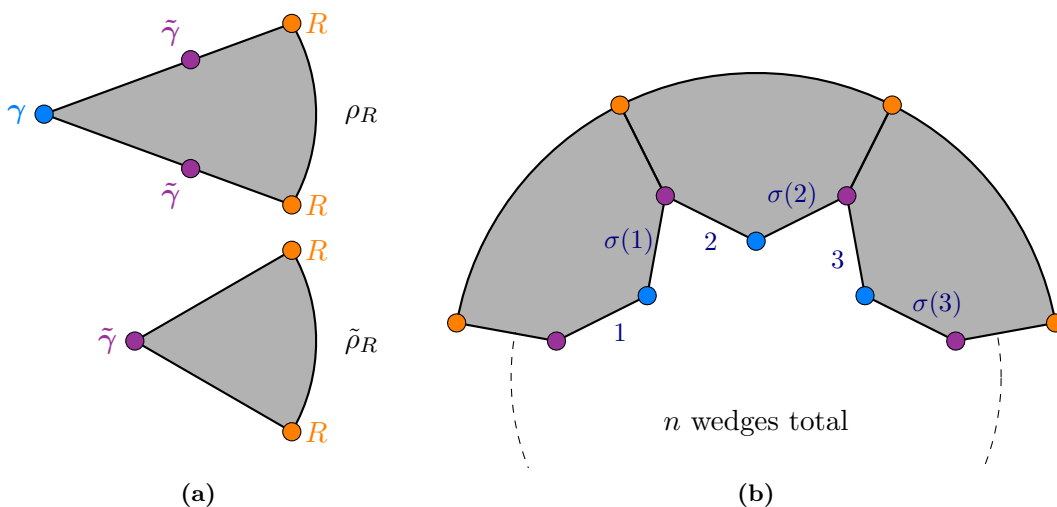


Figure 24. (a) The “pies” of figure 23 unwrapped into “wedges.” (b) A saddle contributing to $\text{Tr}(\rho_R^n)$. The surfaces lying counterclockwise of each blue dot are labeled 1 through n ; for every permutation $\sigma \in S_n$, one can construct a saddle by pasting the surface immediately clockwise of surface k to surface $\sigma(k)$.

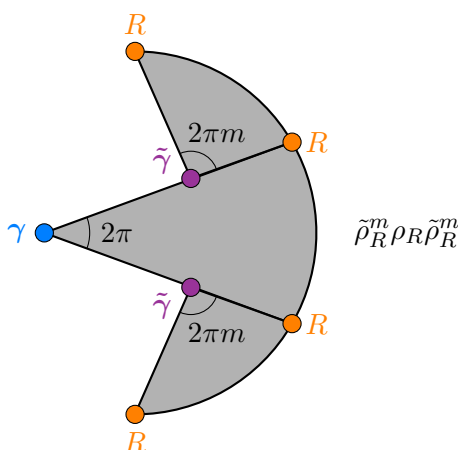


Figure 25. A visual representation of the gravity path integral for the operator $\tilde{\rho}_R^m \rho_R \tilde{\rho}_R^m$. The function $G_{m,n}$ defined in (5.15) is formed by pasting n copies of this path integral together cyclically on the boundary region R .

It was shown in [42] that the total path integral contribution from a saddle with fixed areas A_j and defects ϕ_j is

$$Z = \exp \left[- \sum_j \frac{A_j \phi_j}{8\pi G_N} \right]. \quad (5.19)$$

From this, we get the expression

$$G_{m,n}(\rho_R, \tilde{\rho}_R) = \sum_{\sigma \in S_n} \exp \left[\frac{\text{area}(\gamma)}{4G_N} (C(\sigma) - n) + \frac{\text{area}(\tilde{\gamma})}{4G_N} (C(\tau^{-1} \circ \sigma) - 2mn - n) \right]. \quad (5.20)$$

The largest terms in this sum are the ones with $C(\sigma) + C(\tau^{-1} \circ \sigma) = n + 1$. Terms with $C(\sigma) + C(\tau^{-1} \circ \sigma) < n + 1$ are subleading by factors like $e^{-\text{area}(\gamma)/4G_N}$ or $e^{-\text{area}(\tilde{\gamma})/4G_N}$. As emphasized in [20], these terms are *infinitely* suppressed when the fixed area surfaces have infinite area (i.e., when region R has a boundary); even when the areas are finite, these terms are suppressed nonperturbatively in $1/G_N$. We will neglect such terms — as we have already ignored some nonperturbative corrections to $G_{m,n}$ in assuming that the fixed area path integral is dominated by a sum over replicated saddles — but understanding how such terms contribute to the sum and whether any physics can be extracted from them may be an interesting direction for future work. The permutations σ with $C(\sigma) + C(\tau^{-1} \circ \sigma) = n + 1$ are called *non-crossing permutations* — see [68] for a review — and the number of such permutations with $C(\sigma) = k$ is the Narayana number

$$N(n, k) = \frac{1}{n} \binom{n}{k} \binom{n}{k-1}. \tag{5.21}$$

By summing over only non-crossing permutations, our formula for $G_{m,n}$ simplifies to

$$G_{m,n}(\rho_R, \tilde{\rho}_R) = \sum_{k=1}^n N(n, k) \exp \left[\frac{\text{area}(\gamma)}{4G_N} (k - n) + \frac{\text{area}(\tilde{\gamma})}{4G_N} (1 - k - 2mn) \right]. \tag{5.22}$$

This sum can be computed using Mathematica, giving the expression³⁸

$$G_{m,n}(\rho_R, \tilde{\rho}_R) = e^{-[2mn \text{area}(\tilde{\gamma}) - (1-n) \text{area}(\gamma)]/4G_N} {}_2F_1 \left(1 - n, -n; 2; e^{-[\text{area}(\tilde{\gamma}) - \text{area}(\gamma)]/4G_N} \right). \tag{5.23}$$

The hypergeometric function is analytic in its first two arguments when the third argument is fixed and the fourth argument has magnitude less than one;³⁹ these conditions are satisfied above because the area of $\tilde{\gamma}$ was assumed to be greater than that of γ , so we can freely analytically continue to $m = n = 1/2$ and obtain

$$G_{\frac{1}{2}, \frac{1}{2}}(\rho_R, \tilde{\rho}_R) = e^{-[\text{area}(\tilde{\gamma}) - \text{area}(\gamma)]/8G_N} {}_2F_1 \left(\frac{1}{2}, -\frac{1}{2}; 2; e^{-[\text{area}(\tilde{\gamma}) - \text{area}(\gamma)]/4G_N} \right). \tag{5.24}$$

Plugging this back into equation (5.16), we find the fidelity is

$$F(\rho_R, \tilde{\rho}_R) = e^{-[\text{area}(\tilde{\gamma}) - \text{area}(\gamma)]/4G_N} \left[{}_2F_1 \left(\frac{1}{2}, -\frac{1}{2}; 2; e^{-[\text{area}(\tilde{\gamma}) - \text{area}(\gamma)]/4G_N} \right) \right]^2. \tag{5.25}$$

We stress again that this equation holds only up to corrections nonperturbative in $1/G_N$. With the substitutions $\gamma \rightarrow \text{RT}(A^*)$, $\tilde{\gamma} \rightarrow \text{KRT}(A^*)$, $R \rightarrow ABB^*$, together with the $A \leftrightarrow A^*$, $B \leftrightarrow B^*$ symmetry of the canonical purification, equation (5.25) and inequality (5.9) together reproduce the claimed inequality (5.3).

To conclude, we note that the methods of this subsection can easily be generalized to compute all the sandwiched Rényi relative entropies of ρ_R and $\tilde{\rho}_R$. The sandwiched Rényi relative entropies were defined in [43, 44] by

$$\tilde{D}_\alpha(\rho || \sigma) = \frac{1}{\alpha - 1} \log \text{Tr}[(\sigma^\gamma \rho \sigma^\gamma)^\alpha] \tag{5.26}$$

³⁸We refer the reader to the very nice appendix III of [69] for an explanation of how to do sums like this by hand.

³⁹We refer the reader to section 2.1.6 of [70] for a review of analyticity properties of ${}_2F_1$.

with $\gamma = (1 - \alpha)/2\alpha$. In terms of our function $G_{m,n}$ defined in (5.15), we have

$$\tilde{D}_\alpha(\rho_R||\tilde{\rho}_R) = \frac{1}{\alpha - 1} \log \left[G_{(1-\alpha)/2\alpha,\alpha}(\rho_R, \tilde{\rho}_R) \right]. \quad (5.27)$$

Plugging in the formula for $G_{m,n}$ from equation (5.15) gives

$$\tilde{D}_\alpha(\rho_R||\tilde{\rho}_R) = \frac{\text{area}(\tilde{\gamma}) - \text{area}(\gamma)}{4G_N} + \frac{1}{\alpha - 1} \log \left[{}_2F_1 \left(1 - \alpha, -\alpha; 2; e^{-[\text{area}(\tilde{\gamma}) - \text{area}(\gamma)]/4G_N} \right) \right]. \quad (5.28)$$

The sandwiched Rényi relative entropies for $\alpha \neq 1/2$ are not symmetric in their arguments, but we can compute $\tilde{D}_\alpha(\tilde{\rho}_R||\rho_R)$ quite easily. A similar calculation to the one given for $G_{m,n}$ can be performed for the function

$$H_{m,n}(\rho_R, \tilde{\rho}_R) = \text{Tr}((\rho_R^m \tilde{\rho}_R \rho_R^m)^n), \quad (5.29)$$

yielding

$$H_{m,n}(\rho_R, \tilde{\rho}_R) = \sum_{\sigma \in S_{2mn}} \exp \left[\frac{\text{area}(\gamma)}{4G_N} (C(\sigma) - 2mn) + \frac{\text{area}(\tilde{\gamma})}{4G_N} (C(\tau^{-1} \circ \sigma) - n(2m + 1)) \right]. \quad (5.30)$$

Summing over non-crossing permutations in S_{2mn} gives

$$H_{m,n}(\rho_R, \tilde{\rho}_R) = e^{[-(2mn-1)\text{area}(\gamma) - n\text{area}(\tilde{\gamma})]/4G_N} \times {}_2F_1 \left(-2mn, 1 - 2mn, 2, e^{-[\text{area}(\tilde{\gamma}) - \text{area}(\gamma)]/4G_N} \right). \quad (5.31)$$

Setting $n = \alpha, m = (1 - \alpha)/2\alpha$ yields

$$H_{(1-\alpha)/2\alpha,\alpha}(\rho_R, \tilde{\rho}_R) = e^{-\alpha[\text{area}(\tilde{\gamma}) - \text{area}(\gamma)]/4G_N} \times {}_2F_1 \left(-1 + \alpha, \alpha, 2, e^{-[\text{area}(\tilde{\gamma}) - \text{area}(\gamma)]/4G_N} \right). \quad (5.32)$$

From this, we can compute the sandwiched Rényi relative entropy as

$$\tilde{D}_\alpha(\tilde{\rho}_R||\rho_R) = \frac{1}{\alpha - 1} \log \left[H_{(1-\alpha)/2\alpha,\alpha}(\rho_R, \tilde{\rho}_R) \right]. \quad (5.33)$$

Note, finally, that in the limit $\alpha \rightarrow 1$, the sandwiched Rényi relative entropy approaches the ordinary relative entropy. One can check that $\tilde{D}_\alpha(\rho_R||\tilde{\rho}_R)$ has a finite $\alpha \rightarrow 1$ limit, while $\tilde{D}_\alpha(\tilde{\rho}_R||\rho_R)$ diverges; we take this to mean that the support of $\tilde{\rho}_R$ is strictly larger than the support of ρ_R . This makes sense: very heuristically speaking, we might think of $\tilde{\rho}_R$ as a state constructed from ρ_R by taking the portion of the R -Hilbert space encoding the bulk region lying between γ and $\tilde{\gamma}$ and erasing the information therein by replacing it with the maximally mixed state.

A slightly different relative entropy between fixed area states was computed in [69]; this was a relative entropy between two states where both areas were fixed but where the bulk states contained orthogonal excitations in the region bounded by γ and $\tilde{\gamma}$.

6 Discussion

Our goal in this paper was threefold: (i) we established a connection between the quantity $S_R(A : B) - I(A : B)$ and Markov recovery processes in canonical purifications (section 2), arguing in the process that boundaries in the entanglement wedge cross-section of a holographic state require $S_R - I$ to be nonzero at order $1/G_N$; (ii) we proved inequality (1.12) for time-symmetric states in AdS₃ gravity (section 3), establishing a quantitative bound on the Markov gap in such states, and (iii) we explored a fixed area toy model of the Markov recovery process (section 5) to see how information inequalities partially reproduce the geometric bound on the Markov gap given in equation (1.11). Along the way, we explored generalizations of inequality (1.12) to states with bulk matter and states in higher-dimensional theories of gravity (section 4).

We now comment on some general lessons from our analysis, and some possible directions for future work.

6.1 Generalizing the bound

A natural first step in generalizing the proof of section 3 would be to prove inequality (1.12) for non-time-symmetric classical states in pure AdS₃ gravity. The chief difficulty in pursuing this generalization is that the minimal surfaces under consideration are no longer constrained to lie in a single time slice; one must work in the full, Lorentzian geometry.

The proof of section 3 worked by modeling time-symmetric slices of asymptotically AdS₃ states as quotients of the hyperbolic plane \mathbb{H}_2 . One could imagine that a similar analysis could be undertaken without the assumption of time symmetry, by modeling an arbitrary geometric state in pure AdS₃ gravity as a quotient of global AdS₃. While this problem certainly seems tractable, we think it unlikely that the technique of passing to the universal cover will be particularly useful in the long run. After all, one would like to generalize inequality (1.12) not only to non-time-symmetric states, but also to states with bulk matter; these states are generically not universally covered by AdS₃.

A technique we find more promising, in the sense that it is more easily generalizable, is to reinterpret the proof of inequality (1.12) in terms of Stokes' theorem. We emphasized in the introduction that the surfaces $\text{KRT}(A)$ and $\text{RT}(A)$ are homologous to one another. By expressing the areas of $\text{KRT}(A)$ and $\text{RT}(A)$ as integrals over the corresponding surfaces, one can use Stokes' theorem to write the area difference $[\text{area}(\text{KRT}(A)) - \text{area}(\text{RT}(A))]$ as an integral over the homology region between the two. There are many ways to write this integral — any smooth form on the homology region that interpolates between the volume forms on $\text{KRT}(A)$ and $\text{RT}(A)$ will give rise to a homology integral computing the area difference — but no matter how we write the integral, it must have the property that it diverges at the kinks of $\text{KRT}(A)$. We think it likely that a careful analysis of this problem will yield an alternative proof of (1.12), though we have not yet succeeded in tackling the problem from that direction. An advantage of the Stokes' theorem approach is that it is naturally generalizable: it can be applied to non-time-symmetric states, as well as to states in arbitrary dimensions, and even to states with classical bulk matter provided that one understands how suitable energy conditions on that matter affect the integral on the homology region.

Generalizing inequality (1.12) to states with bulk quantum matter is trickier. As we mentioned in section 4.2, the bulk contribution to the Markov gap $S_{R,\text{bulk}} - I_{\text{bulk}}$ is not necessarily nonnegative on its own; even though the Markov gap is nonnegative for any quantum state, the bulk entropy contribution to the mutual information is not actually the mutual information of a bulk state, so one cannot generally guarantee the inequality $S_{R,\text{bulk}} \geq I_{\text{bulk}}$ — at least not using the techniques of this paper. So proving inequality (1.12) in generality for the area contribution to the Markov gap would not suffice to prove the inequality in the presence of quantum matter. This state of affairs is further complicated by the fact that we don't really expect the area contribution of the Markov gap to universally satisfy inequality (1.12) on its own; it likely satisfies that inequality under suitable energy conditions imposed on classical matter configurations, but such energy conditions are generally violated by quantum matter. If an inequality like (1.12) is to hold for arbitrary states, we expect instead that quantum matter can cause the classical piece of the Markov gap to dip below the lower bound, but that the Markov gap of the quantum matter will make up the difference. One can think of this as a rather delicate energy condition on quantum fields, where quantum field theory states violating some classical energy conditions must have a sufficiently large $S_{R,\text{bulk}} - I_{\text{bulk}}$ to make up for the violation. Any research effort in this direction would likely start with the results of [71–73], where the authors compared reflected entropy and mutual information in simple quantum field theories.

6.2 Multipartite holographic entanglement

In the introduction, we mentioned the result of Akers and Rath [32] that any pure three-party state with a mostly-bipartite entanglement structure has Markov gap close to zero. Our inequality (1.12) can be viewed as a quantitative version of their observation that certain holographic states require tripartite entanglement. Any holographic state $|\Psi_{ABC}\rangle$ for which the entanglement wedge cross-section of ρ_{AB} has a boundary must have tripartite entanglement at order $1/G_N$; the scaling of inequality (1.12) seems to suggest — though not prove — that each boundary of the entanglement wedge cross-section requires some minimal amount of tripartite boundary entanglement to support it.

Let us consider this idea in light of the lore, originating in [40], that geometric connections in bulk states emerge from entanglement patterns in boundary states. What we would like to say is something like, “connections across codimension-2 surfaces are supported by bipartite entanglement in the boundary; connections across codimension-3 surfaces (like boundaries of entanglement wedge cross-sections), are supported by tripartite entanglement”.

At present, we do not have any concrete proposals for how to test this idea or make it more precise. We think it worthwhile to mention, however, a connection to another issue in holography: the difficulties involved in understanding intersections of extremal surfaces. The area operators of intersecting extremal surfaces do not commute. As emphasized in section 7 of [74], this means it is impossible to model the entanglement pattern of intersecting extremal surfaces with a tensor network made up of isometries or with fixed area states. It is generically the case that if one considers n overlapping boundary subregions of a holographic state, the quantum extremal surfaces computing their entropies will have

some complicated pattern of intersections in the bulk. The only tensor network model of holography we know of in which multipartite entanglement patterns have been computed explicitly is that of random stabilizer tensor networks; in that context, it was shown in [75] that tripartite entanglement is scarce. Because tensor network models *also* cannot model intersecting extremal surfaces, we might expect that the geometry of extremal surface intersections emerges from multipartite entanglement in the boundary state. This line of reasoning, if taken seriously, would provide an independent suggestion that codimension-3 and higher features of bulk geometries are signs of multipartite entanglement in the boundary theory.

Unfortunately, quantum extremal surface intersections appear to be a more complicated beast. After this manuscript first appeared online, we learned of forthcoming work [76] that finds a parametrically large Markov gap in *non-stabilizer* random tensor networks, indicating that these tensor networks contain multipartite entanglement. Despite this feature, these tensor networks are not capable of modeling the geometry of intersecting extremal surfaces. We still suspect that multipartite entanglement will be an important part of the story in understanding extremal surface intersections — but the mere existence of multipartite entanglement is not enough to make such intersections emerge.

Acknowledgments

It is a pleasure to thank Chris Akers, Ning Bao, Geoff Penington, Xiaoliang Qi, and Pratik Rath for illuminating conversations. We especially thank Mark Wilde for explaining equation (2.4) during the early stages of this work, and Arvin Shahbazi-Moghaddam for suggesting that bulk strong subadditivity could be used to establish the results of appendix A. This work was partially supported by AFOSR award FA9550-19-1-0369, CIFAR, DOE award DE-SC0019380 and the Simons Foundation.

A KRT surfaces and the quantum maximin formula

In the introduction, in inequality (1.11), we lower-bounded the Markov gap in terms of the generalized entropy differences of KRT and RT surfaces. While the Markov gap is known to be nonnegative from quantum information arguments (see section 2), it is interesting to see how this nonnegativity is enforced in the bulk theory. If we could establish the inequality

$$S_{\text{gen}}(\text{KRT}(A)|A) \geq S_{\text{gen}}(\text{RT}(A)|A), \quad (\text{A.1})$$

and the equivalent inequality for subregion B , then this would immediately imply nonnegativity of the Markov gap via inequality (1.11). Inequality (A.1) certainly seems reasonable, because $\text{RT}(A)$ has minimal generalized entropy among quantum extremal surfaces in its homology class. As emphasized in the introduction, however, $\text{KRT}(A)$ is not a quantum extremal surface, and so some care must be taken in establishing (A.1).

We will establish that the generalized entropy of $\text{KRT}(A)$ is non-increasing for arbitrary null perturbations towards the boundary region A , and then appeal to the quantum focusing conjecture [27] and quantum maximin [28] to establish inequality (A.1). For states

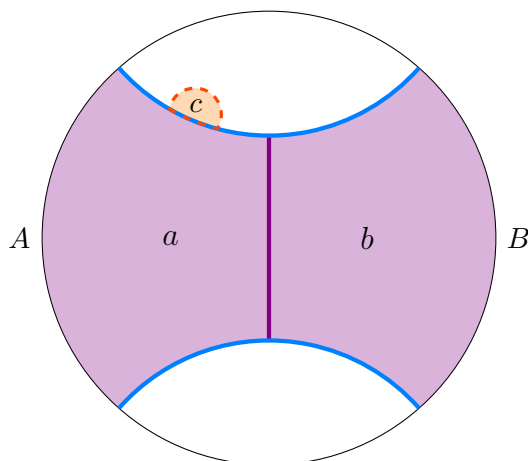


Figure 26. The bulk regions a and b for two intervals in the AdS_3 vacuum, together with a small deformation c localized on the surface $\text{RT}(AB)$. While this figure is drawn as though it lies in a single time-slice of the bulk, the deformation should actually lie in a null direction orthogonal to $\text{RT}(AB)$.

with no quantum matter, the result can be established via the ordinary focusing theorem and the classical maximin formula [8]. The key point will be that the bulk state lying between $\text{KRT}(A)$ and A can be extended into two different entanglement wedges — the AB entanglement wedge in the original state ρ_{AB} and the AA^* entanglement wedge in the canonical purification — that are each bounded by quantum extremal surfaces.

Let us denote by a the wedge lying spacelike between $\text{KRT}(A)$ and A , and by b the wedge lying spacelike between $\text{KRT}(B)$ and B . This is sketched for a simple spacetime in figure 26. Now, consider a null deformation of the portion of $\text{KRT}(A)$ that was inherited from the quantum extremal surface $\text{RT}(AB)$. For reasons that will become clear momentarily, we will choose this deformation to point *away* from A . This amounts to adding a small bulk region c to the bulk region a ; this deformation is also sketched in figure 26. The key observation is that this deformation can also be thought of as a deformation of the bulk region ab , whose boundary $\text{RT}(AB)$ is quantum extremal. The change in the generalized entropy of $\text{KRT}(A)$ under the addition of region c is given by

$$\Delta S_{\text{gen}}(\text{KRT}(A)|A) = S(ac) - S(a) + \frac{1}{4G_N} \Delta \text{area}(\text{KRT}(A)). \quad (\text{A.2})$$

But because the perturbation is localized to the portion of $\text{KRT}(A)$ contained within $\text{RT}(AB)$, we have $\Delta \text{area}(\text{KRT}(A)) = \Delta \text{area}(\text{RT}(AB))$. Furthermore, applying the strong subadditivity inequality to the system abc gives

$$S(ac) - S(a) \geq S(abc) - S(ab). \quad (\text{A.3})$$

Combining these observations yields the inequality

$$\Delta S_{\text{gen}}(\text{KRT}(A)|A) \geq S(abc) - S(ab) + \frac{1}{4G_N} \Delta \text{area}(\text{RT}(AB)) = \Delta S_{\text{gen}}(\text{RT}(AB)|AB). \quad (\text{A.4})$$

But the right-hand side of this inequality vanishes for infinitesimal regions c due to the quantum extremality of $\text{RT}(AB)$. It follows that the change in $S_{\text{gen}}(\text{KRT}(A)|A)$ due to a null deformation away from A is infinitesimally nonnegative, provided that this null deformation lies in $\text{RT}(AB)$. Linearity of perturbations to the entanglement entropy establishes that the change in $S_{\text{gen}}(\text{KRT}(A)|A)$ is infinitesimally *non-positive* for null perturbations on $\text{RT}(AB)$ pointed *toward* A , as claimed in the preceding paragraph.

For null perturbations lying on the portion of $\text{KRT}(A)$ contained within $\text{RT}(AA^*)$, the same exact result can be established using an embedding of the state into the entanglement wedge aa^* in the canonical purification. Putting these results together establishes that the generalized entropy of $\text{KRT}(A)$ is infinitesimally non-increasing for any null deformation toward A that is supported on the smooth parts of $\text{KRT}(A)$. As for null deformations toward A that are supported on the corners of $\text{KRT}(A)$, the right-angle structure of those corners guarantees that the classical expansion of any such deformation is $-\infty$. The infinitesimal change in the bulk entropy, since this is a renormalized quantity that should not be sensitive to corner divergences, is finite. As such, the total quantum expansion of such a perturbation is negative, establishing that the generalized entropy of $\text{KRT}(A)$ is infinitesimally non-increasing for *arbitrary* null deformations toward A . The quantum focusing conjecture [27] implies that if the change in generalized entropy along a null congruence is initially locally nonpositive, then it is nonpositive everywhere on the congruence,⁴⁰ provided that one takes the usual care to remove generators from the congruence after they reach caustic points. This implies that as $\text{KRT}(A)$ is focused along a null congruence leaving $\text{KRT}(A)$ in the direction of A , its generalized entropy is nonperturbatively non-increasing.

We now appeal to the quantum maximin formula [28], which claims the existence of a complete achronal slice of the bulk spacetime on which $\text{RT}(A)$ has minimal generalized entropy. The intersection of the null congruence leaving $\text{KRT}(A)$ with this slice is called the *representative* $\widehat{\text{KRT}}(A)$. We then have the chain of inequalities

$$S_{\text{gen}}(\text{RT}(A)|A) \leq S_{\text{gen}}(\widehat{\text{KRT}}(A)|A) \leq S_{\text{gen}}(\text{KRT}(A)|A), \quad (\text{A.5})$$

establishing inequality (A.1).

Open Access. This article is distributed under the terms of the Creative Commons Attribution License ([CC-BY 4.0](https://creativecommons.org/licenses/by/4.0/)), which permits any use, distribution and reproduction in any medium, provided the original author(s) and source are credited.

References

- [1] S. Dutta and T. Faulkner, *A canonical purification for the entanglement wedge cross-section*, *JHEP* **03** (2021) 178 [[arXiv:1905.00577](https://arxiv.org/abs/1905.00577)] [[INSPIRE](#)].
- [2] S. Ryu and T. Takayanagi, *Holographic derivation of entanglement entropy from AdS/CFT*, *Phys. Rev. Lett.* **96** (2006) 181602 [[hep-th/0603001](https://arxiv.org/abs/hep-th/0603001)] [[INSPIRE](#)].

⁴⁰The quantum focusing conjecture is the statement that the quantum expansion, which is the logarithmic derivative of the generalized entropy under a local null perturbation, is non-increasing on null congruences.

- [3] S. Ryu and T. Takayanagi, *Aspects of Holographic Entanglement Entropy*, *JHEP* **08** (2006) 045 [[hep-th/0605073](#)] [[INSPIRE](#)].
- [4] M. Headrick and T. Takayanagi, *A Holographic proof of the strong subadditivity of entanglement entropy*, *Phys. Rev. D* **76** (2007) 106013 [[arXiv:0704.3719](#)] [[INSPIRE](#)].
- [5] V.E. Hubeny, M. Rangamani and T. Takayanagi, *A Covariant holographic entanglement entropy proposal*, *JHEP* **07** (2007) 062 [[arXiv:0705.0016](#)] [[INSPIRE](#)].
- [6] M. Van Raamsdonk, *Comments on quantum gravity and entanglement*, [arXiv:0907.2939](#) [[INSPIRE](#)].
- [7] 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](#)] [[INSPIRE](#)].
- [8] A.C. Wall, *Maximin Surfaces, and the Strong Subadditivity of the Covariant Holographic Entanglement Entropy*, *Class. Quant. Grav.* **31** (2014) 225007 [[arXiv:1211.3494](#)] [[INSPIRE](#)].
- [9] A. Lewkowycz and J. Maldacena, *Generalized gravitational entropy*, *JHEP* **08** (2013) 090 [[arXiv:1304.4926](#)] [[INSPIRE](#)].
- [10] T. Faulkner, A. Lewkowycz and J. Maldacena, *Quantum corrections to holographic entanglement entropy*, *JHEP* **11** (2013) 074 [[arXiv:1307.2892](#)] [[INSPIRE](#)].
- [11] M. Headrick, V.E. Hubeny, A. Lawrence and M. Rangamani, *Causality & holographic entanglement entropy*, *JHEP* **12** (2014) 162 [[arXiv:1408.6300](#)] [[INSPIRE](#)].
- [12] A. Almheiri, X. Dong and D. Harlow, *Bulk Locality and Quantum Error Correction in AdS/CFT*, *JHEP* **04** (2015) 163 [[arXiv:1411.7041](#)] [[INSPIRE](#)].
- [13] D.L. Jafferis, A. Lewkowycz, J. Maldacena and S.J. Suh, *Relative entropy equals bulk relative entropy*, *JHEP* **06** (2016) 004 [[arXiv:1512.06431](#)] [[INSPIRE](#)].
- [14] X. Dong, A. Lewkowycz and M. Rangamani, *Deriving covariant holographic entanglement*, *JHEP* **11** (2016) 028 [[arXiv:1607.07506](#)] [[INSPIRE](#)].
- [15] N. Engelhardt and A.C. Wall, *Quantum Extremal Surfaces: Holographic Entanglement Entropy beyond the Classical Regime*, *JHEP* **01** (2015) 073 [[arXiv:1408.3203](#)] [[INSPIRE](#)].
- [16] X. Dong, D. Harlow and A.C. Wall, *Reconstruction of Bulk Operators within the Entanglement Wedge in Gauge-Gravity Duality*, *Phys. Rev. Lett.* **117** (2016) 021601 [[arXiv:1601.05416](#)] [[INSPIRE](#)].
- [17] J. Cotler, P. Hayden, G. Penington, G. Salton, B. Swingle and M. Walter, *Entanglement Wedge Reconstruction via Universal Recovery Channels*, *Phys. Rev. X* **9** (2019) 031011 [[arXiv:1704.05839](#)] [[INSPIRE](#)].
- [18] X. Dong and A. Lewkowycz, *Entropy, Extremality, Euclidean Variations, and the Equations of Motion*, *JHEP* **01** (2018) 081 [[arXiv:1705.08453](#)] [[INSPIRE](#)].
- [19] P. Hayden and G. Penington, *Learning the Alpha-bits of Black Holes*, *JHEP* **12** (2019) 007 [[arXiv:1807.06041](#)] [[INSPIRE](#)].
- [20] C. Akers and G. Penington, *Leading order corrections to the quantum extremal surface prescription*, *JHEP* **04** (2021) 062 [[arXiv:2008.03319](#)] [[INSPIRE](#)].
- [21] T. Takayanagi and K. Umemoto, *Entanglement of purification through holographic duality*, *Nature Phys.* **14** (2018) 573 [[arXiv:1708.09393](#)] [[INSPIRE](#)].

- [22] P. Nguyen, T. Devakul, M.G. Halbasch, M.P. Zaletel and B. Swingle, *Entanglement of purification: from spin chains to holography*, *JHEP* **01** (2018) 098 [[arXiv:1709.07424](#)] [[INSPIRE](#)].
- [23] J. Kudler-Flam and S. Ryu, *Entanglement negativity and minimal entanglement wedge cross sections in holographic theories*, *Phys. Rev. D* **99** (2019) 106014 [[arXiv:1808.00446](#)] [[INSPIRE](#)].
- [24] Y. Kusuki, J. Kudler-Flam and S. Ryu, *Derivation of Holographic Negativity in AdS_3/CFT_2* , *Phys. Rev. Lett.* **123** (2019) 131603 [[arXiv:1907.07824](#)] [[INSPIRE](#)].
- [25] Q. Wen, *Balanced Partial Entanglement and the Entanglement Wedge Cross Section*, *JHEP* **04** (2021) 301 [[arXiv:2103.00415](#)] [[INSPIRE](#)].
- [26] V. Chandrasekaran, M. Miyaji and P. Rath, *Including contributions from entanglement islands to the reflected entropy*, *Phys. Rev. D* **102** (2020) 086009 [[arXiv:2006.10754](#)] [[INSPIRE](#)].
- [27] R. Bousso, Z. Fisher, S. Leichenauer and A.C. Wall, *Quantum focusing conjecture*, *Phys. Rev. D* **93** (2016) 064044 [[arXiv:1506.02669](#)] [[INSPIRE](#)].
- [28] C. Akers, N. Engelhardt, G. Penington and M. Usatyuk, *Quantum Maximin Surfaces*, *JHEP* **08** (2020) 140 [[arXiv:1912.02799](#)] [[INSPIRE](#)].
- [29] P. Hayden, M. Headrick and A. Maloney, *Holographic Mutual Information is Monogamous*, *Phys. Rev. D* **87** (2013) 046003 [[arXiv:1107.2940](#)] [[INSPIRE](#)].
- [30] N. Bao, S. Nezami, H. Ooguri, B. Stoica, J. Sully and M. Walter, *The Holographic Entropy Cone*, *JHEP* **09** (2015) 130 [[arXiv:1505.07839](#)] [[INSPIRE](#)].
- [31] V.E. Hubeny, M. Rangamani and M. Rota, *Holographic entropy relations*, *Fortsch. Phys.* **66** (2018) 1800067 [[arXiv:1808.07871](#)] [[INSPIRE](#)].
- [32] C. Akers and P. Rath, *Entanglement Wedge Cross Sections Require Tripartite Entanglement*, *JHEP* **04** (2020) 208 [[arXiv:1911.07852](#)] [[INSPIRE](#)].
- [33] S.X. Cui, P. Hayden, T. He, M. Headrick, B. Stoica and M. Walter, *Bit Threads and Holographic Monogamy*, *Commun. Math. Phys.* **376** (2019) 609 [[arXiv:1808.05234](#)] [[INSPIRE](#)].
- [34] V. Balasubramanian, P. Hayden, A. Maloney, D. Marolf and S.F. Ross, *Multiboundary Wormholes and Holographic Entanglement*, *Class. Quant. Grav.* **31** (2014) 185015 [[arXiv:1406.2663](#)] [[INSPIRE](#)].
- [35] N. Bao and I.F. Halpern, *Conditional and Multipartite Entanglements of Purification and Holography*, *Phys. Rev. D* **99** (2019) 046010 [[arXiv:1805.00476](#)] [[INSPIRE](#)].
- [36] N. Bao, A. Chatwin-Davies and G.N. Remmen, *Entanglement of Purification and Multiboundary Wormhole Geometries*, *JHEP* **02** (2019) 110 [[arXiv:1811.01983](#)] [[INSPIRE](#)].
- [37] N. Bao and N. Cheng, *Multipartite Reflected Entropy*, *JHEP* **10** (2019) 102 [[arXiv:1909.03154](#)] [[INSPIRE](#)].
- [38] N. Bao, A. Chatwin-Davies and G.N. Remmen, *Entanglement Wedge Cross Section Inequalities from Replicated Geometries*, *JHEP* **07** (2021) 113 [[arXiv:2106.02640](#)] [[INSPIRE](#)].
- [39] Y. Zou, K. Siva, T. Soejima, R.S.K. Mong and M.P. Zaletel, *Universal tripartite entanglement in one-dimensional many-body systems*, *Phys. Rev. Lett.* **126** (2021) 120501 [[arXiv:2011.11864](#)] [[INSPIRE](#)].

- [40] M. Van Raamsdonk, *Building up spacetime with quantum entanglement*, *Gen. Rel. Grav.* **42** (2010) 2323 [[arXiv:1005.3035](#)] [[INSPIRE](#)].
- [41] C. Akers and P. Rath, *Holographic Renyi Entropy from Quantum Error Correction*, *JHEP* **05** (2019) 052 [[arXiv:1811.05171](#)] [[INSPIRE](#)].
- [42] X. Dong, D. Harlow and D. Marolf, *Flat entanglement spectra in fixed-area states of quantum gravity*, *JHEP* **10** (2019) 240 [[arXiv:1811.05382](#)] [[INSPIRE](#)].
- [43] M. Muller-Lennert, F. Dupuis, O. Szechr, S. Fehr and M. Tomamichel, *On quantum Renyi entropies: a new generalization and some properties*, *J. Math. Phys.* **54** (2013) 122203 [[arXiv:1306.3142](#)].
- [44] M.M. Wilde, A. Winter and D. Yang, *Strong Converse for the Classical Capacity of Entanglement-Breaking and Hadamard Channels via a Sandwiched Renyi Relative Entropy*, *Commun. Math. Phys.* **331** (2014) 593 [[arXiv:1306.1586](#)] [[INSPIRE](#)].
- [45] P. Hayden, R. Jozsa, D. Petz and A. Winter, *Structure of states which satisfy strong subadditivity of quantum entropy with equality*, *Commun. Math. Phys.* **246** (2004) 359 [[quant-ph/0304007](#)].
- [46] L. Accardi and A. Frigerio, *Markovian cocycles*, *Proc. Roy. Irish Acad. A* **83** (1983) 251.
- [47] D. Petz, *Sufficient subalgebras and the relative entropy of states of a von Neumann algebra*, *Commun. Math. Phys.* **105** (1986) 123 [[INSPIRE](#)].
- [48] O. Fawzi and R. Renner, *Quantum conditional mutual information and approximate Markov chains*, *Commun. Math. Phys.* **340** (2015) 575 [[arXiv:1410.0664](#)].
- [49] M. Junge, R. Renner, D. Sutter, M.M. Wilde and A. Winter, *Universal Recovery Maps and Approximate Sufficiency of Quantum Relative Entropy*, *Annales Henri Poincaré* **19** (2018) 2955 [[arXiv:1509.07127](#)] [[INSPIRE](#)].
- [50] M.M. Wilde, *From Classical to Quantum Shannon Theory*, [arXiv:1106.1445](#) [[INSPIRE](#)].
- [51] M.A. Nielsen and I. Chuang, *Quantum computation and quantum information*, *Am. J. Phys.* **70** (2002) 558.
- [52] J. Maldacena and L. Susskind, *Cool horizons for entangled black holes*, *Fortsch. Phys.* **61** (2013) 781 [[arXiv:1306.0533](#)] [[INSPIRE](#)].
- [53] G. Penington, *Entanglement Wedge Reconstruction and the Information Paradox*, *JHEP* **09** (2020) 002 [[arXiv:1905.08255](#)] [[INSPIRE](#)].
- [54] 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](#)] [[INSPIRE](#)].
- [55] G. Penington, S.H. Shenker, D. Stanford and Z. Yang, *Replica wormholes and the black hole interior*, [arXiv:1911.11977](#) [[INSPIRE](#)].
- [56] A. Almheiri, R. Mahajan, J. Maldacena and Y. Zhao, *The Page curve of Hawking radiation from semiclassical geometry*, *JHEP* **03** (2020) 149 [[arXiv:1908.10996](#)] [[INSPIRE](#)].
- [57] L. Anderson, O. Parrikar and R.M. Soni, *Islands with Gravitating Baths: Towards ER = EPR*, [arXiv:2103.14746](#) [[INSPIRE](#)].
- [58] X. Dong and H. Wang, *Enhanced corrections near holographic entanglement transitions: a chaotic case study*, *JHEP* **11** (2020) 007 [[arXiv:2006.10051](#)] [[INSPIRE](#)].

- [59] D. Marolf, S. Wang and Z. Wang, *Probing phase transitions of holographic entanglement entropy with fixed area states*, *JHEP* **12** (2020) 084 [[arXiv:2006.10089](#)] [[INSPIRE](#)].
- [60] W. Fenchel, *Elementary geometry in hyperbolic space*, De Gruyter (2011).
- [61] J. Sorce, *Holographic entanglement entropy is cutoff-covariant*, *JHEP* **10** (2019) 015 [[arXiv:1908.02297](#)] [[INSPIRE](#)].
- [62] B. Farb and D. Margalit, *A primer on mapping class groups*, Princeton University Press (2011).
- [63] A. Hatcher, *Algebraic topology*, <https://pi.math.cornell.edu/~hatcher/AT/AT.pdf> (2001).
- [64] A. Marden, *Outer Circles: An introduction to hyperbolic 3-manifolds*, Cambridge University Press (2007).
- [65] C. Akers, S. Leichenauer and A. Levine, *Large Breakdowns of Entanglement Wedge Reconstruction*, *Phys. Rev. D* **100** (2019) 126006 [[arXiv:1908.03975](#)] [[INSPIRE](#)].
- [66] W. Fischler, A. Kundu and S. Kundu, *Holographic Mutual Information at Finite Temperature*, *Phys. Rev. D* **87** (2013) 126012 [[arXiv:1212.4764](#)] [[INSPIRE](#)].
- [67] S. Colin-Ellerin, V.E. Hubeny, B.E. Niehoff and J. Sorce, *Large-d phase transitions in holographic mutual information*, *JHEP* **04** (2020) 173 [[arXiv:1911.06339](#)] [[INSPIRE](#)].
- [68] J.A. Mingo and A. Nica, *Annular noncrossing permutations and partitions, and second-order asymptotics for random matrices*, *Int. Math. Res. Not.* **2004** (2004) 1413 [[math/0303312](#)].
- [69] J. Kudler-Flam, *Relative Entropy of Random States and Black Holes*, *Phys. Rev. Lett.* **126** (2021) 171603 [[arXiv:2102.05053](#)] [[INSPIRE](#)].
- [70] H. Bateman, *Higher transcendental functions [volumes I-III]*, vol. I, McGraw-Hill Book Company (1953).
- [71] P. Bueno and H. Casini, *Reflected entropy, symmetries and free fermions*, *JHEP* **05** (2020) 103 [[arXiv:2003.09546](#)] [[INSPIRE](#)].
- [72] P. Bueno and H. Casini, *Reflected entropy for free scalars*, *JHEP* **11** (2020) 148 [[arXiv:2008.11373](#)] [[INSPIRE](#)].
- [73] H.A. Camargo, L. Hackl, M.P. Heller, A. Jahn and B. Windt, *Long Distance Entanglement of Purification and Reflected Entropy in Conformal Field Theory*, *Phys. Rev. Lett.* **127** (2021) 141604 [[arXiv:2102.00013](#)] [[INSPIRE](#)].
- [74] N. Bao, G. Penington, J. Sorce and A.C. Wall, *Beyond Toy Models: Distilling Tensor Networks in Full AdS/CFT*, *JHEP* **11** (2019) 069 [[arXiv:1812.01171](#)] [[INSPIRE](#)].
- [75] S. Nezami and M. Walter, *Multipartite Entanglement in Stabilizer Tensor Networks*, *Phys. Rev. Lett.* **125** (2020) 241602 [[arXiv:1608.02595](#)] [[INSPIRE](#)].
- [76] C. Akers, T. Faulkner, S. Lin and P. Rath, *Reflected entropy in random tensor networks*, to appear.