

Leading order corrections to the quantum extremal surface prescription

Chris Akers^a and Geoff Penington^{b,c}

^aCenter for Theoretical Physics, Massachusetts Institute of Technology,
Cambridge, MA 02139, U.S.A.

^bCenter for Theoretical Physics, University of California,
Berkeley, CA 94720, U.S.A.

^cGoogle,
Mountain View, CA 94303, U.S.A.

E-mail: cakers@mit.edu, geoffp@stanford.edu

ABSTRACT: We show that a naïve application of the quantum extremal surface (QES) prescription can lead to paradoxical results and must be corrected at leading order. The corrections arise when there is a second QES (with strictly larger generalized entropy at leading order than the minimal QES), together with a large amount of highly incompressible bulk entropy between the two surfaces. We trace the source of the corrections to a failure of the assumptions used in the replica trick derivation of the QES prescription, and show that a more careful derivation correctly computes the corrections. Using tools from one-shot quantum Shannon theory (smooth min- and max-entropies), we generalize these results to a set of refined conditions that determine whether the QES prescription holds. We find similar refinements to the conditions needed for entanglement wedge reconstruction (EWR), and show how EWR can be reinterpreted as the task of one-shot quantum state merging (using zero-bits rather than classical bits), a task gravity is able to achieve optimally efficiently.

KEYWORDS: AdS-CFT Correspondence, Nonperturbative Effects, Gauge-gravity correspondence

ARXIV EPRINT: [2008.03319](https://arxiv.org/abs/2008.03319)

Contents

1	Introduction	1
2	Mixtures and contradictions	7
3	Smooth min- and max-entropies	11
4	Replica trick calculations	15
4.1	Replica trick in holography	16
4.2	Fixed-area states and their use	17
4.3	Examples	22
5	Corrections in general holographic states	30
5.1	The regime of validity of naïve QES in fixed-area states	31
5.2	The regime where naïve QES fails	34
5.3	From fixed-area states to general holographic states	35
6	Entanglement wedge reconstruction	38
6.1	State-specific EWR	39
6.2	EWR as one-shot quantum state merging	41
7	Beyond two extremal surfaces	46
7.1	Applying the refined prescription	46
7.2	Properties of the min-EW and max-EW	47
7.3	Full reconstruction outside the max-EW	50
8	Discussion	52
A	Detailed evaluation of the mixture resolvent	55
B	Solving the quadratic resolvent	60
C	Numerics	63
D	One-shot decoupling	63

1 Introduction

The quantum extremal surface (QES) prescription [1] says that the entropy of a boundary region B in AdS/CFT is given by

$$S(B) = \min \text{ext}_\gamma \left[\frac{A(\gamma)}{4G} + S_{\text{bulk}}(\gamma) \right]. \quad (1.1)$$

Here we are extremizing over bulk surfaces γ that are homologous to B , $A(\gamma)$ is the area of the surface γ , G is Newton's constant and $S_{\text{bulk}}(\gamma) = -\text{tr}(\rho \ln(\rho))$ is the von Neumann entropy of the state ρ of the fields in the bulk region (known as the entanglement wedge) bounded by the surface γ and the boundary region B .¹ The combination of the two terms is known as the generalized entropy.

The perceived role of the S_{bulk} term has shifted over time. Because of the explicit factor of $1/G$, the area term becomes very large in the semiclassical limit $G \rightarrow 0$. The bulk entropy term, which has no such factor, was therefore initially regarded as a small, perturbative correction. In the last year or two, this view has changed. It has become clear that the QES prescription is still valid — and indeed plays a crucial role — even in situations where S_{bulk} is very large, and so competes with the area term. In particular, it was shown in [2, 3] that the QES prescription gives a unitary Page curve for the entanglement entropy of an evaporating black hole. This Page transition happens when the bulk entropy of the trivial ‘empty’ QES becomes larger than the area term for a non-trivial QES that lies near the horizon.

However, as we shall see, in such situations considerable care is needed when applying the QES prescription. For many states, a naïve application of the QES prescription gives contradictory answers, which are incompatible with basic properties of von Neumann entropies, even at leading order in $1/G$.

The primary aim of this paper is to (1) show that such contradictions exist, (2) show how the contradictions are resolved by more careful calculations, producing leading order corrections to the QES prescription, and (3) give general conditions for when the naïve QES prescription is valid, and when it needs to be replaced by a more refined version.

The contradictions can arise whenever there are two extremal surfaces, with $\mathcal{O}(1/G)$ bulk entropy in the intermediate region between the two. While common enough in AdS/CFT, this situation is also central to the phase transition that provides the Page curve of the evaporating black hole. Indeed, in section 2 we show that black hole evaporation can still lead to entropies inconsistent with unitarity, when using the naïve QES prescription.

Returning to AdS/CFT, a useful example setup was given in [4]. Consider 2+1d AdS, with the boundary divided into four regions as shown in figure 1. Let two diametrically opposed regions be slightly larger than the other two, such that the union of those two, named B , has a connected entanglement wedge in the absence of bulk entropy. The complement of the boundary region B shall be labelled \bar{B} . There are two extremal surfaces homologous

¹In this paper, we will maintain the traditional fiction that bulk subregions are associated with subsystems of the bulk Hilbert space, and hence that we can construct the reduced state on a subregion by taking a partial trace. In reality, bulk subregions should instead be associated with von Neumann subalgebras of bulk operators.

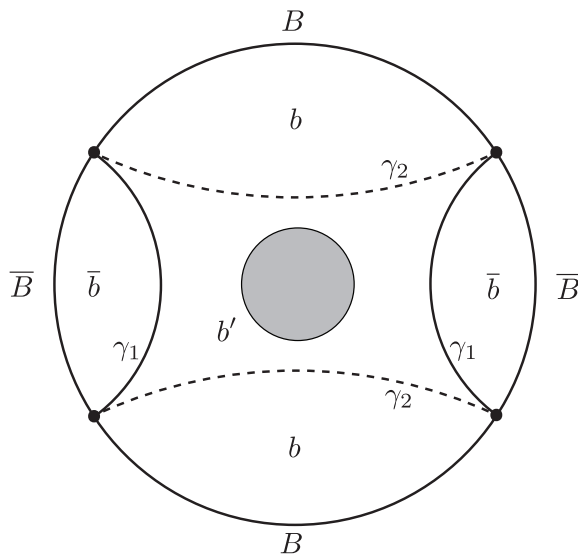


Figure 1. Setup in which we derive a contradiction from a naïve application of the QES prescription. The boundary is divided into two subregions, B and \bar{B} . For both, there are two competing quantum extremal surfaces, γ_1 and γ_2 , with γ_1 homotopic to \bar{B} and γ_2 to B . We take B to be larger, such that the area of γ_2 is bigger than that of γ_1 at $\mathcal{O}(1)$. Between these surfaces is a large amount of matter (the “dustball”), such that some states of the matter have entropy much larger than the difference in areas of the two surfaces.

to B : one homotopic to \bar{B} and labelled γ_1 , and one homotopic to B and labelled γ_2 . These surfaces divide the bulk into three regions: one named b that neighbours B , one named \bar{b} that neighbours \bar{B} , and a central region labelled by b' that is bounded by the two extremal surfaces. Let there be matter in b' with energy $\mathcal{O}(\varepsilon/G)$, for some $\varepsilon \ll 1$.

The backreaction is under control for small enough ε , and the bulk matter can have entropy roughly equal to its energy. We can therefore easily dial the size of B such that some bulk states have a bulk entropy larger than the area difference, while all states have the same approximate classical geometry.

Consider two states: in the first, the bulk matter is in a pure energy eigenstate. The matter therefore does not contribute to S_{bulk} , and the entanglement wedge of B is connected. In the second, the matter is in a thermal state with the same average energy. We tune the region B such that the large entropy of the thermal state causes its entanglement wedge to be disconnected. Hence, the von Neumann entropies are

$$\text{Matter pure:} \quad S(B) = A_1/4G, \tag{1.2}$$

$$\text{Matter thermal:} \quad S(B) = A_2/4G. \tag{1.3}$$

Here A_1 and A_2 are the areas of γ_1 and γ_2 respectively.

Now we can formulate the contradiction. What is $S(B)$ for a state that is a mixture of the pure state and the thermal state? In other words,

$$\rho_{\text{matter}} = p|\psi\rangle\langle\psi| + (1-p)\rho_{\text{thermal}}. \tag{1.4}$$

A naïve application of the QES prescription tells us that, at leading order, the answer is

$$\text{Mixture: } S(B)_{\text{naïve}} = \min \left(A_1/4G + (1-p)S_{\text{thermal}}, A_2/4G \right). \quad (1.5)$$

However, this can't be correct. The AdS/CFT bulk-to-boundary map is linear, so the global boundary state must also be a mixture of the two boundary states. And, if the global state is a mixture of the two states, the reduced state will also be a mixture of the two reduced states. In general, the von Neumann entropy $S(\rho)$ of a mixture of quantum states

$$\rho = \sum_i p_i \rho_i \quad (1.6)$$

of density matrices ρ_i is bounded from above and below by

$$\sum_i p_i S(\rho_i) \leq S(\rho) \leq \sum_i p_i S(\rho_i) - p_i \ln p_i, \quad (1.7)$$

see e.g. [5].² Together, the bounds (1.7) are quite restrictive, forcing a mixture of k states to have entropy at most $\mathcal{O}(\ln k)$ different than the average entropy of those states. In particular, the entropy of a mixture of $\mathcal{O}(1)$ states is within $\mathcal{O}(1)$ of the average entropy within the mixture.

So in our example, it must be the case that the correct (or “refined”) answer is

$$\text{Mixture: } S(B)_{\text{refined}} = pA_1/4G + (1-p)A_2/4G + \mathcal{O}(1). \quad (1.8)$$

This is generally different from the naïve QES at leading order, $\mathcal{O}(1/G)$. See figure 2.

Why does the naïve QES fail for the mixture when (we claim) it gives the correct answer for both the pure state and the thermal state individually? Intuitively, this is because it is a mixture of states that are on different sides of the phase transition. But this notion is not very precise: the thermal state can itself be written as a mixture of states that are on either side of the transition (admittedly in this case one either needs a large number of states or some probabilities in the mixture to be very small), and yet it doesn't receive large corrections.

A more precise answer is that, unlike the pure state and the thermal state, the mixture of the two is not *perfectly compressible*. We say a state ρ is perfectly compressible if we can throw away all but $e^{S(\rho)}$ of the states in its support without changing the state very much. More precisely, there must exist another state σ close to ρ such that $\ln \text{rank}(\sigma) = S(\rho) + \text{subleading}$. In the thermodynamic limit, thermal states are dominated by energies close to the saddle point energy, and are therefore perfectly compressible. A pure state has rank one and hence is trivially perfectly compressible.

A general mixture of the two is not: any state σ close to ρ_{matter} will have almost the same rank as an approximation to the thermal state itself, because

$$\frac{1}{1-p} \sigma - \frac{p}{1-p} |\psi\rangle \langle \psi| \approx \rho_{\text{thermal}}. \quad (1.9)$$

²The lower bound formalizes an intuitive fact: the uncertainty of a mixture of states must be at least as large as the average uncertainty of each of those states. The upper bound is true because ρ must have less entropy than a state that includes a correlated reference system with orthonormal basis $|i\rangle$, $\sum p_i \rho_i \otimes |i\rangle \langle i|$. The lower (upper) bound is saturated if and only if the ρ_i are all identical (all orthogonal).

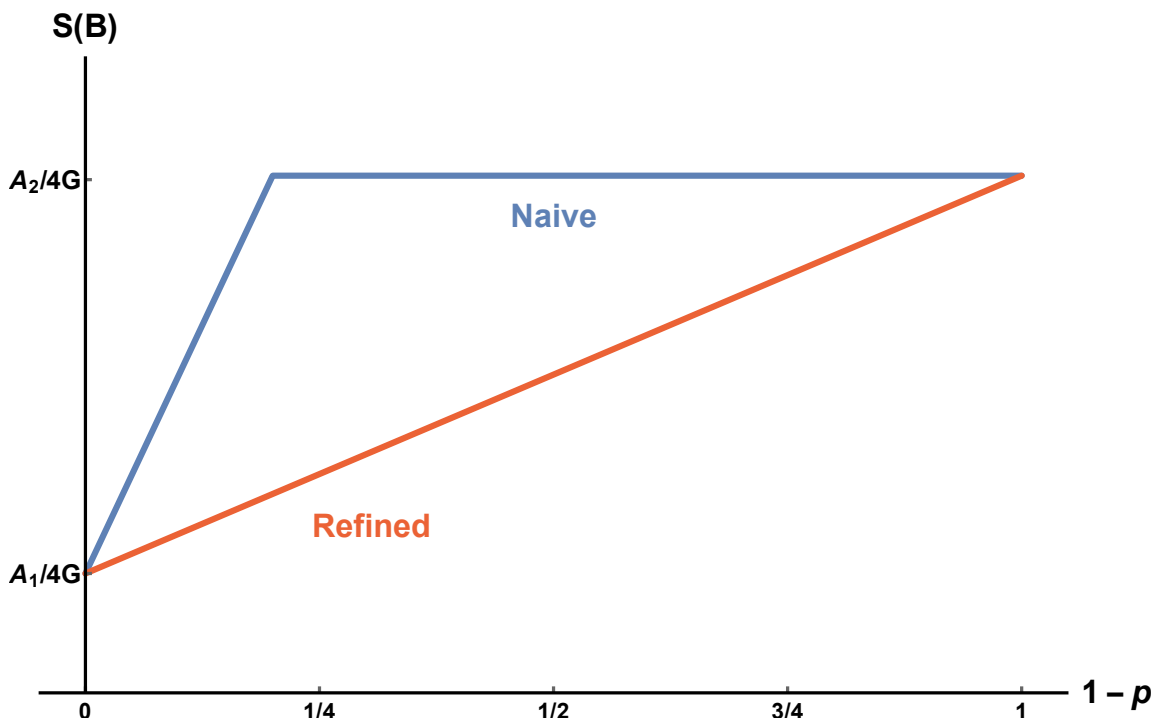


Figure 2. Comparison of the naïve QES entropy to the correct, “refined” answer, for the state (1.4) in the setup of figure 1. While the slope of the refined answer is controlled by $(A_2 - A_1)/4G$, the slope of the naïve answer is controlled by $S_{\text{thermal}} > (A_2 - A_1)/4G$. The naïve answer is in general larger than the refined one by an $\mathcal{O}(1/G)$ amount.

In general the compressibility of a quantum state is characterized not by its von Neumann entropy, but by a quantity known as the smooth max-entropy $H_{\text{max}}^\varepsilon(\rho)$ [6, 7]. This is defined by the fact that you can throw away at most all but $e^{H_{\text{max}}^\varepsilon(\rho)}$ of the states in the support of ρ , without changing ρ very much, for small ε . For thermal and pure states, the smooth max-entropy is approximately equal to the von Neumann entropy — implying those states are perfectly compressible — but in general it can be much larger. For example, in the mixture of a thermal and pure state, we have³

$$H_{\text{max}}^\varepsilon(\rho) \approx \frac{S(\rho)}{1-p}. \tag{1.10}$$

To understand why this should be relevant to the QES prescription, we need to introduce the concept of entanglement wedge reconstruction (EWR) [10–12]. This says that the bulk matter in the entanglement wedge is encoded in the boundary state on the boundary subregion B . “Encoded,” here, means that the set of bulk operators local to the entanglement wedge has a representation on B that acts faithfully on a “code” subspace of BB . It turns out EWR is implied by the QES prescription [8, 12].⁴

³In general, we take ε to be polynomially small in G , though its exact size does not matter much. What’s important is that it’s not exponentially small, because we will need to assume $\ln \varepsilon \ll \mathcal{O}(1/G)$. Physically, this is closely related to the fact that bulk reconstruction necessarily has exponentially small errors [8, 9].

⁴More carefully, EWR (as usually defined) is possible if and only if the QES prescription holds for every state (pure or mixed) in the *code subspace* of states for which the reconstruction is supposed to be valid [8].

EWR (and hence QES) is deeply connected to compressibility. The intuition is that the number of degrees of freedom available in B to describe the bulk state in region b' is given by the difference in areas $(A_2 - A_1)/4G$ between the two extremal surfaces. If the bulk state in region b' cannot be compressed into these degrees of freedom, EWR for region b' cannot be possible, and hence the QES prescription, with γ_1 the minimal QES, cannot be valid, even if γ_1 is the surface with the smallest generalized entropy. One of the main goals of this paper will be to formalize this intuition by showing that EWR can be reinterpreted as a particular information-theoretic task, called *one-shot quantum state merging*, where Alice has to communicate a compressed version of a quantum state to Bob.

To make a precise statement, detailing how the QES prescription needs to be modified given the discussion above, it is helpful to write the naïve QES prescription in the following form:

$$S(B)_{\text{naïve}} = \begin{cases} A_1/4G + S(bb'), & S(b'|b) \leq \frac{A_2 - A_1}{4G} \\ A_2/4G + S(b), & S(b'|b) \geq \frac{A_2 - A_1}{4G}. \end{cases} \quad (1.11)$$

The quantity $S(b'|b) = S(bb') - S(b)$ is the conditional von Neumann entropy.

We will argue that this naïve prescription only works when bulk states are perfectly compressible, because it implies the inclusion (or not) of b' in the bulk entropy term only depends on the von Neumann entropy $S(b'|b)$. In reality, the information from b' is only accessible in B (and hence its entropy is only included in $S(B)$) when the quantum information in b' can be compressed into $(A_2 - A_1)/4 \ln(2)G$ qubits. The relevant bulk entropy is therefore not the conditional von Neumann entropy $S(b'|b)$, but the conditional smooth max-entropy $H_{\text{max}}^\varepsilon(b'|b)$. We'll explain $H_{\text{max}}^\varepsilon(b'|b)$ in detail in section 3 (along with the smooth conditional min-entropy $H_{\text{min}}^\varepsilon(b'|b)$), but, roughly speaking, $H_{\text{max}}^\varepsilon(b'|b)$ characterizes the compressibility of b' when there is entanglement between b' and b (and $H_{\text{min}}^\varepsilon(b'|b)$ is complementary to $H_{\text{max}}^\varepsilon(b'|b)$). For all states, we have $H_{\text{max}}^\varepsilon(b'|b) \geq S(b'|b) \geq H_{\text{min}}^\varepsilon(b'|b)$.

A central result of this paper will be to refine the conditions for the QES prescription (1.11), replacing it by⁵

$$S(B)_{\text{refined}} = \begin{cases} A_1/4G + S(bb'), & H_{\text{max}}^\varepsilon(b'|b) \leq \frac{A_2 - A_1}{4G} \\ \text{(depends on details),} & H_{\text{min}}^\varepsilon(b'|b) \leq \frac{A_2 - A_1}{4G} \leq H_{\text{max}}^\varepsilon(b'|b) \\ A_2/4G + S(b), & H_{\text{min}}^\varepsilon(b'|b) \geq \frac{A_2 - A_1}{4G}. \end{cases} \quad (1.12)$$

We will not give a “one answer fits all” description of the middle regime; it does not admit one as convenient as the naïve QES prescription. The entropy there depends on the details of the bulk entanglement. (That said, one can often estimate the answer by finding the average entropy of a set of constituent states, up to a Shannon term.) This refinement can be derived using replica trick calculations, and resolves the contradictions discussed above.

A heuristic way to understand the difference between these two prescriptions is that our refinement of the QES prescription recognizes that different parts of the wavefunction might be on different sides of the phase transition, whereas the naïve prescription assumes

⁵We defer to section 7 for more general conditions, applicable to setups with more than two competing QES.

that the entire state has to be on one side or the other. The min-/max-entropies appear because they describe the largest/smallest parts of the wavefunction respectively. If the smooth max-entropy is less than $(A_2 - A_1)/4G$, we can be sure that no significant part of the wavefunction has undergone the transition. Similarly, if the smooth min-entropy is greater than $(A_2 - A_1)/4G$, we know that almost the entire wavefunction has undergone the transition. If they straddle $(A_2 - A_1)/4G$, then the entropy will depend on which parts of the wavefunction have crossed the transition.

Overview of paper. The paper is organized as follows.

In section 2, we illustrate the problem with a naïve application of the QES prescription in more detail. We give several closely related examples of the naïve QES prescription violating the bounds on the von Neumann entropy of mixtures of states.

In section 3, we review two quantities that are crucial for understanding the refined QES prescription: the smooth conditional min-entropy $H_{\min}^\varepsilon(A|B)$ and max-entropy $H_{\max}^\varepsilon(A|B)$.

In section 4, we return to the simple examples from sections 1 and 2 and carefully calculate their entropies using the replica trick. By avoiding using the Lewkowycz-Maldacena assumption, we find an answer that disagrees with the naïve QES prescription but is consistent with the bounds on entropy of mixtures. This answer depends on the relative sizes of three quantities: the smooth conditional min- and max-entropy, and the difference in area of the two competing quantum extremal surfaces.

In section 5, we present *general* arguments that justify the conditions given in (1.12) for the existence of large corrections to the naïve QES prescription. We start by arguing this for so-called fixed-area states, and then argue that this extends to general holographic states, up to subleading corrections. A key tool is the connection between gravity calculations in fixed-area states and calculations in random tensor networks.

In section 6, we update the conditions for entanglement wedge reconstruction (EWR), explaining how to generalize the results of Dong, Harlow, Wall [12] and Hayden, Penington [8], given this refinement of the QES prescription. These updated conditions clarify the relationship between EWR and a well-known quantum information task, one-shot quantum state merging. Our results demonstrate that EWR can be a maximally efficient form of one-shot quantum state merging, using zero-bits instead of the usual classical bits.

In section 7, we present a more general refinement of the QES prescription conditions, applying in situations where there are more than two competing extremal surfaces. To do so, we first introduce two interesting new physically relevant subregions of the bulk: the min-entanglement wedge (min-EW) and max-entanglement wedge (max-EW). The naïve QES prescription applies if and only if the min-EW and max-EW are the same.

In section 8, we mention some further implications of these results. In particular, we discuss how the smooth min- and max-entropies should be renormalized to get a UV-finite quantity.

Related work. This paper has some technical overlap with the recent papers [13, 14]. They too find corrections to the QES prescription by carefully including more than one saddle in the replica trick, and they too use fixed-area states to simplify the calculation enough to do so.

There are three key differences between our corrections and theirs. One, the corrections we discuss can be $\mathcal{O}(1/G)$, not just $\mathcal{O}(1/\sqrt{G})$. Two, our corrections can exist for an $\mathcal{O}(1)$ range of $A_2 - A_1$, a window that does not vanish as $G \rightarrow 0$. Finally, our corrections do not arise from fluctuations in the geometry, but rather from the bulk state affecting the boundary entropy in a different way than previously expected. In particular, the corrections in [13, 14] are correctly computed by the expectation of the naïve QES prescription over all the classical geometries that can be created by the fluctuations.

We also provide a different argument justifying the use of fixed-area states in lessons about general states. Our argument also applies to the setups in [13, 14], bounding the error in some of their assumptions.

2 Mixtures and contradictions

In this section, we further illustrate the need for a careful, refined application of the QES prescription, first generalizing the prior example by adding entanglement, then discussing the importance of the refinement for black hole entropy and the unitarity of black hole evaporation.

Contradiction 1: dustball. Our first example is the dustball geometry, which was already presented in the introduction. However, we emphasize that many of the details, as presented there, were unimportant. The contradiction can easily be generalized to higher-dimensions, to mixtures where neither state is pure, or to mixtures of a larger number of states (so long as the number is not exponential in $1/G$).

We also note that we can easily adapt this example to find a similar contradiction where the state in b' is highly entangled with the state in b , eventually illustrating the necessity of using *conditional* min- and max-entropies in (1.12).

The first step is to consider a purification of the mixed dustball state, where the dustball is entangled with a second, identical dustball in a different bulk spacetime, as in figure 3. In other words, where the bulk state is

$$\rho = p \left(|\psi\rangle_{b'} \langle\psi|_{b'} \otimes |\psi\rangle_r \langle\psi|_r \right) + (1 - p) |\Phi\rangle \langle\Phi|, \tag{2.1}$$

and $|\Phi\rangle$ is a purification of ρ_{thermal} . From a boundary perspective, the mixed CFT state is purified by a second identical CFT, which we shall call the reference system R . Consistent with our notational conventions, we use r to denote the bulk Hilbert space associated to the second CFT.

Introducing the reference system R does not change the entropy $S(B)$. However, since the overall state is pure, we have $S(\overline{BR}) = S(B)$. The entropy $S(\overline{BR})$ can also be calculated using the naïve QES prescription. This time, the degrees of freedom in the homology region shared by both extremal surfaces (in this case $\overline{b} \otimes r$) are entangled with the degrees of freedom between the two surfaces (region b' as before). Unsurprisingly, the naïve QES prescription gives the same answers as before, and hence we again find a contradiction.

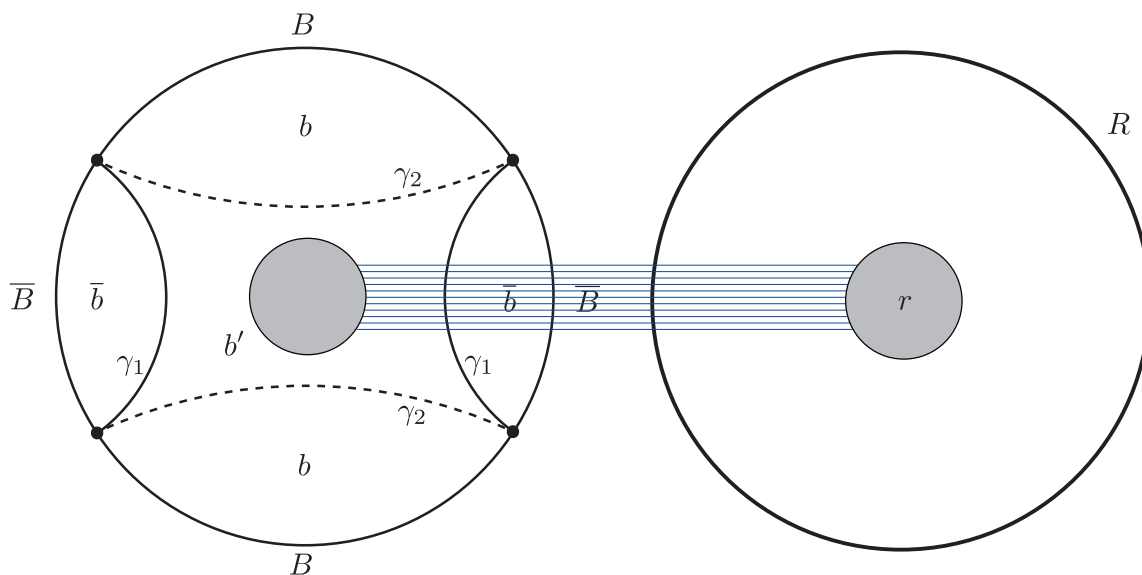


Figure 3. Two entangled dustballs. Like figure 1, but now we consider the entropy of \overline{BR} , where R is an entire extra copy of the boundary, dual to its own dustball. The two dustballs are in a mixture of entangled states, given by (2.1). A naïve application of the QES prescription gives the wrong answer for the entropy $S(\overline{BR})$.

Contradiction 2: black hole. A practically identical setup reaches the same contradiction, if we replace the dustballs with black holes [8]. See figure 4 for the setup with a mixed state black hole (though we also consider two entangled black holes, which would look very similar to figure 3).

The advantage of this setup is that it's familiar to consider black holes with entropy growing with $1/G$. We can, for example, consider all states in an energy band of width $\Delta E \sim \mathcal{O}(1)$, centered on some high energy E . There are $e^{\mathcal{O}(1/G)}$ states in this subspace, and generic density matrices in this band are expected to be black holes. Additionally, unlike the dustball, we can also use a single interval (in $\text{AdS}_3/\text{CFT}_2$) for our boundary region, because the black hole geometry has extremal surfaces on either side of the black hole.

The big disadvantage — indeed the reason we did not lead with this example — is that black hole microstates seem somewhat mysterious. One might worry that mixtures of black hole states, like (2.1), are secretly mixtures of classically distinct geometries, mixtures which people already expected to give averaged answers in the QES prescription. For example, a special case of the mixture of entangled black holes is the mixture of an energy eigenstate and the thermofield double (TFD) state.⁶ The TFD state is

$$|\text{TFD}\rangle = \sum_i e^{-\beta E_i/2} |E_i\rangle_{b'} |E_i\rangle_r, \quad (2.2)$$

for some inverse-temperature β and energy eigenstates $|E_i\rangle$. Two black holes entangled like this are connected by a wormhole [15], and hence there is a nontrivial homology constraint. This is very different from a factorized energy eigenstate, which has trivial

⁶Perhaps projected onto its dominant energy window, to fit into a finite-dimensional subspace.

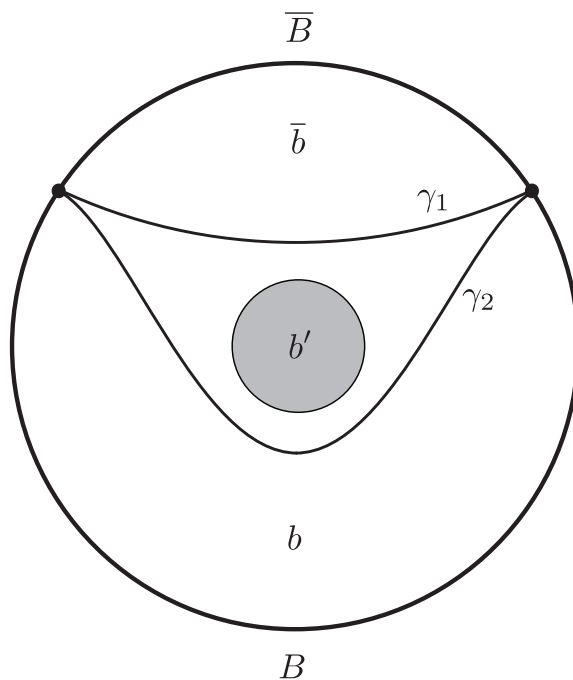


Figure 4. Black hole setup in which we derive a contradiction from the QES prescription. Practically identical to the setup of figure 1, this setup replaces the dustball with a black hole, and its boundary regions B and \bar{B} are now connected.

homology constraints. The mixture of the two,

$$\text{special case: } \rho_{b'r} = p \left(|E\rangle_{b'} \langle E|_{b'} \otimes |E\rangle_r \langle E|_r \right) + (1-p) |\text{TFD}\rangle \langle \text{TFD}|, \quad (2.3)$$

must therefore have the QES prescription applied to it with care, since it is a mixture of two distinct classical geometries. There is a history of speculating that — for this state — the naïve QES prescription gives an $S(B)$ that is indeed the average entropy $(1-p)A/4G$ (see e.g. [16]). The argument was that the area operator is linear, and so its expectation value in this mixture of states must be the average of its expectation value in each.

While that argument is fine, we emphasize that it does not explain away the contradictions we are pointing out. This can be made sharp using the insights from quantum error-correction in [17].⁷ From the quantum error-correction point of view, it is not necessary to count the black hole entropy as part of the “area.” A choice of code subspace that includes the black hole microstates will regard the black hole entropy as part of the matter entropy. This would be inconsistent, giving an answer that does not equal $S(B) = (1-p)A/4G$, if the entropy of a mixture of black hole states is given by the naïve QES answer.

⁷Another way to see this is from the UV-finiteness of the generalized entropy $A/4G + S$, which implies that G acts as a counterterm for the divergences in the von Neumann entropy, suggesting that as one flows to different energy scales, one changes which degrees of freedom contribute to $A/4G$ rather than S . We thank Netta Engelhardt for emphasizing this to us.

Contradiction 3: Hawking radiation. Our final contradiction appears in evaporating black holes. It was shown last year that using the QES prescription allows a gravity calculation of the decrease in entropy of Hawking radiation after the Page time [2, 3]. This goes a long way towards resolving the famous black hole information paradox. However, there’s a lingering paradox in those calculations, if the QES prescription is applied in the naïve way. We demonstrate this now.

Consider a post-Page time black hole B , having already emitted radiation R in state ρ_R . Introduce an ancilla qubit q , and entangle it with R in the following way. First, put q in a superposition

$$\sqrt{1-p}|0\rangle_q + \sqrt{p}|1\rangle_q. \tag{2.4}$$

Then, perform a joint operation on qR , measuring the radiation if q is in state $|1\rangle$, and otherwise doing nothing. This measurement need not be complicated — a factorized measurement on each Hawking photon is simple and will suffice. Given measured state $|\psi\rangle_R$, the reduced state of the radiation becomes

$$\tilde{\rho}_R = p|\psi\rangle_R\langle\psi|_R + (1-p)\rho_R. \tag{2.5}$$

Assuming that the evaporating black hole was following the Page curve, the entropy of the radiation, at leading order, will then be $(1-p)A_{\text{hor}}/4G$ (+ subleading), where A_{hor} is the area of the black hole horizon.

What does the naïve QES prescription say that the entropy will be? As long as we don’t measure the most recent Hawking quanta to escape into R , the locations of the quantum extremal surfaces will be unchanged. The generalized entropy of the empty surface will be $(1-p)S_{\text{rad}}$, where S_{rad} is the semiclassical, thermal entropy of the radiation. The generalized entropy of the nonempty surface near the horizon will be $A_{\text{hor}}/4G$ as before.

As with our previous contradictions, this is just incorrect (assuming unitarity), even at leading order. The naïve QES prescription is giving an answer that is qualitatively just as wrong as the Hawking, information-loss answer. Indeed, for small values of $(1-p)$, the naïve QES prescription answer and the Hawking answer are the same.

A very similar contradiction can be created using purely unitary processes, without any measurements. One just creates an ancilla system A , in the state $|0\rangle$, that is a copy of the radiation Hilbert space R . Then one applies a conditional swap operator (which again factorizes into a product of local interactions) that swaps A and R if and only if the qubit q is in the state $|1\rangle$. Assuming unitarity, the form of the reduced state on R will again be given by (2.5). The generalized entropy of the empty surface will again be $(1-p)S_{\text{rad}}$, while the generalized entropy of the nonempty surface will be $A_{\text{hor}}/4G + pS_{\text{rad}}$. Again, we find a contradiction with unitarity at leading order.

Summary. This section showed classes of examples in which a naïve application of the QES prescription gets the entropy wrong at leading order. In section 4 we do a careful calculation that gets the entropy in these examples right, and then in section 5 we describe more generally when and why there are corrections. First, however, we need to introduce two quantities that will characterize when the naïve QES prescription receives these large corrections.

3 Smooth min- and max-entropies

While naïvely the QES prescription compares only von Neumann entropies to areas, we will find that a more careful prescription compares to the area two other quantities: the smooth conditional min-entropy and smooth conditional max-entropy. These information-theoretic quantities have historically found use in “one-shot” protocols, settings in which only a single copy of a quantum state is used or transferred.

We explain these quantities now, and in all future sections refer to them heavily. We start with the simplest version, the classical min- and max-entropy, and gradually work up to what we really want, the (quantum) smooth conditional min- and max-entropy.

Non-conditional versions. To introduce the idea of one-shot entropies, it is helpful to temporarily forget about quantum mechanics and simply consider classical probability distributions.

Let us first recall the information-theoretic role of the Shannon entropy $S(p)$ of a classical probability distribution $p(x)$ (analogous to the von Neumann entropy in quantum mechanics). Imagine you randomly sample from a large number of copies n of the probability distribution, getting outcomes $\{x_i\}$. You, Alice, now want to communicate those outcomes to your friend Bob.

How much information do you need to send to Bob to do this? To always be successful, for any $\{x_i\}$, you need to send at least $n \log_2 d$ bits, where d is the number of values x can take with nonzero probability. However, if you only insist that the communication succeed with high probability (i.e. succeed for a variety of possible outcomes $\{x_i\}$ that collectively have probability $p > 1 - \varepsilon$ for some small ε), the task becomes much easier. One can show that, at leading order for large n , you only need to send $n S(p) / \ln(2)$ bits. Essentially, this comes from the law of large numbers ensuring that ‘typical’ samples from many copies of the distribution have a probability p such that⁸

$$\ln p = n \langle \ln p(x) \rangle_{p(x)} + o(n) = n S(p) + o(n). \quad (3.1)$$

Hence, you and Bob simply need to agree on a code, in which the $n S(p) / \ln(2)$ bits you send tell Bob which of the $e^{n S(p) + o(n)}$ “typical strings” you sampled.

The story in quantum mechanics is very similar: given any density matrix ρ , we can project $\rho^{\otimes n}$ into a ‘code subspace’, while only changing the state a small amount. This code subspace is just built out of products of states in the Schmidt decomposition of ρ that have typical entropy, as in the classical case. Such states dominate the Schmidt decomposition of $\rho^{\otimes n}$ at large n .

The number of qubits needed for the code subspace grows, in the limit of large n , as $n S(\rho) / \ln(2)$, where $S(\rho)$ is the von Neumann entropy. If Alice has a pure state randomly sampled from $\rho^{\otimes n}$, she can therefore communicate that state to Bob with high success probability, just by sending $n S(\rho) / \ln(2)$ qubits.

However, both in classical probability and in quantum mechanics, we often (perhaps even typically) encounter situations where we don’t have a large number of copies of a

⁸The notation “ $o(n)$ ” represents terms subleading to n , vanishing in $\lim_{n \rightarrow \infty} o(n)/n$.

single density matrix or distribution. Instead, we only have a single state or distribution, which may still be very large in size. An example, of course, is holography. In the limit $G \rightarrow 0$, the boundary Hilbert space dimension blows up exponentially, but this does not mean we have a large number of independent copies of the same state.

In this ‘one-shot’ setting, the von Neumann entropy does not have an important operational role.⁹ It is therefore somewhat surprising that the von Neumann entropy has been playing such a central role in holography, for example in determining whether entanglement wedge reconstruction is possible! As we shall see, the resolution is that the real quantities that are important in holography are smooth max- and min-entropies, which do have a natural operational interpretation in one-shot quantum Shannon theory. It just so happens that these ‘one-shot entropies’ have been approximately equal to the von Neumann entropy, in most of the situations that have been considered in the literature until now.

Suppose we consider the same task as above (sending the outcome of sampling a probability distribution from Alice to Bob), but now we only sample from a single copy of the distribution. How many bits do we need to send to communicate the outcome with high probability? We need to be able to send a distinct message for each outcome that we want to be successfully communicated, and our success probability is maximized by choosing the outcomes with the highest probability of occurring. So the number of bits that need to be sent is $\log_2 N^{(\varepsilon)}$ where $N^{(\varepsilon)}$ is the smallest integer such that

$$\sum_{i=1}^{N^{(\varepsilon)}} p_i > 1 - \varepsilon, \tag{3.2}$$

with the probabilities p_i ordered from largest to smallest.

Again, there is an obvious quantum mechanical generalization, which gives the minimum number of qubits needed to send a quantum state, sampled from a single copy of a density matrix ρ , from Alice to Bob. This is given by

$$H_0^\varepsilon(\rho) = \inf_{\|\tilde{\rho}-\rho\|_1 \leq \varepsilon} H_0(\tilde{\rho}) = \inf_{\|\tilde{\rho}-\rho\|_1 \leq \varepsilon} \ln(\text{Rank}(\tilde{\rho})). \tag{3.3}$$

Let’s unpack this for a moment. We first defined the Rényi 0-entropy (also known as the Hartley entropy) as

$$H_0(\rho) = \ln \text{Rank}(\rho) = \lim_{\alpha \rightarrow 0} \frac{1}{1 - \alpha} \ln \text{tr}(\rho^\alpha), \tag{3.4}$$

and then we ‘smoothed’ this quantity by minimizing it over all $\tilde{\rho}$ close to ρ (which in this case just meant throwing away small eigenvalues). We measured this distance with the trace distance, or Schatten 1-norm, $\|X\|_1 = \text{tr}(\sqrt{X^\dagger X})$.

In fact, (3.3) is the original definition of the smooth max-entropy [7]. It turns out however that $H_0(\rho)$ can be replaced [19] by the Rényi entropy

$$H_\alpha(\rho) = \frac{1}{1 - \alpha} \ln \text{tr}(\rho^\alpha), \tag{3.5}$$

⁹For an exception to this general principle, see [18].

for *any* $\alpha < 1$, while only changing the smooth entropy by a small amount. Specifically,

$$H_0^\varepsilon(\rho) \geq H_\alpha^\varepsilon(\rho) \geq H_0^{2\varepsilon}(\rho) - \frac{1}{1-\alpha} \ln(1/\varepsilon). \quad (3.6)$$

As we shall see below, $H_{1/2}(\rho)$ generalizes better to conditional entropies. It is therefore conventionally used in the modern definition of the smooth max-entropy [20],

$$H_{\max}^\varepsilon(\rho) = \inf_{\tilde{\rho} \in \mathcal{B}^\varepsilon(\rho)} H_{1/2}(\tilde{\rho}), \quad (3.7)$$

where we are taking an infimum over all states $\tilde{\rho}$ within an ε -ball $\mathcal{B}^\varepsilon(\rho)$ of ρ .¹⁰

In summary, the number of qubits needed to send Bob your quantum state with high fidelity, if you only sample the distribution one time, is the smooth max-entropy (3.7), up to the factor of $\ln(2)$. The smooth max-entropy is always greater than or equal to the von Neumann entropy; sending many samples from the distribution can only improve the efficiency of the communication rate.

We can also define a complementary quantity, the *smooth min-entropy*, as

$$H_{\min}^\varepsilon(\rho) = \sup_{\tilde{\rho} \in \mathcal{B}^\varepsilon(\rho)} H_\infty(\tilde{\rho}). \quad (3.8)$$

Again, $H_\infty(\tilde{\rho})$ could be replaced by $H_\alpha(\tilde{\rho})$ for any $\alpha > 1$ while changing the definition by at most $\mathcal{O}(\ln(1/\varepsilon))$. It's operational interpretation is less intuitive than the smooth max-entropy, so we motivate it simply by its relationship to the conditional max-entropy, as we'll explain. Note that the smooth min-entropy is always less than or equal to the von Neumann entropy.

Together, these two quantities establish upper and lower bounds on the confidence interval for the value of (non-negligible) eigenvalues of ρ . The smooth max-entropy encodes the size of the *smallest* eigenvalues in the density matrix (which cannot be thrown away with small error), while the smooth min-entropy captures the size of the *largest* eigenvalues (that cannot be thrown away).

If the spectrum is close to flat (i.e. is dominated by a small range of eigenvalues) then the smooth min- and max- entropies will be close to the von Neumann entropy (which characterizes the *average* (log-)eigenvalue). In particular, thanks to the law of large numbers, this happens at leading order in n when you take a large number of copies $\rho^{\otimes n}$ of a state ρ . This explains the importance of the von Neumann in traditional asymptotic quantum Shannon theory, which deals with exactly this limit.

It is also what has led to the success (so far) of the naïve QES prescription; it's been used for bulk states with an (approximately) flat spectrum, where the smooth min- and max-entropy are roughly the same as the von Neumann entropy.

¹⁰For technical reasons, the distance measure used to define this ε -ball is conventionally the *purified distance*, defined as the minimum trace distance between purifications of ρ and $\tilde{\rho}$. However, again, any reasonable distance measure will work fine (up to unimportant changes in the scaling of ε).

Conditional versions. The most general quantities we will need are the smooth *conditional* min- and max-entropies, which generalize the conditional von Neumann entropy. Unfortunately, the definition of these quantities is somewhat more technical, and somewhat less intuitive, than their unconditional counterparts.

The operational spirit of these quantities is the following. Let us return to the example in which Alice is trying to send a quantum state on A to Bob. However, now the state is sampled from a density matrix ρ_{AB} , where subsystem B is already held by Bob and the two subsystems may be entangled. Can this entanglement help Alice send her part of the state to Bob? It can! For a particular version of this task, called quantum state merging [21], the number of qubits that need to be sent from Alice to Bob is the smooth conditional max-entropy $H_{\max}^\varepsilon(A|B)$, which is generally less than $H_{\max}^\varepsilon(A)$. We discuss quantum state merging in detail in section 6.

Here are the technical definitions. The conditional von Neumann entropy, which the conditional min- and max-entropy generalize, is normally defined as

$$S(A|B) = S(AB) - S(B). \quad (3.9)$$

However, this definition does not generalize well to smooth entropies. Instead, our starting point will be a definition of the conditional entropy in terms of the relative entropy as

$$S(A|B) = -\min_{\sigma_B} D(\rho_{AB}|\mathbb{1}_A \otimes \sigma_B). \quad (3.10)$$

To see that this is equivalent to (3.9), note that

$$\begin{aligned} D(\rho_{AB}|\mathbb{1}_A \otimes \sigma_B) &= \text{tr}(\rho_{AB} \ln \rho_{AB}) - \text{tr}(\rho_B \ln \sigma_B) \\ &= -S(AB) + S(B) + D(\rho_B|\sigma_B) \\ &\geq -S(AB) + S(B) \end{aligned} \quad (3.11)$$

with equality if $\sigma_B = \rho_B$.

Smooth conditional min-entropy. To generalize (3.10) to a smooth conditional min-entropy, we use the fact that there is a unique quantum generalization of the classical Rényi max-divergence $D_\infty(\rho|\sigma)$ which satisfies the data-processing inequality and additivity. This is given by

$$D_\infty(\rho|\sigma) = \inf\{\lambda : \rho \leq e^\lambda \sigma\}. \quad (3.12)$$

In words, the quantum max divergence of ρ relative to σ is the smallest number λ such that $e^\lambda \sigma - \rho$ is positive semi-definite.

We then define the conditional min-entropy as

$$H_{\min}(A|B)_\rho = -\min_{\sigma_B} D_\infty(\rho_{AB}|\mathbb{1}_A \otimes \sigma_B), \quad (3.13)$$

and the smooth conditional min-entropy as

$$H_{\min}^\varepsilon(A|B)_\rho = \sup_{\tilde{\rho} \in \mathcal{B}^\varepsilon(\rho)} H_{\min}(A|B)_{\tilde{\rho}}. \quad (3.14)$$

We can gain some intuition by rewriting the conditional min-entropy as

$$H_{\min}^{\varepsilon}(A|B) = \inf_{\tilde{\rho} \in \mathcal{B}^{\varepsilon}(\rho)} \left(-\ln |A| - \sup_{\Phi_B} \ln F((\mathbb{1}_A \otimes \Phi_B)\tilde{\rho}_{AB}, \tau_{AA'}) \right), \quad (3.15)$$

where F is the fidelity $F(\rho, \tau) = \left(\text{tr} \sqrt{\rho^{1/2} \tau \rho^{1/2}} \right)^2$, $\tau_{AA'}$ is a maximally entangled state on two copies of A , and Φ_B is a completely positive trace preserving map from B to A' . This illustrates that $H_{\min}^{\varepsilon}(A|B)$, in a sense, quantifies how close ρ_{AB} is to a maximally entangled state, equaling its minimum $-\ln |A|$ when A is maximally entangled with B , and its maximum $\ln |A|$ when it's completely decoupled.

Smooth conditional max-entropy. The smooth conditional max-entropy is most cleanly defined as a complement to the smooth conditional min-entropy. Recall that for any purification $|\rho\rangle_{ABC}$ of ρ_{AB} , we have

$$S(A|B) = -S(A|C). \quad (3.16)$$

A generalization of this equality will define the smooth conditional max-entropy. One can show that

$$-H_{\min}(A|C) = \ln |A| + \sup_{\sigma_B} \ln F \left(\rho_{AB}, \frac{\mathbb{1}_A}{|A|} \otimes \sigma_B \right). \quad (3.17)$$

This right hand side is a natural candidate definition for $H_{\max}(A|B)$. We can test this by considering the special case where subsystem B is trivial (i.e. the state on AC is pure). We then have

$$-H_{\min}(A|C) = 2 \ln \text{tr} \rho^{1/2} = H_{1/2}(A). \quad (3.18)$$

Recall that we previously used $H_{1/2}(A)$ in our formal definition of the smooth max-entropy. It is indeed therefore natural to define the smooth conditional max-entropy $H_{\max}^{\varepsilon}(A|B)$ as

$$H_{\max}^{\varepsilon}(A|B) = -H_{\min}^{\varepsilon}(A|C) = \inf_{\tilde{\rho} \in \mathcal{B}^{\varepsilon}(\rho)} \left(\ln |A| + \sup_{\sigma_B} \ln F(\tilde{\rho}_{AB}, \frac{\mathbb{1}_A}{|A|} \otimes \sigma_B) \right). \quad (3.19)$$

This definition provides some intuition for the smooth conditional max-entropy, as quantifying, in a sense, how close ρ_{AB} is to a decoupled state $\mathbb{1}_A/|A| \otimes \sigma_B$, equaling $\ln |A|$ when A is completely decoupled from B , and $-\ln |A|$ when it is maximally entangled with B .

4 Replica trick calculations

The replica trick is a standard technique for computing the von Neumann entropy $S(B)$, based on interpreting $\text{tr}(\rho_B^n)$ as a certain observable in n copies (or replicas) of the system [22]. We first illustrate the standard technique for computing $\text{tr}(\rho^n)$ holographically — using the saddle point approximation and analytically continuing the dominant saddle — which is well-known to give the naïve QES prescription [9, 23–26]. We then do a more careful calculation, where we analytically continue a sum over an entire family of saddles. To make this calculation analytically tractable, we make use of the fixed-area states of [27, 28]. This more careful calculation gives results that differ from the naïve QES answer and avoid any contradictions.

4.1 Replica trick in holography

Given a state $\rho_{B\bar{B}}$, the goal is to compute

$$S(B) \equiv -\text{tr}(\rho_B \ln \rho_B) = \lim_{n \rightarrow 1} \frac{1}{1-n} \ln \text{tr}(\rho_B^n). \quad (4.1)$$

The last equality is useful because $\text{tr}(\rho_B^n)$ can be computed using a path integral. Schematically, the Euclidean path integral preparing $\rho_{B\bar{B}}$ looks like,

$$\rho_{B\bar{B}} = \sum_i \left[\begin{array}{c} \text{shaded hemisphere } B \\ \text{unshaded hemisphere } \bar{B} \end{array} \right] \iff \sum_i \left[\begin{array}{c} \text{circle } B \\ \text{circle } \bar{B} \end{array} \right]. \quad (4.2)$$

The final picture is just a more schematic version of the first. The orange dots, with index i , label a basis of states, prepared by different boundary conditions, that are summed over. This represents the fact that a general density matrix is not just a product of a ket and a bra, but a sum of such products. In future diagrams, we suppress this index and the sum.

To construct the reduced density matrix ρ_B , we glue together \bar{B} in the bras and kets:

$$\rho_B = \left[\begin{array}{c} \text{circle } B \\ \text{circle } \bar{B} \end{array} \right] \iff \left[\begin{array}{c} \text{circle } B \\ \text{circle } B \end{array} \right] \iff \left[\begin{array}{c} \text{arc } B \\ \text{arc } B \end{array} \right] \quad (4.3)$$

Then the path integral for e.g. $\text{tr}(\rho_B^3)$ involves gluing together the different copies of B “cyclically” as

$$\text{tr}(\rho_B^3) = \left[\begin{array}{c} \text{circle } B \\ \text{circle } B \\ \text{circle } B \end{array} \right] \quad (4.4)$$

These boundary conditions can be applied with a “twist operator” τ , which acts on n copies of the B Hilbert space to cyclically permute the state on each copy:

$$\text{tr}(\rho_B^n) = \text{tr}(\rho_{B\bar{B}}^{\otimes n} \tau). \quad (4.5)$$

We call this n -replica geometry \mathcal{M}_n . By evaluating the path integral on \mathcal{M}_n for arbitrary n , and analytically continuing the answer to the limit $n \rightarrow 1$, we can compute the entanglement entropy.

We can map this boundary path integral to a bulk computation using the AdS/CFT dictionary,

$$\text{tr} \left(\rho_{B\bar{B}}^{\otimes n} \tau \right) = \frac{Z_{B,n}}{Z_{B,1}^n}, \tag{4.6}$$

where $Z_{B,n}$ is the bulk partition function, defined by integrating over all bulk geometries with boundary \mathcal{M}_n . In the semiclassical limit, this can be approximated by a sum over classical saddles. Crucially, the saddle-point geometries are not simply n copies of the original geometry glued together. They are whatever the equations of motion provide, given that boundary data.

Partially for this reason, and partially because the number of saddles depends on n , this sum over saddles is generally too difficult to evaluate, let alone analytically continue. So historically, the following trick was used [23]. *Assume* that a replica-symmetric configuration dominates the sum, and that all other contributions to the path integral can be ignored, such that

$$Z_{B,n} \approx e^{-I_{\text{grav}}[g_{s,n}]} Z_{B,n}^{\text{mat}}[g_{s,n}], \tag{4.7}$$

where $g_{s,n}$ is the saddle-point metric, $I_{\text{grav}}[g_{s,n}]$ is the gravitational action, and $Z_{B,n}^{\text{mat}}[g_{s,n}]$ is the matter partition function on this semiclassical background. We shall call this *the Lewkowycz-Maldacena (LM) assumption*. Because the saddle is replica-symmetric, we can equivalently consider the quotient of the saddle-point geometry by the Z_n replica symmetry. This is also a solution to the equations of motion, except at the fixed-points of the Z_n action, where there is a conical singularity with opening angle $2\pi/n$.

We can now analytically continue the quotiented geometry to non-integer values of n . In particular, in the limit $n \rightarrow 1$, the geometry approaches the original *unbackreacted* geometry, with a weak conical singularity at the Z_n fixed-points. The entanglement entropy ends up being the generalized entropy of the Z_n fixed-points [24], which is forced to be a quantum extremal surface by the equations of motion [9, 25, 26]. The dominant semiclassical saddle is the one where the QES has the smallest generalized entropy, leading to the naïve QES prescription.

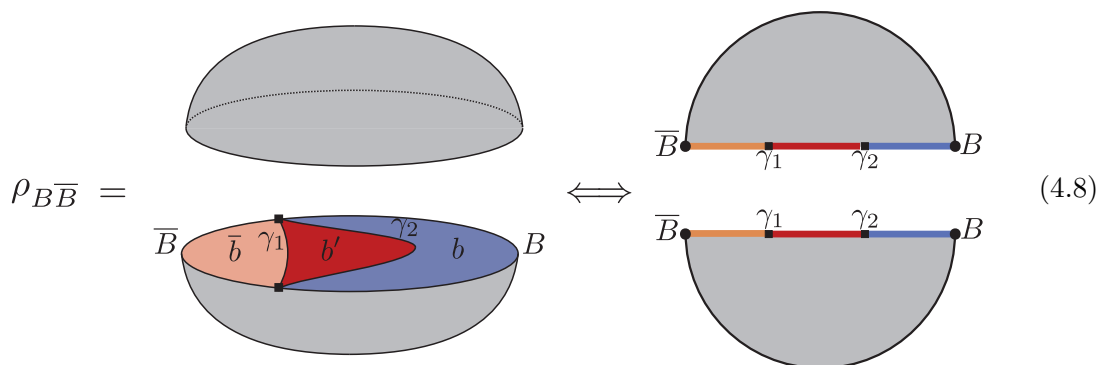
Since this traditional derivation reaches a conclusion that we have shown is contradictory, the obvious next step is to do the replica trick more carefully, without the weak link of the LM assumption. This requires we introduce some other simplifying trick to analytically continue the sum over saddles. This trick will involve the use of fixed-area states, which we now explain.

4.2 Fixed-area states and their use

The fixed-area states of [27, 28] are (approximate) eigenstates of certain area operators. To define such a state, consider the Euclidean path integral that prepares a particular bulk geometry, then insert into that path integral a delta function that fixes the area of

some gauge-invariantly defined surfaces.¹¹ We might physically prepare such a state by measuring the area of these surfaces. Saddle-points of this restricted path integral must satisfy the bulk equations of motion everywhere except at the fixed-area surfaces, where they may have a conical singularity. This is because the conical deficit angle is conjugate to the area operator and is therefore *undetermined* in fixed-area states, due to the uncertainty principle.¹²

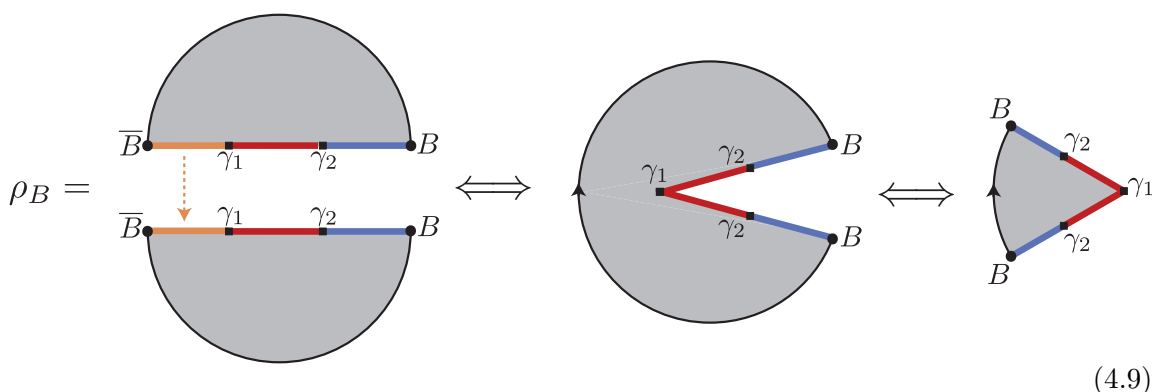
Replica trick for fixed-area states. Consider a state $\rho_{B\bar{B}}$ with two fixed-area surfaces, γ_1 and γ_2 . We depict its path integral as



Fixing the areas in the initial state is a boundary condition and so also fixes the areas of that surface in path integrals featuring any number of replicas of that geometry. This is what makes the sum over geometries in $Z_{B,n}$ doable.

Indeed, we can form geometries that satisfy all boundary conditions of $Z_{B,n}$ — asymptotic boundary \mathcal{M}_n plus fixed areas of all fixed-area surfaces — simply by gluing together n copies of the original $n = 1$ bulk around the fixed-area surfaces.

Since we glue the boundary region \bar{B} together in the bra and the ket path integral to make the density matrix ρ_B , the neighbouring bulk region \bar{b} (shown in orange) is also always glued together



¹¹Of course, in reality the area cannot be measured exactly. Instead, we must specify to what precision we fix the area. It will be sufficient to fix it to be within a window that is polynomially small in G .

¹²This uncertainty in the geometry makes it impossible to fix the area of two overlapping surfaces — but there is no problem simultaneously fixing the area of two surfaces that do not cross (and therefore have commuting area operators) [29, 30].

Similarly, because we glue the boundary regions B together cyclically, the bulk regions b get glued together cyclically. However, because we can have conical singularities at γ_1 and γ_2 , the different copies of the region b' can be glued together using an arbitrary permutation $\pi \in S_n$. To evaluate the full path integral, we sum over all saddles, and hence sum over all permutations π . For example,

$$\frac{Z_{B,3}}{Z_{B,1}^3} = \text{[Diagram 1]} + \text{[Diagram 2]} + \dots + \text{[Diagram 3]} \tag{4.10}$$

Since the replica geometry consists of n copies of the original unbackreacted geometry, the gravitational action away from the fixed-area surfaces cancels between the numerator $Z_{B,n}$ and the denominator $Z_{B,1}^n$. The only contribution to the gravitational action that doesn't cancel out is the contribution to the Einstein-Hilbert action from the conical singularities, which are different in the numerator and the denominator. Each conical singularity gives a contribution equaling $(\phi - 2\pi)A/8\pi G$, where ϕ is the opening angle of the conical singularity.

If the b' regions are glued together using a permutation π , the full contribution to the action from the conical singularities in the replica geometry is therefore

$$(n\phi_1 - 2\pi C(\pi)) \frac{A_1}{8\pi G} + (n\phi_2 - 2\pi C(\tau^{-1} \circ \pi)) \frac{A_2}{8\pi G}, \tag{4.11}$$

where $C(g)$ is the number of cycles in the permutation g , and ϕ_1, ϕ_2 are the conical singularity angles associated to γ_1, γ_2 in the unreplicated geometry. After normalization, the dependence on ϕ cancels. Including the matter partition function, we are then left with

$$\frac{Z_{B,n}}{Z_{B,1}^n} = \text{tr}(\rho_B^n) = \sum_{\pi \in S_n} e^{(C(\pi)-n)A_1/4G + (C(\tau^{-1} \circ \pi)-n)A_2/4G} \text{tr}(\rho_{bb'}^{\otimes n} \tau_b \pi_{b'}). \tag{4.12}$$

This further simplifies because we do not need to sum over all S_n . Any permutation that does not maximize $C(\pi) + C(\tau^{-1} \circ \pi)$ corresponds to an action subleading by factors of the area. The areas A_1 and A_2 are IR divergent, so those permutations are *infinitely* suppressed. The remaining permutations lie on the geodesic in the Cayley graph (i.e. shortest path in permutation space, where each step is a transposition) connecting τ and the identity. These are the so-called “non-crossing” permutations NC_n , which all satisfy $C(\pi) + C(\tau^{-1} \circ \pi) = n + 1$ (see e.g. [31]).

Without the $\text{tr}(\rho_{bb'}^{\otimes n} \tau_b \pi_{b'})$ factor, we could evaluate this sum explicitly. The number of non-crossing permutations with $C(\tau^{-1} \circ \pi) = k$ is the Narayana number $N(n, k)$. With that, we could organize the terms into a sum over k and get an analytic answer in terms of hypergeometric functions.

The bulk term interferes because it depends not just on the number of cycles $C(\tau^{-1} \circ \pi)$, but also on the number of elements *per* cycle. Fortunately, there is a way, presented in [9], to reorganize this sum into one over the number of elements per cycle.

Resolvent method. To make the calculation tractable, we will need to assume that the entropy of b is small and can be ignored,¹³ so that (4.12) becomes

$$\text{tr}(\rho_B^n) = \sum_{\pi \in NC_n} e^{(C(\pi)-n)A_1/4G+(1-C(\pi))A_2/4G} \text{tr}(\rho_{b'}^{\otimes n} \pi_{b'}). \quad (4.13)$$

Here, we have used the fact that $C(\pi) + C(\tau^{-1} \circ \pi) = n + 1$ for non-crossing permutations to rewrite the formula without $C(\tau^{-1} \circ \pi)$.

Define the resolvent R_{ij} of ρ_B as

$$R_{ij}(\lambda) = \left(\frac{1}{\lambda \mathbb{1} - \rho_B} \right)_{ij} \quad (4.14)$$

This contains all the data about the eigenvalues of ρ_B . For example, the density $D_B(\lambda)$ of eigenvalues of ρ_B is

$$D_B(\lambda) = -\frac{1}{\pi} \lim_{\epsilon \rightarrow 0^+} \text{Im} R(\lambda + i\epsilon), \quad (4.15)$$

where R is the trace of the resolvent.

We compute this as follows, heavily using the fact that $\rho_{B\bar{B}}$ is a fixed-area state. First, Taylor expand (4.14) around $\rho_B = 0$ to obtain

$$R(\lambda)_{ij} = \frac{1}{\lambda} \delta_{ij} + \sum_{n=1}^{\infty} \frac{1}{\lambda^{n+1}} (\rho_B^n)_{ij}. \quad (4.16)$$

We can visualize this as

$$\text{Diagrammatic expansion of } R \text{ as a sum of terms with increasing powers of } \rho_B. \quad (4.17)$$

¹³This is a stronger assumption than necessary, but is valid in the examples we'll care about. More generally, this method can work as long as either ρ_b or $\rho_{\bar{b}}$ has a flat Renyi spectrum.

Each dashed line comes with a factor of $1/\lambda$. Then substitute for ρ_B^n equation (4.9),

The diagram shows the expansion of a circular trace R (represented by a circle with a dashed arrow) into a sum of terms. The first term is a single dashed line. The second term is a triangle with red solid lines and blue dashed lines. The third term is a figure-eight shape with red solid lines and blue dashed lines. The fourth term is a more complex shape with red solid lines and blue dashed lines. The expansion continues with an ellipsis.

$$(4.18)$$

Taking the trace of this quantity — visualized as simply connecting the dangling blue arrows into a closed loop — gives the equation

$$R = \frac{\text{rank}(\rho_B)}{\lambda} + \sum_{n=1}^{\infty} \sum_{\pi \in NC_n} \frac{1}{\lambda^{n+1}} e^{(C(\pi)-n)A_1/4G+(1-C(\pi))A_2/4G} \text{tr}(\rho_{b'}^{\otimes n} \pi_{b'}). \quad (4.19)$$

We can reorganize these sums in a convenient way, to get a Schwinger-Dyson equation:

The diagram shows the Schwinger-Dyson equation for R . On the left is a circle with a dashed arrow labeled R . This is equal to a sum of terms: a dashed rectangle, a triangle with red solid lines and blue dashed lines, a figure-eight shape with red solid lines and blue dashed lines, and a more complex shape with red solid lines and blue dashed lines. The expansion continues with an ellipsis.

$$(4.20)$$

On the right hand side, the second term sums all non-crossing geometries in which the first replica of b' is glued to no other replicas. The third term sums all non-crossing geometries in which the first replica of b' is glued to exactly one other replica. And so on.

We now formally explain the diagrammatic expansion (4.20) in terms of equations. Starting with (4.19), decompose the sums into a sum over the number of elements m in the cycle of π that includes the first element (the “primary cycle”), as well as the number

of elements n_i between the i th and $i + 1$ th element of the primary cycle: that is,

$$\sum_{n=1}^{\infty} \sum_{\pi \in NC_n} \rightarrow \sum_{m=1}^{\infty} \sum_{n_1=0}^{\infty} \dots \sum_{n_m=0}^{\infty} \sum_{\pi_1 \in NC_{n_1}} \dots \sum_{\pi_m \in NC_{n_m}} . \quad (4.21)$$

The primary cycle is always cyclic, but the other permutations π_i may not be. We note that $n = m + \sum n_i$ and $C(\pi) = 1 + \sum C(\pi_i)$. Also,

$$\text{tr}(\rho_{b'}^{\otimes n} \pi_{b'}) = \text{tr}(\rho_{b'}^{\otimes m} \tau_m) \prod_{i=1}^m \text{tr}(\rho_{b'}^{\otimes n_i} \pi_{n_i}) . \quad (4.22)$$

Plugging these into the formula (4.19) for the resolvent gives

$$R = \frac{\text{rank}(\rho_B)}{\lambda} + \sum_{m=1}^{\infty} \frac{e^{A_1/4G} \text{tr}(\rho_{b'}^m)}{\lambda e^{m(A_1+A_2)/4G}} \prod_{i=1}^m \left(\frac{e^{A_2/4G}}{\lambda} + \sum_{n_i=1}^{\infty} \sum_{\pi_i \in NC_{n_i}} \frac{1}{\lambda^{n_i+1}} e^{(C(\pi_i)-n_i)A_1/4G+(1-C(\pi_i))A_2/4G} \text{tr}(\rho_{b'}^{\otimes n_i} \pi_i) \right) . \quad (4.23)$$

The part in parenthesis is R itself, from (4.19), assuming $\text{rank}(\rho_B) = e^{A_2/4G}$.¹⁴ Therefore,

$$R = \frac{e^{A_2/4G}}{\lambda} + \sum_{m=1}^{\infty} \frac{e^{A_1/4G} \text{tr}(\rho_{b'}^m) R^m}{\lambda e^{m(A_1+A_2)/4G}} . \quad (4.24)$$

We are now ready to work out some specific examples.

4.3 Examples

Example 1: mixed states.

Setup. Consider the setup with a dustball or a black hole, from sections 1 or 2, depicted in figures 1 and 4 respectively. We will simultaneously compute $S(B)$ in both cases, first modifying the setups slightly by fixing the areas of γ_1 and γ_2 to A_1 and A_2 . All parameters we mention below apply equally well to both: e.g. $\rho_{b'}$ is the state of either the dustball or the black hole.

The same calculations also give the entropy of the *black hole* in our third contradiction from section 2. A particularly concrete example, where the full non-perturbative path integral can be evaluated and agrees with the answer that we find below, is the JT gravity plus end-of-the-world (EOW) brane model of [9]. In this case, working with fixed-area states is equivalent to working in the microcanonical ensemble (note that A_2 here is the horizon area of the black hole, while $A_1 = 0$), and the only bulk degrees of freedom are on the EOW brane in region b' , which is the assumption that we needed above to make the resolvent calculation possible.

¹⁴Since we can shift R by A/λ for a real constant A , without changing $D(\lambda)$ away from $\lambda = 0$, we are always free to assume this.

Consider two bulk states, $\rho_{b',1} = |\psi\rangle\langle\psi|$ pure and $\rho_{b',2}$ an arbitrary orthogonal mixed state of entropy S . We will assume that the state $\rho_{b',2}$ has a flat spectrum, and hence is perfectly compressible. We will compute the entropy $S(B)$ for their mixture,

$$\rho_{b'} = p|\psi\rangle\langle\psi| + (1-p)\rho_{b',2}. \quad (4.25)$$

To keep the example as simple as possible, we assume

$$(1-p)e^{-S} \ll p. \quad (4.26)$$

This ensures that the bulk density matrix eigenvalues p and $(1-p)e^{-S}$ are separated by a large multiplicative factor.¹⁵ The von Neumann, min-, and max-entropies of this state are

$$\begin{aligned} H_{\min}^\varepsilon(b') &\approx -\ln(p), \\ S(b') &= (1-p)S - p\ln(p) - (1-p)\ln(1-p), \\ H_{\max}^\varepsilon(b') &\approx S. \end{aligned} \quad (4.27)$$

The naïve QES prescription says

$$S_{\text{naïve}}(B) = \begin{cases} \frac{A_1}{4G} + S(b'), & S(b') \ll \frac{A_2 - A_1}{4G} \\ \frac{A_2}{4G}, & S(b') \gg \frac{A_2 - A_1}{4G}. \end{cases} \quad (4.28)$$

We will see that the correct answer, up to $\mathcal{O}(1)$ corrections, is

$$S_{\text{refined}}(B) = \begin{cases} \frac{A_1}{4G} + S(b'), & H_{\max}^\varepsilon(b') \ll \frac{A_2 - A_1}{4G} \\ p\frac{A_1}{4G} + (1-p)\frac{A_2}{4G}, & H_{\min}^\varepsilon(b') \ll \frac{A_2 - A_1}{4G} \ll H_{\max}^\varepsilon(b') \\ \frac{A_2}{4G}, & H_{\min}^\varepsilon(b') \gg \frac{A_2 - A_1}{4G}. \end{cases} \quad (4.29)$$

Calculation. Plug (4.25) into (4.24) and evaluate the two geometric sums to arrive at

$$\lambda R = e^{A_2/4G} + \frac{pR}{e^{A_2/4G} - \frac{p}{e^{A_1/4G}}R} + \frac{(1-p)R}{e^{A_2/4G} - \frac{(1-p)}{e^{A_1/4G+S}}R}. \quad (4.30)$$

The roots are the function $R(\lambda)$ that we seek. As a cubic equation, its roots can be written analytically but are difficult to integrate to compute the entropy. Fortunately, we can find a simple approximate solution, by using the assumption (4.26).

We expand (4.30) in two different ways, which are valid at large R (and hence small λ) and small R (large λ) respectively. A full treatment, including proofs of all claims, is in appendix A. The two expansions are as follows.

- *Expansion 1.* For sufficiently large R , we have

$$\lambda R = e^{A_2/4G} - e^{A_1/4G} + \frac{(1-p)R}{e^{A_2/4G} - \frac{(1-p)}{e^{A_1/4G+S}}R} + \mathcal{O}\left(\frac{e^{A_1/4G}e^{(A_1+A_2)/4G}}{pR}\right). \quad (4.31)$$

¹⁵For simplicity, we also assume $p, 1-p = \mathcal{O}(1)$, so that the corrections can be leading order. We remove this assumption in our more careful treatment in appendix A.

- *Expansion 2.* For sufficiently small R , we have

$$\lambda R = e^{A_2/4G} + \frac{pR}{e^{A_2/4G} - \frac{p}{e^{A_1/4G}}R} + \frac{(1-p)R}{e^{A_2/4G}} + \mathcal{O}\left(\frac{(1-p)R}{e^{A_2/4G}} \frac{(1-p)R}{e^{(A_1+A_2)/4G+S}}\right). \quad (4.32)$$

The condition (4.26) ensures there is overlap in the conditions where the two expansions are valid, implying that some expansion is valid for all values of R and λ . Each expansion gives a quadratic equation for the resolvent $R(\lambda)$, which can be easily solved and has a single branch cut, where the eigenvalue density $D_B(\lambda)$ is nonzero. Both branch cuts are within the respective regimes of validity of the corresponding expansion, and so we find two distinct sets of eigenvalues. The eigenvalues in Expansion 1 come from the $\rho_{b',2}$, while the eigenvalues in Expansion 2 come from the $|\psi\rangle\langle\psi|$ part of the state.

The entropy is given by

$$S(B) = - \int d\lambda \lambda \ln(\lambda) D_B(\lambda) = \frac{1}{\pi} \int d\lambda \lambda \ln(\lambda) \text{Im}R(\lambda + i\varepsilon), \quad (4.33)$$

where we include both sets of eigenvalues in the integral. The answer depends on how $H_{\min}^\varepsilon(b')$ and $H_{\max}^\varepsilon(b')$ from (4.27) compare to $\Delta A \equiv A_2 - A_1$.

Regime 1: $H_{\min}^\varepsilon, H_{\max}^\varepsilon \ll \Delta A/4G$. In this regime, the naïve QES prescription gives the right answer. Expansion 1 has a peak of eigenvalues at $\lambda \approx (1-p)e^{-A_1/4G-S}$, and Expansion 2 has a peak of eigenvalues at $\lambda \approx pe^{-A_1/4G}$. Both are within the regime of validity of their expansion. See the top plot of figure 5. Combined these peaks give entropy

$$S(B) = \frac{A_1}{4G} + (1-p)S + \dots, \quad (4.34)$$

where “...” represents terms subleading at large S , A_1 , and A_2 . This includes the Shannon entropy term $-p \ln(p) - (1-p) \ln(1-p)$. Recall that equation (4.34) is the naïve QES answer because $S(b') \lesssim H_{\max}^\varepsilon(b') \ll \Delta A/4G$.

Regime 2: $H_{\min}^\varepsilon \ll \Delta A/4G \ll H_{\max}^\varepsilon$. Here there are large corrections to the naïve QES prescription. Expansion 2 describes the same peak it did in Regime 1, giving eigenvalues at $\lambda \approx pe^{-A_1/4G}$. Expansion 1 now describes eigenvalues that have crossed the phase transition, which are therefore at $\lambda \approx (1-p)e^{-A_2/4G}$. Both peaks are still well-separated, and the expansions continue to be valid at the peaks. See the middle plot of figure 5. The entropy comes out to

$$S(B) = p\frac{A_1}{4G} + (1-p)\frac{A_2}{4G} + \dots, \quad (4.35)$$

and again we have dropped subleading terms, including the Shannon term.

Note that this entropy is different from the naïve QES answer. While the naïve answer only cares about the relative sizes of $S(b')$ and $\Delta A/4G$, this answer is independent of those relative sizes! Indeed, by dialing S , we can place $S(b')$ on either side of $\Delta A/4G$, as we please:

$$H_{\min}^\varepsilon, S(b') \ll \frac{\Delta A}{4G} \ll H_{\max}^\varepsilon, \quad \text{or} \quad H_{\min}^\varepsilon \ll \frac{\Delta A}{4G} \ll S(b'), H_{\max}^\varepsilon \quad (4.36)$$

The entropy $S(B)$ equals (4.35) in both cases, while the naïve QES prescription gives totally different formulas for the two cases! In both cases, the naïve QES prescription gives an answer that is larger (at leading order) than the correct answer.

The naïve QES prescription failed because it treated the bulk eigenvalues in an all-or-nothing way, stubbornly refusing to acknowledge that some of the eigenvalues are much larger than the phase transition value $e^{-A_2/4G}$, even though many others are small enough to have crossed the phase transition.

Regime 3: $\Delta A/4G \ll H_{\min}^\epsilon, H_{\max}^\epsilon$. The naïve QES prescription is back to receiving no corrections. In this regime, Expansion 1 is never valid, while Expansion 2 describes a peak of eigenvalues that have crossed the phase transition, sitting at $\lambda \approx e^{-A_2/4G}$. See the bottom plot of figure 5. We obtain an entropy

$$S(B) = \frac{A_2}{4G} + \dots, \tag{4.37}$$

again letting “...” represent subleading terms. There is no Shannon term in this regime. This agrees with the answer from the naïve QES prescription because $S(b') \gtrsim H_{\min}^\epsilon(b') \gg \Delta A/4G$.

Higher Renyis. So far we have computed the von Neumann entropy in each regime, finding large corrections to the naïve holographic prescription in Regime 2. What about the higher Renyi entropies? There is a holographic way to compute them as well [32]; does that also have large corrections in some regime?

The answer is that their corrections are generally much smaller, even in Regime 2. This is straightforward to derive with the resolvent approximations we have given. For integer Renyi entropies with $n > 1$, this is fairly self-explanatory. These can be computed directly using n replicas without the need for any analytic continuation, and so can always be computed in the semiclassical limit using a saddle point approximation.

More interestingly, the corrections are also nonperturbatively small for non-integer Renyi entropies with $n > 1$ (and $n < 1$), so long as $(n - 1)$ is finite in the semiclassical limit. The large corrections to the von Neumann entropy come from the limit $n \rightarrow 1$ not commuting with the semiclassical limit $G \rightarrow 0$, unless we keep track of nonperturbatively small corrections.

Example 2: entangled states.

Setup. This next example demonstrates the role of *conditional* min-/max-entropy. It closely resembles the previous one, but now the dustball or black hole b' is entangled with another dustball or black hole that is always in the entanglement wedge. For evaporating black holes (or their JT + EOW brane cousins), it calculates the entropy of the Hawking radiation rather than the black hole. The setups are detailed in section 2 and the dustball version is depicted in figure 3. Again those setups are simplified by fixing the areas of γ_1 and γ_2 to A_1 and A_2 .

We emphasize again that this setup is, quite literally, the complement of the first one. In that example, while we imagined a CFT $B\bar{B}$ in a mixed state, we could have instead

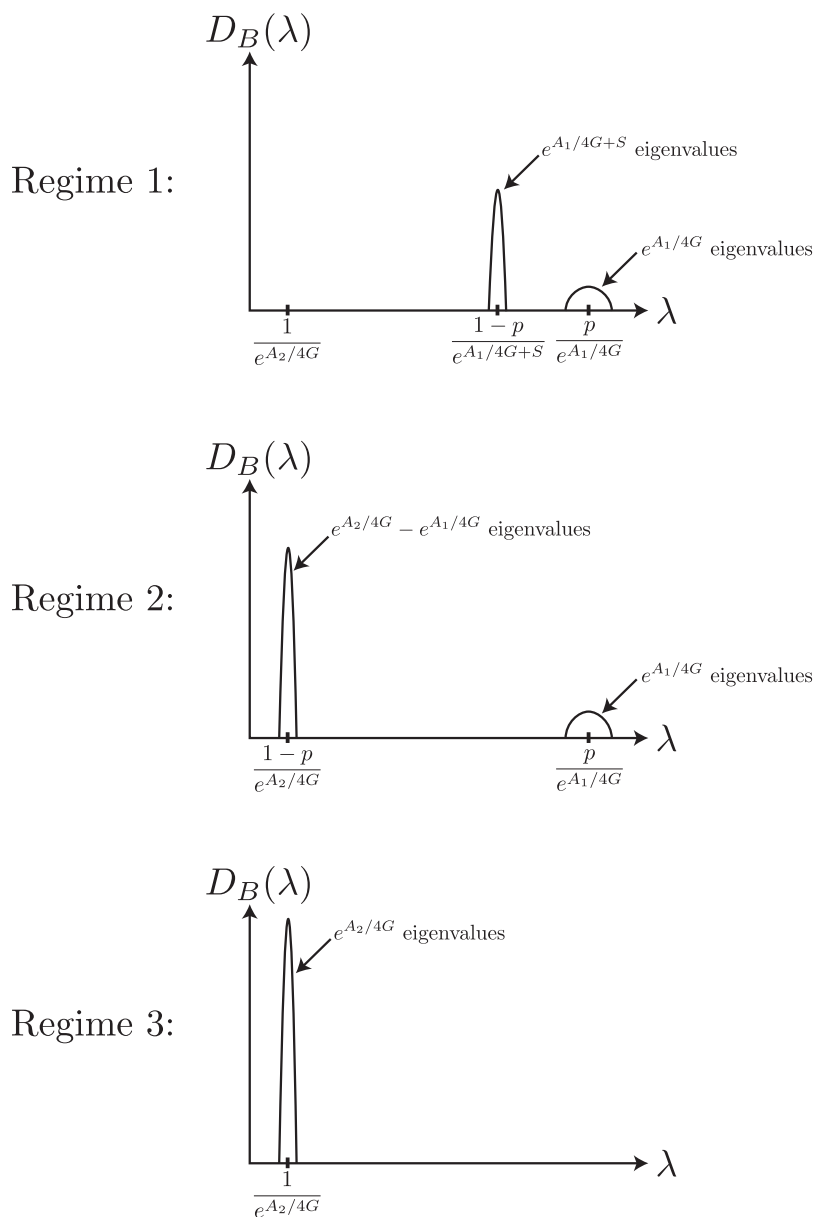


Figure 5. Eigenvalue density for the three Regimes in section 4.3. In Regime 1, there are two peaks of eigenvalues, each associated to one of the two states in the mixture, and each much greater than the critical value $1/e^{A_2/4G}$. Hence the naïve QES prescription is correct. In Regime 2, one of the peaks has shifted to the critical value, while the other remained where it was, leading to large corrections in the naïve QES prescription. In Regime 3, both peaks have moved to the critical value, and the naïve prescription is valid again. Note the agreement with numerical results for the analogous random tensor network in appendix C.

imagined it purified by some reference system R . Introducing R changes nothing about that calculation. Nonetheless, it is useful because $S(\overline{BR}) = S(B)$ regardless of the makeup of R . Here we imagine R to be an identical copy of $B\overline{B}$, with the same size dustball or black hole in its bulk dual r . For notational simplicity, we shall combine R into \overline{B} , so that we are just computing $S(\overline{B})$. Similarly in the bulk we combine r into \overline{b} .

We consider the following two states of the dustballs or black holes. One is a pure, factorized state $|\psi\rangle_{b'} \otimes |\psi\rangle_{\overline{b}}$. The other is a pure, maximally entangled state, $|\Phi\rangle_{b'\overline{b}}$, with entanglement entropy S across the partition. We will compute the entropy of their superposition,

$$\sqrt{p} |\psi\rangle_{b'} \otimes |\psi\rangle_{\overline{b}} + \sqrt{1-p} |\Phi\rangle_{b'\overline{b}}. \tag{4.38}$$

Again, we keep this example simple by assuming (4.26). The conditional von Neumann, min-, and max-entropies of this state are

$$\begin{aligned} H_{\min}^\varepsilon(b'|\overline{b}) &\approx -S, \\ S(b'|\overline{b}) &= -(1-p)S, \\ H_{\max}^\varepsilon(b'|\overline{b}) &\approx \ln(p). \end{aligned} \tag{4.39}$$

All of these are negative numbers, because of the entanglement. The entropy $S(\overline{B})$ depends on their comparison to $(A_1 - A_2)/4G$, which is itself big and negative. The naïve QES prescription says

$$S_{\text{naïve}}(\overline{B}) = \begin{cases} \frac{A_2}{4G} + S(b'\overline{b}), & S(b'|\overline{b}) \ll \frac{A_1 - A_2}{4G} \\ \frac{A_1}{4G}, & S(b'|\overline{b}) \gg \frac{A_1 - A_2}{4G}. \end{cases} \tag{4.40}$$

The correct answer up to $\mathcal{O}(1)$, we will see, is

$$S_{\text{refined}}(\overline{B}) = \begin{cases} \frac{A_2}{4G}, & H_{\max}^\varepsilon(b'|\overline{b}) \ll \frac{A_1 - A_2}{4G} \\ p \frac{A_1}{4G} + (1-p) \frac{A_2}{4G}, & H_{\min}^\varepsilon(b'|\overline{b}) \ll \frac{A_1 - A_2}{4G} \ll H_{\max}^\varepsilon(b'|\overline{b}) \\ \frac{A_1}{4G} + S(b'), & H_{\min}^\varepsilon(b'|\overline{b}) \gg \frac{A_1 - A_2}{4G}. \end{cases} \tag{4.41}$$

Calculation. Rather than write out a resolvent like we did before, we will use a trick to compute the entropy in each of these three regimes. Notice that these smooth conditional min- and max-entropies (4.39) equal minus the max- and min-entropies (4.27) respectively, from Example 1. This was the general rule, from section 3: for a pure state on ABC ,

$$H_{\min}(A|B) = -H_{\max}(A|C). \tag{4.42}$$

Since the system b that we conditioned on in Example 1 was trivial, we have

$$H_{\min}^\varepsilon(b'|\overline{b}) = -H_{\max}^\varepsilon(b'), \tag{4.43}$$

$$H_{\max}^\varepsilon(b'|\overline{b}) = -H_{\min}^\varepsilon(b'). \tag{4.44}$$

So we can compute the entropy in the three regimes as follows. Consider, for example, the regime in which both the conditional min- and max-entropy are less than $\Delta A/4G \equiv$

$(A_1 - A_2)/4G$. This corresponds exactly to the regime in Example 1 where both min- and max-entropy were *greater* than $(A_2 - A_1)/4G$. So, using purity of $B\bar{B}$, the entropy $S(\bar{B})$ in this regime equals $S(B)$ from that regime. Thus the entropy $S(\bar{B})$ is completely deducible from the results of Example 1.

The key lesson is this: there is an important role played by bulk entanglement, encapsulated by the *conditional* min- and max-entropy. That's the only way this setup would be consistent with the complementary answers from the previous example.

Regime 1: $H_{\min}^\varepsilon(b'|\bar{b}), H_{\max}^\varepsilon(b'|\bar{b}) \ll \Delta A/4G$. This is Regime 3 of Example 1. Therefore,

$$S(\bar{B}) = \frac{A_2}{4G} + \dots, \tag{4.45}$$

where “...” represents subleading terms. The naïve QES prescription gives the same answer because $S(b'|b) \lesssim H_{\max}^\varepsilon(b'|\bar{b}) \ll \Delta A/4G$.

Regime 2: $H_{\min}^\varepsilon(b'|\bar{b}) \ll \Delta A/4G \ll H_{\max}^\varepsilon(b'|\bar{b})$. This is Regime 2 of Example 1, so again here there are large corrections to the naïve QES prescription. The entropy comes out to

$$S(\bar{B}) = p \frac{A_1}{4G} + (1-p) \frac{A_2}{4G} + \dots, \tag{4.46}$$

and again we have dropped subleading terms, including the $\mathcal{O}(1)$ Shannon term. This entropy is different than the naïve QES answer.

Regime 3: $\Delta A/4G \ll H_{\min}^\varepsilon(b'|\bar{b}), H_{\max}^\varepsilon(b'|\bar{b})$. This is Regime 1 of Example 1, and hence the naïve QES prescription is back to receiving no corrections. The entropy is

$$S(\bar{B}) = \frac{A_1}{4G} + (1-p)S + \dots, \tag{4.47}$$

where again “...” includes the Shannon term. This matches the naïve QES answer because $S(b'|b) \gtrsim H_{\min}^\varepsilon(b'|\bar{b}) \gg \Delta A/4G$.

Example 3: arbitrary entanglement spectra. What about more general bulk states $\rho_{b'}$, which aren't simply the mixture of two states with (approximately) flat entanglement spectra (again forgetting about entanglement, for now)? For an arbitrary bulk state $\rho_{b'}$ with eigenvalue density $D_{b'}(\lambda_{b'})$, the resolvent recursion relation (4.24) becomes

$$\lambda R = e^{A_2/4G} + \int d\lambda_{b'} \frac{\lambda_{b'} D_{b'}(\lambda_{b'}) R}{e^{A_2/4G} - \lambda_{b'} e^{-A_1/4G} R}. \tag{4.48}$$

We will not be able to solve this equation as precisely as we were able to calculate the resolvents in the preceding examples, but we will still have sufficient control to calculate the von Neumann entropy up to $\mathcal{O}(1)$ corrections.

Calculation. Our strategy will closely mirror the strategy used to calculate corrections to the von Neumann entropy near the Page transition in [9], and we refer the reader to that paper (and in particular appendix F) for more detailed justifications. We will perturbatively approximate the resolvent for $\lambda \gg e^{-A_2/4G}$, and argue that there are no eigenvalues with

$\lambda \ll e^{-A_2/4G}$ whenever the smooth max-entropy is sufficiently large. Combined these two results will enable us to calculate the von Neumann entropy up to $\mathcal{O}(1)$ corrections.

For $\lambda \gg e^{-A_2/4G}$, we treat the second term in (4.48) as a small perturbation. At leading order, the resolvent is given by $R_0(\lambda) = e^{A_2/4G}/\lambda$. The leading contribution to the density of states comes from the first perturbative correction

$$R_1(\lambda) = e^{A_1/4G} \int d\lambda_{b'} \frac{\lambda_{b'} D_{b'}(\lambda_{b'})}{\lambda (e^{A_1/4G} \lambda - \lambda_{b'})}. \quad (4.49)$$

To justify this perturbative approximation, we assume λ has a small imaginary part $i\epsilon$, with $\lambda \gg \epsilon \gg e^{-A_2/4G}$. Hence

$$|R_1(\lambda)| \leq \int d\lambda_{b'} \frac{\lambda_{b'} D_{b'}(\lambda_{b'})}{\epsilon \lambda} = \frac{1}{\epsilon \lambda} \ll R_0(\lambda), \quad (4.50)$$

as desired.

Suppose we are in Regime 1, where the smooth max-entropy $H_{\max}^\epsilon(b') \ll (A_2 - A_1)/4G$. Then we can approximate our bulk state by a nearby state with $D_{b'}(\lambda_{b'}) = 0$, except when $\lambda_{b'} \gg e^{-A_2/4G}$.¹⁶ For $\lambda \lesssim e^{-A_2/4G}$, we can ignore the first term in the denominator of (4.48) to get the self-consistent approximation

$$R \approx \frac{1}{\lambda} \left[e^{A_2/4G} - \int d\lambda_{b'} D_{b'}(\lambda_{b'}) \right]. \quad (4.51)$$

Note that going to higher orders in perturbation theory will not introduce a nonzero eigenvalue density, because there are no poles in (4.48) for these values of λ . We conclude that in Regime 1 we have $D(\lambda) = e^{A_1/4G} D_{b'}(e^{A_1/4G} \lambda)$, and hence $S(B) = \frac{A_1}{4G} + S(b')$.

What about when $H_{\max}^\epsilon(b') \gg (A_2 - A_1)/4G$? We want to argue that there are no eigenvalues with $\lambda \ll \epsilon e^{-A_2/4G}$, and hence that $R(\lambda)$ is negative and real. To do so, we rewrite (4.48) to give λ as a function of R

$$\lambda = \frac{e^{A_2/4G}}{R} + \int d\lambda_{b'} \frac{\lambda_{b'} D_{b'}(\lambda_{b'})}{e^{A_2/4G} - \lambda_{b'} e^{-A_1/4G} R}. \quad (4.52)$$

For small negative R , λ is large and negative, since the first term dominates. When R is very large and negative however, the second term dominates (thanks to our assumptions about the smooth max-entropy), and so λ is positive. There will be some intermediate R where λ is maximal, which gives the bottom of the entanglement spectrum.

To lower bound this maximum, we choose some $R \gg e^{A_2/2G}$. Then the second term in (4.52) dominates and we find

$$\lambda \gtrsim e^{-A_2/4G} \int_{\lambda_{b'} \ll e^{(A_1+A_2)/4G}/R} d\lambda_{b'} \lambda_{b'} D_{b'}(\lambda_{b'}) \gtrsim \mathcal{O}(\epsilon e^{-A_2/4G}). \quad (4.53)$$

The last approximation again follows from our assumption about the size of the smooth max-entropy. We therefore conclude that there are no eigenvalues with $\lambda \ll e^{-A_2/4G}$, as expected.

¹⁶The effect of this $\mathcal{O}(\epsilon)$ approximation to the state on the von Neumann entropy is controlled by Fannes inequality [33]. See (5.14).

We can now calculate the entropy $S(B)$. Since we know the eigenvalue density for both $\lambda \gg e^{-A_2/4G}$ and $\lambda \ll \varepsilon e^{-A_2/4G}$, we know the remaining eigenvalues must all have $\varepsilon e^{-A_2/4G} \lesssim \lambda \lesssim e^{-A_2/4G}$. Up to $\mathcal{O}(\ln \varepsilon)$ corrections, this means that

$$S(B) = \frac{A_1}{4G} + \int d\lambda_{b'} \lambda_{b'} D_{b'}(\lambda_{b'}) \min \left(\ln \lambda_{b'}, \frac{A_2 - A_1}{4G} \right). \quad (4.54)$$

When $H_{\min}^\varepsilon(b') \gg (A_2 - A_1)/4G$, we can ignore the first term in the minimization and we recover the naïve QES prescription result $S(B) = A_2/4G$. However, when $H_{\min}^\varepsilon(b') \ll (A_2 - A_1)/4G$, we find leading order corrections. Our results agree with the refined QES prescription in all three regimes.

Summary Let us summarize what we learned in this section. Doing a careful calculation — without the LM assumption — reveals a refinement of the naïve QES formula, which can differ from the naïve one at leading order. This refined QES prescription compares the smooth conditional min- and max-entropies to the difference in areas. We have seen this in three examples, all of which were fixed-area states, and all of which had particular simple bulk states. Unfortunately, states where there is a large amount of entropy in all the bulk regions and an arbitrary entanglement structure, and states where the areas are not fixed, are beyond the current technology we have for computing the replica trick without using the LM assumption. However, in the next section, we will derive the QES refinement more generally, beyond these particular bulk states and beyond fixed-area states, by using a more indirect approach.

5 Corrections in general holographic states

We start by arguing that the naïve QES prescription *is* valid, whenever the smooth conditional min- and max-entropy are safely on the same side of $(A_2 - A_1)/4G$. This generalizes half the pattern from our examples, now showing that for *any* state on $bb\bar{b}$ the naïve QES prescription can be trusted when the min- and max-entropy are on the same side of the area difference, though we emphasize that we still limit ourselves to two competing fixed-area QES, as in figure 1.

Then in section 5.2 we prove that there *are* generally large corrections to the naïve QES prescription in the regime where we did not prove the corrections are small. I.e. there are large corrections when the min- and max-entropy are on different sides of the area difference.

Finally, in section 5.3 we remove the fixed-area requirement, demonstrating that more general geometries follow the same pattern, up to a relatively small difference, $\mathcal{O}(\ln G)$.

Altogether, our argument shows that there are large corrections in general holographic states if and only if the bulk min- and max-entropy straddle the difference in areas between the two competing QES.

5.1 The regime of validity of naïve QES in fixed-area states

We first argue that the naïve QES prescription is valid, up to $o(1)$ corrections, for general fixed-area states with two extremal surfaces, so long as either

$$\frac{A_2}{4G} \ll \frac{A_1}{4G} + H_{\min}^\varepsilon(b'|b), \tag{5.1}$$

in which case the minimal QES is the surface γ_2 , or

$$\frac{A_2}{4G} \gg \frac{A_1}{4G} + H_{\max}^\varepsilon(b'|b), \tag{5.2}$$

in which case the minimal QES is the surface γ_1 . By “much greater than”, \gg , we mean a difference that is much larger than $\mathcal{O}(\ln G)$. Therefore, large corrections can only exist if

$$\frac{A_1}{4G} + H_{\max}^\varepsilon(b'|b) \gtrsim \frac{A_2}{4G} \gtrsim \frac{A_1}{4G} + H_{\min}^\varepsilon(b'|b). \tag{5.3}$$

We will later argue in section 5.2 that significant corrections (at least $\mathcal{O}(1)$ in size) in fact *always* exist when

$$\frac{A_1}{4G} + H_{\max}^\varepsilon(b'|b) \gg \frac{A_2}{4G} \gg \frac{A_1}{4G} + H_{\min}^\varepsilon(b'|b). \tag{5.4}$$

Our strategy will be to make use of the correspondence between the nonperturbative corrections to the replica trick entropy in a) fixed-area states in gravity and b) single-tensor random tensor networks (RTNs) [34], a nonperturbative equivalence first noted in [9].

Let us start by reviewing that correspondence. We have already evaluated $\text{tr}(\rho^n)$ for fixed-area states in section 4. We found that for a general normalized bulk state $\rho_{bb\bar{b}}$, the dual normalized boundary state $\rho_{B\bar{B}}$ satisfied

$$\text{tr}(\rho_B^n) = \sum_{\pi} \text{tr}(\tau_b \pi_{b'} \rho_{bb'\bar{b}}^{\otimes n}) \exp\left(\left[C(\tau^{-1} \circ \pi) - n\right] \frac{A_2}{4G} + \left[C(\pi) - n\right] \frac{A_1}{4G}\right), \tag{5.5}$$

where the fixed permutation τ is cyclic, the sum is over permutations π that maximize $C(\tau^{-1} \circ \pi) + C(\pi)$, and the operators $\tau_b, \pi_{b'}$ permute the n copies of their respective subsystem.

We want to show that one finds the same formula in RTNs, where

$$\rho_{B\bar{B}} = V_B V_{\bar{B}} V \rho_{bb'\bar{b}} V^\dagger V_B^\dagger V_{\bar{B}}^\dagger \tag{5.6}$$

and $V : b' \rightarrow B' \otimes \bar{B}'$, $V_B : b \otimes B' \rightarrow B$ and $V_{\bar{B}} : \bar{b} \otimes \bar{B}' \rightarrow \bar{B}$ are random isometries. This is shown graphically in figure 6. Here the subsystems B' and \bar{B}' have dimensions $d_{B'} = \exp(A_2/4G)$ and $d_{\bar{B}'} = \exp(A_1/4G)$ respectively.

We can write [34]

$$\text{tr}(\rho_B^n) = \text{tr}(\tau_b \tau_{B'} (U V_0 \rho_{bb'} V_0^\dagger U^\dagger)^{\otimes n}). \tag{5.7}$$

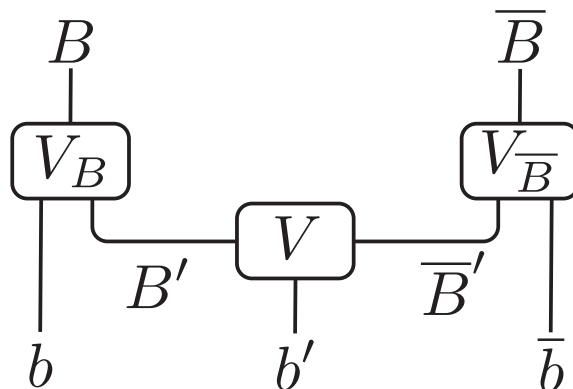


Figure 6. Simple random tensor network analogous to figure 1 with fixed areas. The “bulk legs” b, b', \bar{b} are related by isometries $V, V_B, V_{\bar{B}}$ to “boundary legs” B, \bar{B} . The (log of the) dimensions of the legs B' and \bar{B}' play the role of the areas in figure 1.

Here we have written $V = UV_0$ for a fixed isometry V_0 and a Haar random unitary U . Now, we can use the formula [35]

$$\int dU U_{i_1 j_1} \dots U_{i_n j_n} U_{j'_1 i'_1}^\dagger \dots U_{j'_n i'_n}^\dagger = d^{-n} \sum_{\pi} \delta_{i_1 i'_{\pi(1)}} \delta_{j_1 j'_{\pi(1)}} \dots \delta_{i_n i'_{\pi(n)}} \delta_{j_n j'_{\pi(n)}} + \mathcal{O}(d^{-n-1}), \quad (5.8)$$

where $\pi(i) \in \{1, \dots, n\}$ represents the arbitrary permutation π acting on the i -th element of an n -element set. Combine with $d = d_{B'} d_{\bar{B}'}$ to obtain

$$\text{tr}(\rho_B^n) = d^{-n} \sum_{\pi} \text{tr}(\tau_b \pi_{b'} \rho_{bb'}) \text{tr}(\tau_{B'} \pi_{B'}) \text{tr}(\pi_{\bar{B}'}) = \sum_{\pi} \text{tr}(\tau_b \pi_{b'} \rho_{bb'}) d_{B'}^{C(\tau^{-1} \circ \pi) - n} d_{\bar{B}'}^{C(\pi) - n}. \quad (5.9)$$

This is exactly the result that we found in the gravity calculation.

Since we can (in principle if not in practice) calculate the entropy $S(B)$ simply by analytically continuing $\text{tr}(\rho_B^n)$, the RTN must have the same entanglement entropy as the gravity calculation. Armed with this knowledge, we can calculate the entanglement entropy in the RTN, *using any technique we want*, and thereby find the gravitational answer as well.

The key result that we will use is the one-shot decoupling theorem, Theorem III.1 of [36] (see appendix D for our summary of the proof), which says that for $V_0 \rho_{bb'} V_0^\dagger$, then so long as

$$\ln d_{B'} < \ln d_{\bar{B}'} + H_{\min}(b'|b)_\rho - 2 \ln \frac{1}{\varepsilon}, \quad (5.10)$$

we have

$$\int dU \left\| \text{tr}_{\bar{B}'}(UV_0 \rho_{bb'} V_0^\dagger U^\dagger) - \rho_b \otimes \frac{\mathbb{1}_{B'}}{d_{B'}} \right\|_1 \leq \varepsilon. \quad (5.11)$$

What does this theorem mean? It states that (5.10) is a sufficient condition to ensure

$$\rho_B \approx \tilde{\rho}_B = V_B \rho_b \otimes \mathbb{1}_{B'} / d_{B'} V_B^\dagger. \quad (5.12)$$

Moreover, the condition (5.10) can be weakened, replacing the min-entropy $H_{\min}(b'|b)_\rho$ by its smooth version $H_{\min}^\varepsilon(b'|b)_\rho$, with only a small degradation in the quality of the approximation, as follows from the definition of smoothing.

The state on the right hand side of (5.12) has two essential features. The first is that it depends only on the reduced state ρ_b , and is completely independent of b' . The second is that its entropy

$$S(\tilde{\rho}_B) = \ln d_{B'} + S(\rho_b) \tag{5.13}$$

corresponds in gravity to the generalized entropy of the surface γ_2 .

We want to use this to bound the entropy of the state ρ_B itself. To do so, we need the Fannes' inequality [33]

$$|S(\rho) - S(\sigma)| \leq \frac{1}{2} \|\rho - \sigma\|_1 \ln d + S_2\left(\frac{1}{2} \|\rho - \sigma\|_1\right). \tag{5.14}$$

Here $S_2(p) = -p \ln p - (1-p) \ln(1-p)$ is the Shannon entropy of the probability distribution $(p, 1-p)$.

Applying this inequality to the states in (5.11), we find that

$$S(\rho_B) = S(\tilde{\rho}_B) + \mathcal{O}(\varepsilon \ln d'_B) + \mathcal{O}(\varepsilon \ln \varepsilon) = \frac{A_2}{4G} + S(\rho_b) + \mathcal{O}\left(\frac{\varepsilon}{4G}\right) + \mathcal{O}(\varepsilon \ln \varepsilon). \tag{5.15}$$

If we take ε (in both (5.11) and the smooth min-entropy) to be polynomially small in G (say $\mathcal{O}(G^2)$), then (5.10) is satisfied whenever

$$\frac{A_1}{4G} + H_{\min}^\varepsilon(b'|b) - \frac{A_2}{4G} \gg 0 \tag{5.16}$$

and this difference grows faster than $\ln(1/G)$ in the semiclassical limit $G \rightarrow 0$. Moreover, (5.15) says $S(\rho_B)$ is given by the generalized entropy of the quantum extremal surface γ_2 , up to a perturbatively small ($\mathcal{O}(G)$) correction.

We also want to show the QES prescription is valid, this time with minimal QES γ_1 , so long as

$$\frac{A_1}{4G} + H_{\max}^\varepsilon(b'|b) \ll \frac{A_2}{4G}. \tag{5.17}$$

It turns out that this follows by the same arguments used above, applied to the complementary boundary region.

We first consider some arbitrary purification $|\psi\rangle_{bb'\bar{b}R}$ of the bulk state $\rho_{bb'\bar{b}}$. We now want to calculate the entanglement entropy $S(\bar{B}R)$ of the corresponding pure boundary state. This is, of course, equal to the entropy $S(\rho_B)$ that we are really interested in.

As can be seen immediately from the tensor network picture, this is exactly the same situation that we considered before, except that B has been replaced by $\bar{B} \otimes R$, b has been replaced by $\bar{b} \otimes R$, and the areas A_1 and A_2 have been exchanged. It follows that

$$S(B) = S(\bar{B}R) \approx \frac{A_1}{4G} + S(\bar{b}R)_\psi = \frac{A_1}{4G} + S(bb')_\rho, \tag{5.18}$$

so long as

$$\frac{A_2}{4G} + H_{\min}^\varepsilon(b'|\bar{b}R)_\psi - \frac{A_1}{4G} \gg 0. \tag{5.19}$$

Since $|\psi\rangle_{bb'\bar{b}R}$ is pure, $H_{\min}^\varepsilon(b'|\bar{b}R)_\psi = -H_{\max}^\varepsilon(b'|b)_\rho$ and so this is exactly (5.17).

It is worth briefly commenting on whether an equivalent formula to (5.11) could be directly shown in gravity. The proof of (5.11) is reviewed in appendix D and involves calculating the Hilbert-Schmidt distance between the entangled and product states, and then using the Hilbert-Schmidt norm to bound the trace-norm. Since the Hilbert-Schmidt norm can be computed directly using a path integral (without analytic continuation), it should in principle be possible to evaluate in gravity. Just like for Rényi entropy calculations, for fixed-area states the gravity answer should agree with the random tensor network answer. However, in order to derive a gravitational decoupling theorem, one would still need to use some quantum information tricks (basically a clever application of the Cauchy-Schwarz inequality) in order to eventually bound the trace-norm. The derivation would therefore still not be a completely direct gravity calculation.

5.2 The regime where naïve QES fails

So far we have only argued that there *aren't* significant correction to the QES prescription, so long as we *don't* have

$$\frac{A_1}{4G} + H_{\max}^\varepsilon(b'|b) \gtrsim \frac{A_2}{4G} \gtrsim \frac{A_1}{4G} + H_{\min}^\varepsilon(b'|b). \tag{5.20}$$

In this section, we argue that there *do* exist significant corrections (at least $\mathcal{O}(1)$ in size) when

$$\frac{A_1}{4G} + H_{\max}^\varepsilon(b'|b) \gg \frac{A_2}{4G} \gg \frac{A_1}{4G} + H_{\min}^\varepsilon(b'|b). \tag{5.21}$$

Here, ε can be relatively small, but should be parametrically $\mathcal{O}(1)$ in the semiclassical limit. Note that we do not have a general argument that these corrections need to be *leading order*, although we strongly expect that this is the case so long as (5.21) holds at leading order, as we found in the simple examples in section 4. Finding a proof that this is true in full generality is an important task for future work.

Our main tool will be the converse one-shot decoupling theorem of [37]. Applied to the random tensor network in figure 6, this says that, if (5.21) holds, then

$$\|\rho_{\bar{B}'\bar{b}R} - \rho_{\bar{B}'} \otimes \rho_{\bar{b}R}\|_1 = \mathcal{O}(1) \tag{5.22}$$

and

$$\|\rho_{B'b} - \rho_{B'} \otimes \rho_b\|_1 = \mathcal{O}(1). \tag{5.23}$$

Using Pinsker's inequality, these lower bound the relative entropies

$$S(\rho_{B'b} || \rho_{B'} \otimes \rho_b) = I(B' : b) \tag{5.24}$$

and

$$S(\rho_{\bar{B}'\bar{b}R} || \rho_{\bar{B}'} \otimes \rho_{\bar{b}R}) = I(\bar{B}' : \bar{b}R). \quad (5.25)$$

There is therefore at least an $\mathcal{O}(1)$ amount of mutual information both between B' and b and between \bar{B}' and $\bar{b} \otimes R$. Since there is no upper bound on the relative entropy from the trace distance, the mutual information can, of course, be parametrically larger than $\mathcal{O}(1)$, which we expect to happen when the inequalities holds at leading order.

The naïve QES prescription says that

$$S(B) = S(B'b) = S(\bar{B}R) = S(\bar{B}'\bar{b}R) = \min \left(\ln d_{B'} + S(b), \ln d_{\bar{B}'} + S(bb') \right). \quad (5.26)$$

However

$$\ln d_{B'} + S(b) > S(B') + S(b) = S(B'b) + I(B' : b), \quad (5.27)$$

and

$$\ln d_{\bar{B}'} + S(bb') > S(B') + S(\bar{b}R) \geq S(B'b) + I(\bar{B}' : \bar{b}R). \quad (5.28)$$

Since there is $\mathcal{O}(1)$ mutual information in each case, we find that the naïve QES prescription receives at least $\mathcal{O}(1)$ corrections. Since the same replica trick calculation gives the entropy of both the random tensor network and the corresponding fixed-area state, the same is true for fixed-area states.

5.3 From fixed-area states to general holographic states

Of course, most of the time we are not interested in fixed-area states. Instead the states of interest (vacuum-AdS, perturbative excitations above the vacuum, thermofield double states etc.) generally have small ($\mathcal{O}(\sqrt{G})$) fluctuations in the area of the extremal surface(s).

In this section, we argue that, up to small $\mathcal{O}(\ln G)$ corrections, the entropies of such states can be calculated by expanding the state as a superposition of fixed-area states, and then taking an expectation of the entropies of the states in the superposition.¹⁷

From our point of view, the primary importance of this result is that, at leading order, the entanglement entropy of a generic holographic state is the same as the entropy of a fixed-area state with the same classical area. Therefore general holographic states inherit the leading order corrections we found for fixed-area states. It also shows that the corrections to the naïve QES prescription are small, for general holographic states, so long as

$$\frac{A_2 - A_1}{4G} - H_{\max}^\varepsilon(b'|b) \gg \sqrt{\frac{\ln G}{G}}, \quad (5.29)$$

and similarly for $H_{\min}^\varepsilon(b|b)$.

¹⁷For simplicity, in this section we only consider pure states. If the state of interest is mixed, one can simply first purify it using a reference system and then replace all references to the complementary region \bar{B} in the argument below by $\bar{B} \otimes R$.

That said, our argument also has other technical applications, for example bounding the error in the assumptions used in [14] to calculate the $\mathcal{O}(1/\sqrt{G})$ corrections to the entanglement entropy near a QES phase transition.

Our starting point is that the general holographic state $|\psi\rangle$ can be written as a superposition over fixed-area states $|A_1, A_2\rangle$ as

$$|\psi\rangle = \sum_{A_1, A_2} \sqrt{p(A_1, A_2)} |A_1, A_2\rangle. \quad (5.30)$$

The fluctuations in the area are Gaussian (in the semiclassical limit) with width $\mathcal{O}(\sqrt{G})$ (see [14] for detailed calculations), so we can approximate the state up to any polynomially small error (w.r.t. G) by a state with support only on an $\mathcal{O}(\sqrt{G \ln G})$ range of values.¹⁸

As with any continuously valued measurement operator, it is not well defined to measure the area exactly. Instead, the area operator should be viewed as a projection-valued measure (PVM), and the states $|A_1, A_2\rangle$ should be viewed as the outcome of measuring the area to some precision δ . We shall take δ to be polynomially small with respect to G .

It follows from the preceding two paragraphs that the number of distinct fixed-area states in the superposition scales as $\mathcal{O}(G \ln G / \delta^2)$. (Note that we are taking a superposition over states with both A_1 and A_2 fixed, which squares the number of terms in the superposition.) Crucially this means that the number of states is polynomial in $1/G$.

We are now almost ready to consider the reduced density matrix $\rho_B = \text{tr}_{\bar{B}} |\psi\rangle \langle \psi|$. However, as an intermediate step we first consider taking a superposition over only states with different values of A_1 , for some fixed A_2 . In other words, we have

$$\rho_B(A_2) = \sum_{A_1, A'_1} \sqrt{p(A_1|A_2)p(A'_1|A_2)} \text{tr}_{\bar{B}} \left(|A_1, A_2\rangle \langle A'_1, A_2| \right). \quad (5.31)$$

The first thing to observe is that the bulk operator \hat{A}_1 is always reconstructable on the boundary region \bar{B} . Hence the only terms that survive the partial trace have $A_1 = A'_1$.

We therefore find that $\rho_B(A_2)$ can be written as the incoherent mixture

$$\rho_B(A_2) = \sum_{A_1} p(A_1|A_2) \text{tr}_{\bar{B}} \left(|A_1, A_2\rangle \langle A_1, A_2| \right). \quad (5.32)$$

However, as discussed in section 2, we can bound the entropy of such a mixture from above and below by

$$\sum_{A_1} p(A_1|A_2) S(B)_{|A_1, A_2\rangle} \leq S(\rho_B(A_2)) \leq \sum_{A_1} p(A_1|A_2) S(B)_{|A_1, A_2\rangle} - p(A_1|A_2) \ln p(A_1|A_2). \quad (5.33)$$

¹⁸The error in neglecting the tail of a Gaussian outside a window Δx goes like $e^{-\mathcal{O}(\Delta x/\sigma)^2}$. So if Δx equals k standard deviations, the error goes like $e^{-\mathcal{O}(k^2)}$. Hence the \sqrt{G} is for the standard deviation, and the $\sqrt{\ln G}$ ensures we capture a greater number of standard deviations as $G \rightarrow 0$, such that the error tends to zero polynomially in G .

The difference between the upper and lower bounds is an $O(\ln G)$ entropy of mixing term (because there were $\mathcal{O}(\sqrt{G \ln G}/\delta)$ distinct states in the superposition) and hence can be ignored at leading order (and for calculating the $O(1/\sqrt{G})$ corrections discussed in [13, 14]).

Now we need to take a superposition over different values of A_2 . Because all the states involved are pure, $S(\rho_B) = S(\rho_{\bar{B}})$, and, for any A_2 , $S(\rho_B(A_2)) = S(\rho_{\bar{B}}(A_2))$. We can therefore compute the entropy of the reduced state on \bar{B} rather than B .

Since A_2 can always be reconstructed on B , this is again an incoherent mixture

$$\rho_{\bar{B}} = \sum_{A_2} p(A_2) \rho_{\bar{B}}(A_2). \tag{5.34}$$

Hence we have

$$\sum_{A_2} p(A_2) S(\rho_B(A_2)) \leq S(\rho_B) \leq \sum_{A_2} p(A_2) S(\rho_B(A_2)) - p(A_2) \ln p(A_2). \tag{5.35}$$

Again the difference between the lower and upper bounds is $O(\ln G)$ and so can be ignored in lower order calculations.

Altogether, we therefore find

$$S(\rho_B) = \sum_{A_1, A_2} p(A_1, A_2) S(B)_{|A_1, A_2} + O(\ln G), \tag{5.36}$$

which is exactly what we set out to show. In particular, the QES prescription is valid at leading order for general holographic states, whenever it is valid for the corresponding fixed-area states. Moreover, the QES prescription receives leading order corrections, whenever there are leading order corrections to the entropy of corresponding fixed-area states. When the difference in areas is smaller than the fluctuations in this difference, we also find the ($\mathcal{O}(\sqrt{1/G})$) corrections from [9, 13, 14].

There's one remaining remark to make. The fluctuations in the areas A_1 and A_2 are formally divergent when we take the radial cut-off to infinity. This leads to a natural question of whether we were justified in treating the potential entropy of mixing terms as smaller than $\mathcal{O}(1/G)$ but non-divergent corrections.

The short answer is that this subtlety does not matter for our purposes. The IR fluctuations create a *constant* (divergent) difference between the entropy in fixed-area states and entropy in general states, independent of which QES is dominant, independent of the bulk state. Hence the entropy $S(B)$ in a general bulk state can be computed as the expectation of the entropies of the fixed-area states in its superposition, as we already argued, *plus* a constant shift. This shift is currently underappreciated and deserves more study, but it does not affect our ability to infer general corrections from fixed-area states.

The longer answer is as follows. First note that the IR pieces of A_2 and A_1 are the same, so the fluctuations of A_1 , in states where the area of A_2 is fixed, do not diverge [14]. This implies the difference between the lower and upper bounds in (5.33) is genuinely finite. Moreover, this implies that $A_2 - A_1$ is independent of this IR subtlety, implying the condition for corrections (5.3) remains well-defined.

The important effect of this IR subtlety is in the entropy of mixing term in (5.35), and is indeed divergent. However, it represents a large constant shift — not a large window —

because the lower bound can be strengthened to include this divergence as well. This works as follows. Let there be some fixed radial cutoff ε , such that A_2 diverges in the $\varepsilon \rightarrow 0$ limit. Group fixed-area states into blocks corresponding to some $\mathcal{O}(1)$ range of areas. There are a polynomial in $1/G$ number, $\mathcal{O}(1/\delta)$, of fixed-area states in each block. This number grows as $G \rightarrow 0$. As the IR cutoff is taken away, the number of such blocks grows as $1/\varepsilon$ to some power.

We can separate the Shannon term associated to the mixing of these blocks from the Shannon term associated to the mixing of the $\mathcal{O}(1/\delta)$ states within each block. The first Shannon term does not depend on G , only on ε .

This IR Shannon term, crucially, can be included in the above lower bound of (5.35). The resulting inequality is true because the different blocks are distinguishable on both B and \bar{B} . Indeed, A_1 , a quantity known to \bar{B} , takes on vastly different values in the different blocks (because its IR value matches that of A_2).

Therefore, the entropy of mixing associated to the IR fluctuations of area can be understood as a constant shift to the entropy, present even if there is just a single fixed-area surface. This concludes the argument.

6 Entanglement wedge reconstruction

This refinement of the QES prescription brings with it a refinement of the condition for entanglement wedge reconstruction (EWR).

We show in section 6.1 the refined conditions are the following. A region B of the boundary will be able to reconstruct the state of a region b' of the bulk, as in figure 1, given a bulk state ρ , if and only if¹⁹

$$\frac{A_2}{4G} \gg \frac{A_1}{4G} + H_{\max}^\varepsilon(b'|b)_\rho. \tag{6.1}$$

This condition is similar to that from Hayden and Penington²⁰ [8] (see also [4]), but builds on it in a key way. The similarity is that both depend at some level on the comparison between $\Delta A/4G$ and $H_{\max}(b')_\rho$ (though [8] did not say it this way).

The key difference is that (6.1) tells you whether B can reconstruct *the particular state* ρ . The condition from [8] tells you whether there exists a single reconstruction procedure that works for *any state* in a code subspace that contains ρ .

This difference shows up in two places: the smoothing of H_{\max} , and the conditioning on b . The smoothing allows us to only care about the *approximate* dimension of $\rho_{b'}$, formalizing the intuitive notion that we can ignore small pieces of the wavefunction and still approximately reconstruct the state. The conditioning on b quantifies how entanglement in ρ helps B reconstruct b' , formalizing the intuition that bulk entanglement between b and b' can aid reconstruction.

In section 6.2, we explain that this new, state-specific formulation of EWR (6.1) is equivalent to a well-known quantum information task, one-shot quantum state merging.

¹⁹We discuss setups with more than two candidate QES in section 7.

²⁰See also Dong, Harlow, and Wall [12], which first derived EWR in settings with small code subspaces, where the minimal QES is determined up to perturbative corrections by the area term.

Furthermore, we explain that the AdS/CFT dictionary performs this task *maximally efficiently*. EWR is just very efficient one-shot quantum state merging.²¹

6.1 State-specific EWR

Let us first carefully define what we mean by EWR for an arbitrary, single bulk state ρ .

Traditionally, EWR has been defined not for a single (mixed) state ρ , but for a *code subspace* of states $\mathcal{H}_{\text{code}}$. There are then two definitions of what it means for EWR to be possible, depending on whether we work in the Schrödinger or Heisenberg picture. In the Schrödinger picture, we need to find a quantum channel $\mathcal{R} : B \rightarrow b \otimes b'$ that recovers the reduced bulk state on $b \otimes b'$ from the reduced boundary state on B , for any state in the code subspace. In the Heisenberg picture, for any bulk operator acting on $b \otimes b'$, we need to find an operator O_B , acting only on the boundary region B , whose action is the same as the action of the bulk operator, when applied to any state in the code subspace.²²

We can replace this definition with a definition that considers only a single state, by utilizing a canonical purification $|\psi\rangle_{bb'\bar{b}R}$ of the maximal mixed state within the code subspace. In this language, EWR is possible if and only if, for any bulk operator on $b \otimes b'$, there exists an operator reconstruction on B that has the correct action on $|\psi\rangle_{bb'\bar{b}R}$. Similarly, in the Schrödinger picture, EWR is possible — in this single-state language — if and only if it is possible to recover a canonical purification of the bulk state on $\bar{b} \otimes R$ from the boundary state on region B .

If EWR were exact, this single-state definition would be exactly equivalent to the traditional, code subspace definition. However, because EWR is in practice only approximate, there is a slight difference. In the traditional definition, the error is commonly defined as the ‘worst-case’ error, i.e. the largest output error for any input state. The error when acting on a maximally entangled state is more like an ‘average-case’ error: the reconstruction can do a lot worse on particular input states, as long as it does well for most input states. (See e.g. the discussion in [38].)

An advantage of this new, single-state definition is that it very naturally generalizes to mixed bulk states ρ with an entanglement spectrum that isn’t flat. Again, we simply say that EWR is possible if a boundary operator exists with the correct action on a (canonical) purification of the bulk state ρ . When the state ρ is unentangled, this just means that we are taking a ‘weighted-average’ error, where $\rho_{b'}$ tells us how different states should be weighted. However, when the state $\rho_{bb'}$ is entangled, we can take advantage of that entanglement to make reconstruction easier. This has no classical analogue.

When is EWR of region b' — using this more general definition — possible? We start by considering the tensor network shown in figure 6. In this setup, a necessary and sufficient

²¹Let us make a helpful distinction. The term “entanglement wedge reconstruction” usually means two things at the same time: the task of encoding b' into B (and then decoding), and also the particular protocol implicit in the AdS/CFT dictionary, the protocol that performs the task. The task, we will explain, is a special case of quantum state merging. The protocol, we will argue, is a very efficient way to perform quantum state merging.

²²We also require that, when the bulk operator is Hermitian, the boundary reconstruction is also Hermitian, and, when the bulk operator is unitary, the boundary operator is also unitary.

condition for EWR is *approximate decoupling* [39, 40]. Namely, that

$$\left\| \text{tr}_{B'} \left(V \rho_{b'\bar{b}R} V^\dagger \right) - \rho_{\bar{b}R} \otimes \frac{\mathbb{1}_{\bar{B}'}}{d_{\bar{B}'}} \right\|_1 \leq \varepsilon. \quad (6.2)$$

Roughly speaking, the intuition for this is that all purifications are equivalent up to unitaries, and B purifies $\bar{B}' \otimes \bar{b} \otimes R$. It follows that, if (and only if) the reduced state on $\bar{B}' \otimes \bar{b} \otimes R$ is (approximately) the product of a state on \bar{B}' and a state on $\bar{b} \otimes R$, then we can extract a purification of $\bar{b} \otimes R$ from B . As discussed above, this is just the Schrödinger picture definition of EWR.

As discussed in section 5.1, (6.2) holds if and only if

$$-H_{\min}^\varepsilon(b'|\bar{b} \otimes R) = H_{\max}^\varepsilon(b'|b) \ll \ln \frac{d_{B'}}{d_{\bar{B}'}} = \frac{A_2 - A_1}{4G}. \quad (6.3)$$

In other words, the condition for EWR of region b' is exactly the condition for the QES prescription to be valid, with minimal QES γ_1 (and hence region b' is ‘in the entanglement wedge’).

We would like to show that the same condition holds for EWR in gravity. Given our discussion in section 5.1 about the close connections between random tensor networks and fixed-area states, it should be unsurprising that this indeed the case.

The simplest argument for this is to use the Petz map reconstruction [41–43]. This is an explicit general-purpose construction for reconstructing operators that is known to be close to optimal. Specifically, using the Petz map (with reference state $\rho_{bb'} \otimes \sigma_{\bar{b}}$ for any full-rank state $\sigma_{\bar{b}}$) will give a reconstruction error that is at most twice the optimal error [38, 41]. Hence, for the random tensor network the Petz map reconstruction will work with small error, if and only if (6.3) holds.

However, Petz map matrix elements can be computed using a replica trick [9]. And, as for the von Neumann entropy, the replica trick calculation is identical for both fixed-area states and random tensor networks [9, 44]. We can therefore use the known results for random tensor networks to do the analytic continuation and conclude that the Petz map reconstruction succeeds (and hence EWR is possible at all) if and only if (6.3) holds.

What about EWR in states where the extremal surface areas are not fixed? Since the area operator A_2 can always be measured on B , we are free to consider states of fixed A_2 . If entanglement wedge reconstruction is possible for all values of the area A_2 , it must also be possible for states that involve superpositions over A_2 , because we can reconstruct an operator $\phi_{b'}$ as

$$\phi_B = \sum_{A_2} \Pi_{A_2} \phi_B^{(A_2)} \Pi_{A_2}, \quad (6.4)$$

where the sum is over possible values of the area A_2 , $\phi_B^{(A_2)}$ is a reconstruction of $\phi_{b'}$ for states with area A_2 , and Π_{A_2} is a projector onto the area being A_2 .

In general, we can’t do the same thing for the area A_1 , since it is not always measurable from B . However, if the region b' is reconstructable on B for all states in the superposition,

then A_1 can be reconstructed in B for all the states, and we can use exactly the same argument to construct operators that work for superpositions of eigenstates of A_1 . We therefore conclude that entanglement wedge reconstruction is possible so long as

$$\frac{A_2 - A_1}{4G} - H_{\max}^\epsilon(b'|b) \gg \mathcal{O}\left(\sqrt{\frac{\ln G}{G}}\right). \tag{6.5}$$

The above argument was somewhat sloppy. Our previous argument for EWR of the region b' , in fixed-area states, involved operators that acted within a single fixed-area code subspace (as in the tensor networks). The operator A_1 instead compares code subspaces with different areas. How do we know that it has the same reconstruction conditions?

Again, we can turn to the Petz map. To reconstruct the operator A_1 using the Petz map, we need to consider a reference state that involves a mixture of states with different areas A_1 . In the replica trick calculation of the Petz map matrix elements, the area A_1 in each replica has to be the same, whenever the different replicas are glued together at the surface γ_1 . If this is the case, the parts of the mixed reference state with the ‘wrong’ area A_1 will not contribute to the operator action, and the reconstruction will succeed. If some of the replicas are instead glued together at the surface γ_2 , then the areas A_1 do not need to be the same, and the reconstruction will fail.

The same statement is also true for the Petz reconstruction of ordinary bulk operators in region b' [9]: the reconstruction succeeds if and only if the contribution from saddles where replicas are glued together at γ_2 is small (and so can be safely ignored while doing the analytic continuation). We already argued that those reconstructions succeed when (6.3) holds. Hence, when (6.3) holds, gluing at γ_1 must dominate the analytic continuation, and hence the operator A_1 must also be reconstructible.

6.2 EWR as one-shot quantum state merging

This single-state reformulation of EWR is a special case of a ubiquitous information-theoretic task, known as one-shot quantum state-merging [21, 45, 46].

In quantum state-merging, Alice and Bob share a quantum state. This state is chosen from some arbitrary ensemble of pure states with density matrix ρ_{AB} . Alternatively, we can consider a single purification $|\psi\rangle_{ABR}$ of ρ_{AB} . The objective of the task is to transfer Alice’s part of the state to Bob while sending as few qubits from Alice to Bob as possible. In other words, to produce an output state $|\psi'\rangle_{A'BR} \approx |\psi\rangle_{ABR}$ where the A' and B subsystems are both held by Bob. Equivalently, the average error between the initial state (shared between Alice and Bob) and the final state (held only by Bob), should be small, where the average is over the pure states in the ensemble with density matrix ρ_{AB} .

It should be clear that this task is closely related to state-specific EWR. There too, the bulk state either is chosen from some ensemble $\rho_{bb\bar{b}}$, or is purified by a reference system as $|\psi\rangle_{bb\bar{b}R}$. The part of the state that is held by Bob corresponds to the part of the state that is encoded in the boundary region B . The bulk region b is always encoded in the boundary region B ; this corresponds to the part of the state that is *initially* held by Bob.

In the one-shot setting (where Alice and Bob are trying to merge a single copy of the state), it is known that the minimum number of qubits required for state merging is

$H_{\max}(A|B)/\ln(2)$. Remarkably, this is exactly how many qubits gravity seems to require! The number of qubits from region b' that can be decoded in region B is $\Delta A/4\ln(2)G$, as stated in (6.3). Hence it seems that EWR can be explained not just as a special case of quantum state merging, but as an optimal implementation of it!

However, there is an important caveat that we have ignored until now. In quantum state merging as traditionally defined, it is crucial that unlimited *classical* information can be sent from Alice to Bob [21]. Without this classical communication, significantly more quantum communication would be required.

Holography does *not* transfer large amounts of classical information from b' to B . Indeed, the amount of transferred classical information is bounded by the Holevo information, which is also equal to $\Delta A/4G$ [47]. That is, the total number of transferred qubits *plus* bits is bounded by $\Delta A/4G$. There is no additional classical communication that can make state merging achievable.

So if EWR is accomplishing state merging, why did our results from section 6.1 suggest that we only need

$$\frac{\Delta A}{4G} > H_{\max}^{\varepsilon}(A|B) \tag{6.6}$$

for EWR to be possible? It turns out that the full power of classical communication is unnecessary for quantum state merging. Instead, a weaker communication primitive, known as zero-bit communication, is sufficient [48]. The number of zero-bits communicated from region b' encoded in region B is not constrained by ΔA , and it is this additional information that allows the state merging protocol to succeed when (6.6) holds.

To understand this, we start with the resource inequality governing a highly efficient, rather general quantum protocol, the ‘‘one-shot mother protocol,’’ also known as (one-shot) quantum state transfer or fully quantum Slepian-Wolf [49, 50]. The inequality states that

$$\langle \psi_{ABR} \rangle + \frac{[H_0^{\varepsilon}(A)_{\psi} + H_{\max}^{\varepsilon}(A|B)_{\psi}]}{2 \ln(2)} \text{ qubits} \geq \frac{[H_0^{\varepsilon}(A)_{\psi} - H_{\max}^{\varepsilon}(A|B)_{\psi}]}{2 \ln(2)} \text{ ebits} + \langle \psi_{A'BR} \rangle. \tag{6.7}$$

At first glance, this inequality is somewhat terrifying. Let’s take some time to unpack it. The whole statement relates the relative usefulness of different quantum communication resources. On the left, we start with the state $|\psi\rangle$, which is shared between Alice, Bob, and the reference R . Alice also has the ability to send $[H_0^{\varepsilon}(A)_{\psi} + H_{\max}^{\varepsilon}(A|B)_{\psi}]/2 \ln(2)$ qubits to Bob.

The claim is that this is more useful to Alice and Bob than the resources on the right hand side, because the resources on the left can be used to *create* the resources on the right (up to some small error). What are the resources on the right? We still have the state $|\psi\rangle$, but it has now been successfully ‘merged,’ so that everything except the reference is now in system $A'B$, held entirely by Bob. Alice and Bob have also gained $[H_0^{\varepsilon}(A)_{\psi} + H_{\max}^{\varepsilon}(A|B)_{\psi}]/2 \ln(2)$ Bell pairs or ‘ebits’.

For clarity of presentation, we dropped additional terms in (6.7) of size $\mathcal{O}(\ln \varepsilon)$, terms correcting the number of qubits required and ebits produced. These corrections are sub-leading for appropriate choices of ε in the limit where the entropies are large. We note that

the inequality is optimal in the following sense: in any protocol for one-shot quantum state transfer, the number of qubits communicated, minus the ebits of entanglement gained, will be at least

$$H_{\max}^{\varepsilon'}(A|B)_\psi + \mathcal{O}(\ln(\varepsilon')),$$

for a particular ε' that is controlled by the protocol error.

How does this relate to quantum state merging? In the language of resource inequalities, *quantum teleportation* states that

$$1 \text{ ebit} + 2 \text{ cbits} \geq 1 \text{ qubit}, \tag{6.8}$$

where a cbit is a classical bit. Substituting this inequality into (6.7), and recalling that classical communication is free in traditional quantum state merging, we find that the number of qubits that need to be sent is $H_{\max}^\varepsilon(A|B)$. Hence unlimited classical communication does allow Alice to give her state to Bob, just by using the mother protocol and transferring $H_{\max}^\varepsilon(A|B)$ qubits.

As an aside: note that quantum conditional entropies can be negative. What does it mean if only a negative number of qubits need to be sent from Alice to Bob? The answer is that the communication cost in state merging is defined *catalytically*. If the protocol produces Bell pairs, these can be stored, ready to use, together with the free classical communication, to produce quantum communication in the future. We can end up with more ability to communicate than we started with!

Returning to the main point, we emphasize that classical bits are not actually required to do teleportation. Zero-bits are sufficient. We have

$$1 \text{ ebit} + 2 \text{ zero-bits} \stackrel{(a)}{=} 1 \text{ qubit}, \tag{6.9}$$

where the (a) means that (6.9) only holds at leading order in the limit where we have a large number of each type of bit. Note that, unlike (6.8), (6.9) is an equality, not an inequality. Zero-bits are the minimal resource required for teleportation.

Therefore, with enough zero-bits communicated from Alice to Bob, Alice can give Bob her state with just $H_{\max}^\varepsilon(A|B)/\ln(2)$ qubits, using the mother protocol. To see this, substitute (6.9) into (6.7), finding that

$$\langle \psi_{ABR} \rangle + \frac{H_{\max}^\varepsilon(A|B)_\psi}{\ln(2)} \text{ qubits} + \frac{[H_0^\varepsilon(A)_\psi - H_{\max}^\varepsilon(A|B)_\psi]}{\ln(2)} \text{ zero-bits} \geq \langle \psi_{A'BR} \rangle. \tag{6.10}$$

State merging is just as easy with free zero-bit communication as with free classical communication.

How many zero-bits *are* communicated from b' to B ? More broadly, what is the total amount of information about b' encoded in region B ? These questions were answered in [8].²³ For $\Delta A > 0$, region B encodes the ‘ α -bits’ of region b' for

$$\alpha = \frac{\Delta A}{4GS_0}. \tag{6.11}$$

²³The total information was correctly computed in [8], even though they used the naïve prescription, because the relevant state for computing the transferred information is the maximally-mixed state, which is perfectly compressible.

Here $S_0 = \ln(d_{b'})$ is the thermodynamic entropy in region b' . So, for example, when the code space states in region b' are the possible microstates of a black hole with horizon area A_{hor} , we have $\alpha = \Delta A/A_{\text{hor}}$.

We can convert α -bits into a mixture of qubits and zero-bits using another resource equality from [48], namely

$$1 \text{ } \alpha\text{-bit} = \alpha \text{ qubits} + (1 - \alpha) \text{ zero-bits} . \tag{6.12}$$

We therefore find that region B can receive

$$\frac{1}{\ln(2)} S_0 \text{ } \alpha\text{-bits} = \frac{\Delta A}{4 \ln(2) G} \text{ qubits} + \frac{1}{\ln(2)} \left[S_0 - \frac{\Delta A}{4G} \right] \text{ zero-bits} \tag{6.13}$$

from region b' . This is worth emphasizing: the AdS/CFT dictionary transfers more than $\Delta A/4 \ln(2)G$ qubits of information from b' to B . It also transfers many zero-bits, precisely $\left[S_0 - \frac{\Delta A}{4G} \right] / \ln(2)$.

That was for $\Delta A > 0$; what about $\Delta A < 0$? In this case, region B encodes no physical information about region b' (if b' is not heavily entangled with b). Nonetheless, the right hand side of (6.13) still formally defines the amount of information from b' accessible in B . This is important, for example, if we start adding bulk entanglement, as in the following scenario. Imagine that more than $|\Delta A|/4 \ln(2)G$ Bell pairs are shared between regions b' and b . Then the zero-bits of the remaining degrees of freedom in b' will be encoded in B . This follows from the associated phase transition in the minimal QES. This phase transition is reflected in (6.13) in the following way. Converting qubits into ebits and zero-bits using (6.9), the right hand side of (6.13) says that $|\Delta A|/4 \ln(2)G$ ebits allow $S_0 - |\Delta A|/4 \ln(2)G$ zero-bits to be transferred from b' to B , which is exactly what we just found. (Any additional ebits will continue to combine with those zero-bits to form qubits of communication, reflecting the fact that adding more and more entanglement between b' and b allows B to recover larger and larger subspaces of b' .)

As an aside, we emphasize that (6.13) allowing additional zero-bits (on top of $\Delta A/4 \ln(2)G$ qubits) from region b' to be encoded in region B is not some strange phenomenon that only happens in quantum gravity. Instead, it happens very generically whenever you have a noisy quantum channel. Consider the well-known properties of the quantum capacity of a channel, i.e. the number of qubits that can be communicated through that channel. The quantum capacity of a noisy channel is given by the so-called maximal regularized coherent information. However, the *entanglement-assisted* quantum capacity is given by half the maximal mutual information, and is generically strictly larger. The difference comes from the channel having an additional zero-bit capacity. Free entanglement allows the zero-bits to be ‘upgraded’ to qubits, giving additional qubit capacity.

Let’s see how the same phenomenon manifests itself in gravity. Suppose we have $S_0 > \Delta A/4G$. In this case, without using entanglement, we can learn, at most, $\Delta A/4 \ln(2)G$ qubits in b' from B . Not all the information is encoded there. However, let’s imagine we entangle $(S_0 - \Delta A/4G)/2 \ln 2$ Bell pairs between region b and b' . If we do this, all the information about the remaining $(S_0 + \Delta A/4G)/2 \ln 2$ qubits in region b' will be successfully encoded in region B (the entanglement wedge will have expanded to include b'). By using

entanglement between b and b' , we have increased the amount of information about region b' that is accessible in region B . This increase in information capacity from entanglement assistance comes from the extra zero-bits in (6.13).

Having understood the information transferred from bulk to boundary, we are now ready to interpret the conditions for EWR that we found in section 6.1. We first note that for any state $|\psi\rangle$, we have

$$H_0^\varepsilon(b')_\psi \leq H_0(b')_\psi = \ln \text{Rank}(\psi_{b'}) \leq \ln d_{b'} = S_0. \tag{6.14}$$

It therefore follows from (6.10) and (6.13) that there are sufficient qubits and zero-bits for state merging, and hence the encoding (and reconstruction) of region b' from region B using *any* protocol, if and only if

$$H_{\max}^\varepsilon(b'|b)_\psi < \frac{\Delta A}{4G}. \tag{6.15}$$

This is exactly what we found in section 6.1.

To summarize, we noted that the task of encoding b' in B is the same as the task in quantum state merging. This simply followed from definitions. We were then led to ask how *efficiently* AdS/CFT *performs* this task, requiring us to carefully account for exactly how much information is transferred from b' to B by the AdS/CFT dictionary. The total information, we noted, is $\Delta A/4 \ln(2)G$ qubits *plus* additional zero-bits (6.13). This is just enough transferred information for the most efficient state-merging protocol (the mother protocol) to work. I.e. one could not transfer the bulk information in b' to B using any fewer resources. It's remarkable that AdS/CFT encodes b' in B exactly when just enough information is transferred from b' to B for *any* protocol to do it. EWR is a maximally efficient state merging protocol.

In contrast, the naïve QES prescription suggests that AdS/CFT exceeds the maximal efficiency bound, performing state merging as though every state were perfectly compressible.²⁴

We emphasize that the arguments in this section should not be interpreted as an independent proof of the results from section 5. A channel having sufficient capacity to carry out some task does not automatically mean that any (possibly inefficient) protocol using that channel will actually perform the task. Conversely, one could worry that region B might encode some other form of information about region b' , distinct from both qubits and zero-bits, which could help make state merging possible even when the zero-bits and qubits alone would be insufficient.

Instead, our point was to make precise the relationship between entanglement wedge reconstruction (and other questions in AdS/CFT) and standard protocols in quantum information, such as state merging, which may not have been clear to members of either community.

²⁴This was the realization that led to this work. From [8] we knew the amount of information being transferred from bulk to boundary. Seemingly in contradiction was the fact that the naïve QES prescription implies EWR for any state with small enough von Neumann entropy [11, 12, 17]. (Taking into account reconstruction errors means that this is only true using the state-specific definition from section 6.1.) The resolution is that the naïve prescription needs to be refined.

In particular, we want to emphasize that the relevant quantum information protocols are always *one-shot* protocols. After all, in AdS/CFT, one only typically considers a single copy of a holographic state, rather than a large number of identical copies. The only reason that the von Neumann entropy has proven relevant is that until now people have generally only considered states where the von Neumann entropy is equal to the one-shot entropies, at least at leading order. Once you consider states where this is not the case, it should not be surprising that it is one-shot entropies which play the crucial role.

7 Beyond two extremal surfaces

So far we have presented refined conditions for the QES prescription when there are exactly two competing surfaces, (1.12). In this section, we discuss the natural generalization of this rule which considers *all* bulk surfaces homologous to B .

The upshot is that the condition for large corrections is no longer two simple inequalities; it becomes a family of inequalities. Together these inequalities determine what information is actually transmitted to B .

All the claims about reconstruction in this section can be shown in random tensor networks using a careful application of the one-shot decoupling theorem. We expect based on our arguments from section 5 that they should also be true in AdS/CFT.

7.1 Applying the refined prescription

The refined way to find the entanglement wedge (EW) is as follows.²⁵

Step 1: find the *max*-entanglement wedge (max-EW). The max-EW is intuitively the bulk region that B can definitely reconstruct with small error. In this sense, it most closely resembles the traditional operational definition of the entanglement wedge.

We define the max-EW as the largest region b that satisfies all of the following inequalities:

$$\forall b' \subset b, H_{\max}^\epsilon(b - b'|b') < \frac{A(b') - A(b)}{4G}, \tag{7.1}$$

where $b - b'$ is the complement of b' in b .

This definition implicitly assumes that there exists some ‘largest’ region satisfying (7.1) that contains all other regions satisfying (7.1). We shall prove in the next subsection that this is indeed the case. The essential intuition is that, if we can reconstruct region b_1 , and we can reconstruct region b_2 , then we should also be able to reconstruct their union. Having access to additional degrees of freedom can only make reconstruction easier.

In principle, (7.1) requires checking infinitely many subregions b' . However, in practice, except in situations where the bulk entropy gradients can become very large (such as evaporating black holes) it should be sufficient to only check regions where $\partial b'$ is perturbatively close to a classical extremal surface. This is because the classical area gradient

²⁵These are the refined conditions for moments of time symmetry. We expect there exists a covariant generalization, in the way HRT [51] generalized RT [52]. This may well require use of the maximin formalism [53, 54].

must be $\mathcal{O}(G)$ at minima of $(A(b') - A(b))/4G - H_{\max}^\varepsilon(b - b'|b')$. This justifies the simple conditions given in (1.12) when only two extremal surfaces exist.

Step 2: find the *min*-entanglement wedge (min-EW). The min-EW is the complement of region B definitely knows no information about. In other words, it is the region that region B may know at least some information about. For pure states, it is the complement of the max-EW of \bar{B} . For mixed states, it can be smaller.

We define the min-EW as the smallest region b that satisfies all of the following inequalities, for \bar{b} the complement of b :

$$\forall \bar{b}' \subset \bar{b}, H_{\min}^\varepsilon(\bar{b}'|b) > \frac{A(b) - A(b\bar{b}')}{4G}, \tag{7.2}$$

where $b\bar{b}'$ is the union of b and \bar{b}' . Again, the existence of a smallest such region is nontrivial, and is equivalent to the existence of a max-EW for the purification of B , namely $\bar{B}R$.

Step 3: define EW as min-EW = max-EW. In general, the max-EW is contained in the min-EW, as we will prove in the next subsection. In the special case in which they are the same, we can define the EW to be equal to both of them, and the entropy $S(B)$ equals the generalized entropy of this EW.

However, if the min-EW contains a region that the max-EW doesn't, then B may have partial information about that region. In general in such cases, the entanglement entropy $S(B)$ will not be equal to the generalized entropy of any single surface.

7.2 Properties of the min-EW and max-EW

In this subsection, we prove several important properties of the min-EW and max-EW. To do so, we will need certain inequalities that are satisfied by smooth min- and max-entropies. The first is that both the min- and max-entropies satisfy strong subadditivity

$$H_{\min/\max}^\varepsilon(A|B) \geq H_{\min/\max}^\varepsilon(A|BC). \tag{7.3}$$

Secondly, the smooth min- and max-entropies satisfy a number of approximate chain rule inequalities [55]. Most importantly for our purposes, we have

$$H_{\max}^\varepsilon(AB|C) \leq H_{\max}^\varepsilon(A|BC) + H_{\max}^\varepsilon(B|C) + \mathcal{O}(\ln \varepsilon). \tag{7.4}$$

Property 1: existence of the min/max-EW. We will show that, given any two regions satisfying (7.1), their union will also satisfy (7.1). This immediately implies the existence of the max-EW, and implies the existence of the min-EW by the equivalence with the min-EW of $\bar{B}R$. To prove this, we need to consider three overlapping regions: the two original regions, and an arbitrary subregion b' of their union. These three overlapping regions can be decomposed into six disjoint regions, which we label $b_0, b'_0, b_1, b'_1, b_2, b'_2$, as shown in figure 7. The original two regions are given by $b_0b'_0b_1b'_1$ and $b_0b'_0b_2b'_2$. Their union is then $b = b_0b'_0b_1b'_1b_2b'_2$. We need to show that

$$H_{\max}^\varepsilon(b - b'|b') < \frac{A(b') - A(b)}{4G} \tag{7.5}$$

for the arbitrary region $b' = b'_0b'_1b'_2$ in b .

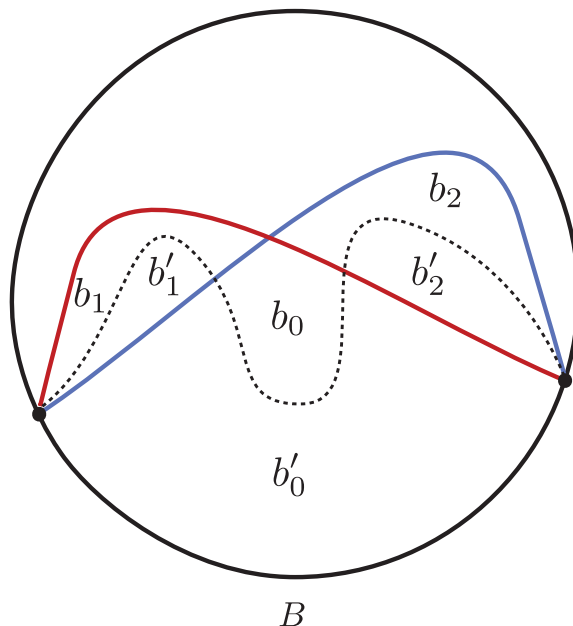


Figure 7. Setup in the proof of Property 1. Both the region bounded by the blue surface and that bounded by the red surface are assumed to satisfy (7.1). We prove that therefore the union of those regions satisfies (7.1), by showing it to hold for an arbitrary choice $b' = b'_0 b'_1 b'_2$, depicted here bounded by the dashed black line.

Because the original two regions satisfied (7.1), we know that

$$H_{\max}^\varepsilon(b_0 b_i | b'_0 b'_i) < \frac{A(b'_0 b'_i) - A(b_0 b'_0 b_i b'_i)}{4G}, \quad (7.6)$$

for $i = 1$ or 2 , as well as

$$H_{\max}^\varepsilon(b_i | b_0 b'_0 b'_i) < \frac{A(b_0 b'_0 b'_i) - A(b_0 b'_0 b_i b'_i)}{4G}. \quad (7.7)$$

Adding together these four inequalities (two for each of the two regions) and comparing the area terms, we find

$$H_{\max}^\varepsilon(b_0 b_1 | b'_0 b'_1) + H_{\max}^\varepsilon(b_1 | b_0 b'_0 b'_1) + H_{\max}^\varepsilon(b_0 b_2 | b'_0 b'_2) + H_{\max}^\varepsilon(b_2 | b_0 b'_0 b'_2) \leq 2 \frac{A(b') - A(b)}{4G}. \quad (7.8)$$

We can then simplify the left hand side, using

$$H_{\max}^\varepsilon(b_0 b_1 | b'_0 b'_1) + H_{\max}^\varepsilon(b_2 | b_0 b'_0 b'_2) \geq H_{\max}^\varepsilon(b_0 b_1 | b'_0 b'_1 b'_2) + H_{\max}^\varepsilon(b_2 | b_0 b'_0 b_1 b'_1 b'_2) \quad (7.9)$$

$$\geq H_{\max}^\varepsilon(b_0 b_1 b_2 | b'_0 b'_1 b'_2) + \mathcal{O}(\ln \varepsilon). \quad (7.10)$$

The first inequality uses SSA and the second uses the chain rule (7.4). Together with a similar set of inequalities with 1 and 2 exchanged, this gives

$$2H_{\max}^\varepsilon(b_0 b_1 b_2 | b'_0 b'_1 b'_2) + \mathcal{O}(\ln \varepsilon) \leq 2 \frac{A(b') - A(b)}{4G}. \quad (7.11)$$

The max- and min-EW are therefore well-defined, up to $\mathcal{O}(\ln \varepsilon)$ corrections (which is the same entropy difference that was required for EWR and the QES prescription to hold safely, anyway).

Property 2: min-/max-EW nesting. Almost the exact same argument shows that the max-EW and min-EW satisfy nesting. That is, a boundary region $B_1 \subseteq B_2$ must have a max-EW (min-EW) that is entirely contained in the max-EW (min-EW) of B_2 .

To prove this for the max-EW, once again let the regions $b_0, b'_0, b_1, b'_1, b_2, b'_2$ be disjoint, with the max-EW of B_1 given by $b_0 b'_0 b_1 b'_1$ and the max-EW of B_2 given by $b_0 b'_0 b_2 b'_2$. Their union is $b = b_0 b'_0 b_1 b'_1 b_2 b'_2$. We need to show that

$$H_{\max}^\varepsilon(b - b'|b') < \frac{A(b') - A(b)}{4G} \tag{7.12}$$

for an arbitrary region $b' = b'_0 \cup b'_1 \cup b'_2$ in b . This will imply that the max-EW of B_2 should have included $b_1 b'_1$ since the beginning. The proof, given this setup, is identical to the previous one.

The proof for the min-EW follows from nesting of the max-EW of the complement plus a purifying reference system.

Property 3: max-EW \subseteq min-EW. The max-EW is always contained in the min-EW. Intuitively this must be true if, as we claim, the max-EW characterizes the region that B has (approximately) *all* information about, while the min-EW characterizes the region that B has *any* information about.

To prove this, we assume for contradiction that there is some region b' that is contained in the max-EW, but not in the min-EW. Let b be the intersection of the max- and min-EWs, let \bar{b}' be the region contained in the min-EW, but not the max-EW and let \bar{b} be the complement of the union of the two wedges.

Then it must both be true that

$$H_{\max}^\varepsilon(b'|b) < \frac{A(b) - A(b'b)}{4G}, \tag{7.13}$$

and that

$$H_{\min}^\varepsilon(b'|b\bar{b}') > \frac{A(b\bar{b}') - A(b\bar{b}'b')}{4G}. \tag{7.14}$$

However,

$$\frac{A(b) - A(b'b)}{4G} \leq \frac{A(b\bar{b}') - A(b\bar{b}'b')}{4G}, \tag{7.15}$$

while

$$H_{\max}^\varepsilon(b'|b) \geq H_{\max}^\varepsilon(b'|b\bar{b}') \geq H_{\min}^\varepsilon(b'|b\bar{b}'). \tag{7.16}$$

We therefore have our desired contradiction.

Property 4: max-EW = min-EW only at minimal generalized entropy surfaces.

In the special case that the min-EW and max-EW equal the same region b , they must be bounded by a surface that minimizes $A(b)/4G + S(b)$.

Consider a general deformation $b' = b_0b_2$ of $b = b_0b_1$. We want to show

$$\frac{A(b_0b_2)}{4G} + S(b_0b_2) > \frac{A(b_0b_1)}{4G} + S(b_0b_1). \tag{7.17}$$

From (7.1) we know

$$\frac{A(b_0) - A(b_0b_1)}{4G} \geq H_{\max}^\epsilon(b_1|b_0). \tag{7.18}$$

This implies

$$\frac{A(b_0b_2) - A(b_0b_1b_2)}{4G} > H_{\max}^\epsilon(b_1|b_0b_2) \tag{7.19}$$

$$\geq S(b_0b_1b_2) - S(b_0b_2), \tag{7.20}$$

where in the first line we used SSA of both area and max-entropy, and in the second line we used $H_{\max}^\epsilon(A|B) \geq S(A|B)$. Meanwhile, (7.2) tells us

$$\frac{A(b_0b_1) - A(b_0b_1b_2)}{4G} < H_{\min}^\epsilon(b_2|b_0b_1) \tag{7.21}$$

$$\leq S(b_0b_1b_2) - S(b_0b_1), \tag{7.22}$$

where in the second line we used $H_{\min}^\epsilon(A|B) \leq S(A|B)$. Combining these two inequalities gives (7.17), where the inequality must be strict for non-trivial b_1b_2 because (7.18) and (7.21) are strict for non-trivial b_1 and b_2 respectively. This is what we set out to show.

The converse is not true. A minimal generalized entropy surface will not in general satisfy all of (7.1) and (7.2). However, if all states were perfectly compressible, then this converse would be true, and therefore the naïve QES prescription would hold. Indeed, (conditional) perfect compressibility implies $H_{\min}^\epsilon = S = H_{\max}^\epsilon$, and equations (7.1) and (7.2) would both be satisfied only by the minimal generalized entropy surface.

This is one way to understand the refined conditions for the QES prescription.

7.3 Full reconstruction outside the max-EW

Everything in the max-EW can be fully reconstructed from B . Similarly no information reaches B from degrees of freedom outside the min-EW. However, the converses of these statements are not necessarily true. There can be regions outside the max-EW which can be fully reconstructed; and regions inside the min-EW that cannot. Nonetheless, when the min-EW and max-EW are not equal, there is always *some* nonempty intermediate region that is partially, but not fully, reconstructible.

A tensor network example will make this clearer. Consider m tripartite random tensors arranged in a line, each with bulk leg b_i for $i \in \{1, \dots, m\}$, each connected to the tensors to its left and right by maximally entangled “in-plane” legs B'_i , of dimension $e^{A_i/4G}$, except for the first and last tensor, which have one dangling in-plane leg each (the “boundary”

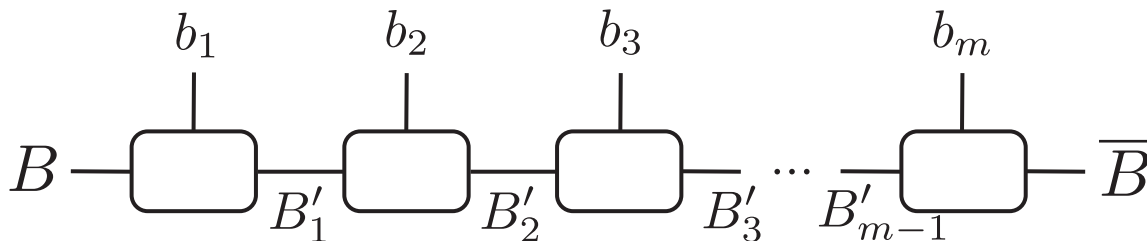


Figure 8. Random tensor network used to demonstrate reconstruction outside the max-EW. Each box is a tripartite random tensor with “bulk” leg b_i , connected to the two tensors on either side of it with maximally entangled “in-plane” legs B'_{i-1} and B'_i . The tensors on the ends each have a single “boundary” leg, B or \bar{B} . Which bulk legs are reconstructable on B depends on two things: the dimensions of the in-plane legs and the state on the bulk legs. A lower bound on the size of this reconstructable region is the max-EW, but it could be larger, depending on how much helpful information there is from the legs in the min-EW.

legs). Let B be the name of the left boundary leg and \bar{B} be the right one, with dimensions much larger than any $e^{A_i/4G}$. See figure 8.

Consider a bulk state that is a mixture of a) a pure state with a large amount of entanglement (with entanglement entropy S) between b_2 and b_3 and b) a pure state on b_2 and a highly mixed state (with entropy S) on b_3 . If the extremal surface areas satisfy,

$$\frac{A_3}{4G} - S \ll \frac{A_2}{4G} \ll \frac{A_3}{4G} \ll \frac{A_1}{4G} \ll \frac{A_2}{4G} + S, \tag{7.23}$$

then we find that the max-EW is $b = b_1$, while the min-EW is $\hat{b} = b_1 b_2 b_3$.

However, the two states in the mixture are perfectly compressible, with EWs that consist of $b_1 b_2 b_3$ and $b_1 b_2$ respectively. Hence b_2 is reconstructible in both. And it is easy to check that the two states must be close to orthogonal on B (e.g. their entropies differ at $\mathcal{O}(1/G)$ and both have approximately flat spectra). So b_2 must be reconstructible for a mixture of the two states.

How is this possible? The answer is that the max-EW is the largest region that can be reconstructed without knowing anything about the state outside that region. However, because (and only because) the min-EW is larger than the max-EW, some partial information from outside the max-EW makes it through the tensor network legs. In particular, some information from b_3 makes it through the in-plane leg B'_2 (namely the part of the state which is entangled with b_2). And this additional information makes it possible for all the information in b_2 to reach B .

More formally, even though

$$H_{\max}^\varepsilon(b_2|b_1) \gg (A_1 - A_2)/4G, \tag{7.24}$$

we have

$$H_{\max}^\varepsilon(b_2 B'_2|b_1) \ll \frac{A_1}{4G}, \tag{7.25}$$

where the state on B'_2 is the state produced by the entire network to its right (and then tracing over \overline{B}). If the min-EW did not contain anything outside $b_1 b_2$, this state would be maximally mixed and (7.25) would reduce to (7.24). When this is not the case, the in-plane legs can expand the fully reconstructable region.

While sensible in a tensor network, it is not clear how we should define a quantity analogous to $H_{\max}^\varepsilon(b_2 B'_2 | b_1)$ in AdS/CFT, except by explicitly converting the calculation into one involving tensor networks. Hence if the min-EW and max-EW are not equal, it may be hard to identify with certainty the full region of the bulk where everything can be reconstructed in B . It will be at least as big as the max-EW, defined by (7.1), but could be larger (because of additional information from outside the max-EW). Likewise (by looking at the complementary region in a purification, as usual), if the min-EW and max-EW are not equal, the full region of the bulk that B has any information about may be smaller (but not larger) than the min-EW, defined by (7.2).

8 Discussion

Refining the QES prescription. In this paper, we have introduced a refinement of the usual QES prescription. This refinement is both necessary for the boundary entanglement entropies to be self-consistent, and follows from careful application of the replica trick. Without our refinements, the QES prescription would only be valid for the limited subclass of states that are perfectly compressible.

Specifically, we have strengthened the conditions required for the entropy $S(B)$ to be given by the generalized entropy of the minimal QES. In the language of section 7, this is only true when the max- and min-entanglement wedges coincide (perhaps up to perturbative corrections). When the two wedges do not coincide, the entropy $S(B)$ is much more complicated. This is closely related to the breakdown of complementary reconstruction, with a large region that cannot be fully reconstructed from either region B or from its complement.

Fundamental lesson: EWR as one-shot state merging. In many ways, this second point about entanglement wedge reconstruction (EWR) is the more fundamental one. For pedagogical reasons, our presentation was, in a certain sense, inverted. We led by demonstrating the large corrections to the QES prescription, in sections 2 through 5. Only then in section 6 did we explain that EWR should be understood through the lens of one-shot quantum state merging, necessitating the refined conditions for reconstruction.

The QES prescription is just a rule for computing one particular boundary quantity (the von Neumann entropy of a reduced state). This is just one measure of the boundary entanglement structure (albeit a very simple and useful one). EWR is stronger, telling us a deep fact about how information in the bulk is distributed on the boundary that is *independent* of the particular measure (Petz map operators, relative entropies, modular flows etc.) that one might use to probe it.

As we argued in section 6, the information theoretic task of encoding the bulk into the boundary is manifestly a form of one-shot state merging, albeit one that uses zero-bits

rather than the traditional classical bits. Just from this, one can see that the naïve QES prescription implied EWR conditions that were too powerful. There simply is not enough information transferred from the bulk to the boundary via the AdS/CFT dictionary. No quantum information protocol could encode the bulk in the boundary in the way implied by the naïve QES prescription; it is incompatible with quantum Shannon theory.

This reinterpretation of EWR in terms of one-shot quantum state merging seems likely to have important future consequences. For one thing, it opens the door to connecting QES and quantum error-correction [56], providing an understanding of the QES prescription that doesn't come from the Euclidean path integral. This might shed light on how to modify Hawking's calculation of non-unitary black hole evaporation. Indeed — the new arguments from the QES prescription [2, 3, 9, 26] give a unitary answer, but — unlike Hawking — make vital use the Euclidean path integral. A Hilbert space understanding of the QES prescription may connect the calculations.

Generalized min-/max-entropy. That we know of, these refinements are the first example of a generalization of the generalized entropy that replaces the von Neumann entropy by a new entropy measure (in this case the smooth min-/max-entropy). The generalized entropy of a codimension-2 surface, defined as the area plus matter von Neumann entropy,²⁶ is believed to be a well-defined continuum quantity, having passed many non-trivial checks. It is UV finite, is scheme independent, and seems to correctly generalize the classical area in many classical general relativity theorems [1, 59–61]. It therefore made perfect sense to promote extremal area surfaces to extremal generalized entropy surfaces, in the naïve QES prescription.

In contrast, the refined QES prescription asks us to do something new: to add the smooth min-entropy or max-entropy of the bulk fields to the area. The arguments from this paper suggest that this must be equally well-defined. In particular, there should be an appropriate renormalization procedure that makes these differences UV-finite.

The leading UV-divergence in the smooth min- and max-entropy of a subregion in quantum field theory is proportional to the area (just like for the von Neumann entropy). This is essentially because the UV-divergent parts of the subregion states are thermal Rindler modes, and hence are perfectly compressible.

However, the difference between the von Neumann entropy and the smooth min-/max-entropy will still be $\mathcal{O}(\sqrt{S})$ and hence UV-divergent [29, 62]. This means that the smooth min-/max-entropies cannot be renormalized by the same quantity as the von Neumann entropy.

This is OK. As discussed in section 5.3, the relevant area difference is not the expectation of the difference in area, but a *lower confidence bound* on the difference in areas. This differs from the expectation of the difference by $\mathcal{O}(\sqrt{G})$, which is the correct scaling to renormalize the difference between von Neumann entropies and min/max-entropies.

It is therefore natural to hope that the generalized smooth min- and max-entropies, defined as

$$H_{\min/\max}^\epsilon(b) + \frac{A_{\min/\max}^\epsilon(b)}{4G} \tag{8.1}$$

²⁶Or more generally, the gravitational entropy [57, 58] plus matter von Neumann entropy.

with $A_{\min/\max}^\varepsilon$ respectively lower and upper-confidence bounds on the area, should be a UV-finite quantity.

If this is indeed the case, we should also expect that the conditional generalized smooth max-entropy

$$H_{\max}^\varepsilon(b'|b) + \frac{[A(bb') - A(b)]_{\max}^\varepsilon}{4G} \quad (8.2)$$

should also be UV-finite (note $[A(bb') - A(b)]_{\max}^\varepsilon$ is again an upper confidence bound on $A(bb') - A(b)$). There are two sets of modes that give divergent contributions to $H_{\max}^\varepsilon(b'|b)$: modes near the boundary of bb' , and modes near the boundary between b and b' . The contribution to the divergence of $H_{\max}^\varepsilon(b'|b)$ from UV-modes near $\partial(bb')$ will be the same as for the smooth max-entropy $H_{\max}^\varepsilon(b'|b)$ (because these modes are unentangled with b). Meanwhile the divergence from UV-modes near ∂b will be the same as for the smooth min-entropy $H_{\min}^\varepsilon(b|\bar{b})$, except with the opposite sign, because $H_{\max}^\varepsilon(b'|b) = -H_{\min}^\varepsilon(b|\bar{b})$. Hence we should expect the total divergence to be renormalized by $[A(bb') - A(b)]^\varepsilon$.

As usual, the UV-finiteness of the smooth conditional generalized min-entropy also follows by considering complementary subsystems.

Note that the conditional generalized smooth min- and max- entropies should also be IR finite (just like the conditional generalized entropy $(A(bb') - A(b))/4G + S(b'|b)$). This follows from $[A(bb') - A(b)]^\varepsilon$ and $H_{\max}^\varepsilon(b'|b)$ being separately IR-finite. The refined conditions for the QES prescription (1.12) can therefore be written in terms of the sign of the (finite) conditional generalized smooth min- and max-entropies. So instead of (1.12), we should really write

$$S(B)_{\text{refined}} = \begin{cases} \langle A_1 \rangle / 4G + S(bb'), & H_{\max}^\varepsilon(b'|b) + \frac{[A_1 - A_2]_{\max}^\varepsilon}{4G} \leq 0 \\ \text{(depends on details),} & H_{\min}^\varepsilon(b'|b) + \frac{[A_1 - A_2]_{\min}^\varepsilon}{4G} \leq 0 \leq H_{\max}^\varepsilon(b'|b) + \frac{[A_1 - A_2]_{\max}^\varepsilon}{4G} \\ \langle A_2 \rangle / 4G + S(b), & H_{\min}^\varepsilon(b'|b) + \frac{[A_1 - A_2]_{\min}^\varepsilon}{4G} \geq 0. \end{cases} \quad (8.3)$$

This formulation naturally unifies the corrections discussed in this paper with the corrections from [13, 14], which considered situations in which $H_{\max}^\varepsilon = H_{\min}^\varepsilon = 0$, while $[A_1 - A_2]_{\min}^\varepsilon < 0 < [A_1 - A_2]_{\max}^\varepsilon$.

Bit threads. The bit threads paradigm [63], to the extent that it continues to be useful with large bulk entropies, should have a matching refinement. A good first step towards finding it is to incorporate bulk entropy, possibly by allowing threads to end on a “reference system” understood to purify the bulk matter.

A more sophisticated modification that is sometimes mentioned is to allow threads to ‘pass through’ entanglement, effectively using the bulk Bell pairs as ‘Planckian wormholes.’ For this to be consistent with our refinement of the QES prescription, the number of bit threads that can pass through these Planckian wormholes should be controlled by the conditional min- and max-entropy, not the von Neumann entropy.

It would also be interesting to incorporate zero-bits into this framework, allowing bit threads to more precisely depict the total flow of information in AdS/CFT.

Other future work. We have not given a direct path integral argument for these QES refinements for general bulk states. Our argument was more indirect: we proved it for RTN using linear algebra. Because the RTN entropy can be computed using the replica trick, the replica trick must enforce these refinements. The RTN replica trick is identical to the fixed area state replica trick, and so the same results must be true in fixed-area states. More typical (non fixed-area) states have the same entropy as the average of fixed area states that comprise them, plus subleading corrections. Although we think this argument is compelling, it is very indirect. There should be some way to relate bulk min- and max-entropy to the holographic calculation, allowing a direct replica trick proof of our result. In particular, we should be able to directly see why they are the quantities that determine whether the LM assumption is valid.

These results should also be generalized to von Neumann algebras. We have discussed subregions (b, b' , etc.) instead of subalgebras, only for simplicity. The smooth conditional min- and max-entropy admit algebraic definitions, which is a better language for bulk reconstruction.

We also didn't give any description as convenient as the naïve QES prescription when $H_{\min}^\varepsilon < \Delta A/4G < H_{\max}^\varepsilon$. We did provide useful *bounds*, for example that the entropy of a state in that regime is less than the average of the entropies of any mixture comprising that state, plus $\mathcal{O}(\ln d)$, where d is the number of states in the mixture. But getting something stronger, such as an explicit formula, may be too much to hope for. Any formula would need to encode the details of the entanglement structure of the mixed state $\rho_{bb'}$. This is known to be very hard to characterize.

Acknowledgments

It's a pleasure to thank Netta Engelhardt, Daniel Harlow, Patrick Hayden, Tom Faulkner, Pratik Rath, and Arvin Shahbazi-Moghaddam for discussions. CA is supported by the US Department of Energy grants DE-SC0018944 and DE-SC0019127, and also the Simons foundation as a member of the It from Qubit collaboration. GP is supported in part by AFOSR award FA9550-16-1-0082 and DOE award DE-SC0019380.

A Detailed evaluation of the mixture resolvent

Here we elaborate on the calculations in section 4.3, detailing how to go from the cubic resolvent

$$\lambda R = e^{A_2/4G} + \frac{pR}{e^{A_2/4G} - \frac{p\lambda_1}{e^{A_1/4G}}R} + \frac{(1-p)R}{e^{A_2/4G} - \frac{(1-p)\lambda_2}{e^{A_1/4G}}R}, \tag{A.1}$$

to the eigenvalues in each regime. Note the differences between this resolvent and (4.30). Here, the bulk state is

$$\rho_{b'} = \begin{pmatrix} p\lambda_1 \mathbb{1}_{\frac{1}{\lambda_1}} & 0 \\ 0 & (1-p)\lambda_2 \mathbb{1}_{\frac{1}{\lambda_2}} \end{pmatrix}. \tag{A.2}$$

This is a slight generalization of the state from section 4.3, in that we don't require one state to be pure. However, we still require both to have flat spectra. We recover the state from section 4.3 by setting $\lambda_1 = 1$ and $\lambda_2 = e^{-S}$.

Recall that we assumed, for simplicity,

$$(1-p)\lambda_2 \ll p\lambda_1, \quad A_1 \ll A_2. \quad (\text{A.3})$$

This condition ensures that our small and large R expansions have overlapping regimes of validity. Unlike in section 4.3, we will not assume that $p, 1-p = \mathcal{O}(1)$. This will require us to introduce a third expansion that is valid for sufficiently small R and very small p .

Here are the three expansions we use, plus details about their associated spectra, along with information that will be useful in evaluating their regime of validity. These details are computed with the help of appendix B.

The expansions.

Expansion 1. Consider the large R expansion

$$\lambda R = e^{A_2/4G} - \frac{e^{A_1/4G}}{\lambda_1} + \frac{(1-p)R}{e^{A_2/4G} - \frac{(1-p)\lambda_2}{e^{A_1/4G}} R} + \mathcal{O}\left(\frac{e^{A_1/4G}}{\lambda_1} \frac{e^{(A_1+A_2)/4G}}{p\lambda_1 R}\right). \quad (\text{A.4})$$

Using the results of appendix B, this leads to

$$\text{Number of eigenvalues} = \begin{cases} e^{A_2/4G} - \frac{e^{A_1/4G}}{\lambda_1}, & \frac{1}{\lambda_2} \gg e^{(A_2-A_1)/4G} \gg \frac{1}{\lambda_1} \\ \frac{e^{A_1/4G}}{\lambda_2}, & \frac{1}{\lambda_2} \ll e^{(A_2-A_1)/4G} \gg \frac{1}{\lambda_1} \end{cases} \quad (\text{A.5})$$

of average size

$$\lambda_{\text{avg}} = \begin{cases} (1-p)e^{-A_2/4G}, & \frac{1}{\lambda_2} \gg e^{(A_2-A_1)/4G} \\ (1-p)\lambda_2 e^{-A_1/4G} \left(1 - \frac{e^{-(A_2-A_1)/4G}}{\lambda_1}\right), & \frac{1}{\lambda_2} \ll e^{(A_2-A_1)/4G}. \end{cases} \quad (\text{A.6})$$

To analyze when this expansion is valid, it is useful to know the value of the resolvent. At λ_{avg} , the resolvent is

$$R(\lambda_{\text{avg}}) = \begin{cases} -i \frac{e^{(A_1+A_2)/4G}}{(1-p)\lambda_2} \sqrt{\lambda_2 e^{(A_2-A_1)/4G} - \frac{\lambda_2}{\lambda_1} + \frac{e^{2A_2/4G} - e^{(A_1+A_2)/4G}/\lambda_1}{2(1-p)}} + \dots, & \frac{1}{\lambda_2} \gg e^{(A_2-A_1)/4G} \\ \frac{e^{(A_1+A_2)/4G}}{(1-p)\lambda_2} \left(1 - i \left(\lambda_2 e^{(A_2-A_1)/4G} - \frac{\lambda_2}{\lambda_1}\right)^{-1/2} + \dots\right), & \frac{1}{\lambda_2} \ll e^{(A_2-A_1)/4G}. \end{cases} \quad (\text{A.7})$$

Expansion 2. Consider the small R expansion

$$\lambda R = e^{A_2/4G} + \frac{pR}{e^{A_2/4G} - \frac{p\lambda_1}{e^{A_1/4G}} R} + \frac{(1-p)R}{e^{A_2/4G}} + \mathcal{O}\left(\frac{(1-p)R}{e^{A_2/4G}} \frac{(1-p)\lambda_2 R}{e^{(A_1+A_2)/4G}}\right). \quad (\text{A.8})$$

This results in

$$\text{Number of eigenvalues} = \begin{cases} e^{A_2/4G}, & \frac{1}{\lambda_1} \gg e^{(A_2-A_1)/4G} \\ \frac{e^{A_1/4G}}{\lambda_1}, & \frac{1}{\lambda_1} \ll e^{(A_2-A_1)/4G} \end{cases} \quad (\text{A.9})$$

of average size

$$\lambda_{\text{avg}} = \begin{cases} e^{-A_2/4G}, & \frac{1}{\lambda_1} \gg e^{(A_2-A_1)/4G} \\ p\lambda_1 e^{-A_1/4G} + (1-p)e^{-A_2/4G}, & \frac{1}{\lambda_1} \ll e^{(A_2-A_1)/4G}. \end{cases} \quad (\text{A.10})$$

At λ_{avg} , the resolvent is

$$R(\lambda_{\text{avg}}) = \begin{cases} -i \frac{e^{(A_1+A_2)/4G}}{p\lambda_1} \sqrt{\lambda_1 e^{(A_2-A_1)/4G} + \frac{e^{2A_2/4G}}{2p}} + \dots, & \frac{1}{\lambda_1} \gg e^{(A_2-A_1)/4G} \\ \frac{e^{(A_1+A_2)/4G}}{p\lambda_1} \left(1 - i \left(\lambda_1 e^{(A_2-A_1)/4G} \right)^{-1/2} + \dots \right), & \frac{1}{\lambda_1} \ll e^{(A_2-A_1)/4G}. \end{cases} \quad (\text{A.11})$$

For very small p , the second term on the right hand side of (A.8) can become smaller than the terms that were dropped. It is therefore helpful to use a slightly adapted version of Expansion 2, namely

$$\lambda R = e^{A_2/4G} + \frac{pR}{e^{A_2/4G} - \frac{p\lambda_1}{e^{A_1/4G}} R} + \frac{e^{-A_2/4G}(1-p)R}{1 - \frac{(1-p)\lambda_2}{p\lambda_1}} + \mathcal{O}\left(\frac{(1-p)^2\lambda_2 R \left(R - \frac{e^{(A_1+A_2)/4G}}{p\lambda_1}\right)}{e^{(A_1+2A_2)/4G}}\right). \quad (\text{A.12})$$

The only effect of this change is that now

$$\lambda_{\text{avg}} = p\lambda_1 e^{-A_1/4G} + \frac{(1-p)e^{-A_2/4G}}{1 - \frac{(1-p)\lambda_2}{p\lambda_1}}. \quad (\text{A.13})$$

Finally, we note that for $1/\lambda_1 \ll e^{(A_2-A_1)/4G}$, and for values of λ where $D(\lambda) \neq 0$, we have

$$R - \frac{e^{(A_1+A_2)/4G}}{p\lambda_1} \leq \mathcal{O}\left(\frac{e^{(A_1+A_2)/8G}}{\lambda_1^{1/2}}\right). \quad (\text{A.14})$$

This will again be important when considering small values of p .

Expansion 3. Finally, we can use an alternative small R expansion, where we expand *both* the λ_1 and λ_2 terms up to $\mathcal{O}(R^2)$,

$$\lambda R = e^{A_2/4G} + \frac{R}{e^{A_2/4G}} + \frac{p^2\lambda_1 + (1-p)^2\lambda_2}{e^{(A_1+2A_2)/4G}} R^2 + \mathcal{O}\left(\frac{(p^3\lambda_1^2 + (1-p)^3\lambda_2^2)R^3}{e^{(2A_1+3A_2)/4G}}\right). \quad (\text{A.15})$$

This results in

$$D(\lambda) = \frac{e^{(A_1+2A_2)/4G}}{2\pi(p^2\lambda_1 + (1-p)^2\lambda_2)} \sqrt{\frac{p^2\lambda_1 + (1-p)^2\lambda_2}{e^{(A_1+A_2)/4G}} - (\lambda - e^{A_2/4G})^2}, \quad (\text{A.16})$$

which gives $e^{A_2/4G}$ eigenvalues with average eigenvalue $\lambda_{\text{avg}} = e^{-A_2/4G}$. Finally,

$$R(\lambda_{\text{avg}}) = ie^{A_2/4G} \sqrt{\frac{e^{(A_1+A_2)/4G}}{p^2\lambda_1 + (1-p)^2\lambda_2}}. \quad (\text{A.17})$$

This expansion is important because when p is very small, the $\mathcal{O}(R^2)$ correction from the λ_2 term may be larger than the corresponding correction from the λ_1 term, even though $p\lambda_1 \gg (1-p)\lambda_2$.

The regimes. Here are the three regimes, each defined by the relative size of $\Delta A/4G \equiv (A_2 - A_1)/4G$ and

$$\begin{aligned} H_{\min}^\varepsilon(b') &= \ln\left(\frac{1}{p\lambda_1}\right), \\ H_{\max}^\varepsilon(b') &\approx \ln\left(\frac{1}{\lambda_2}\right). \end{aligned} \tag{A.18}$$

There are corrections to the naïve QES prescription only in Regime 2, when $H_{\min}^\varepsilon(b')$ and $H_{\max}^\varepsilon(b')$ are on different sides of $\Delta A/4G$.

Regime 1: $H_{\min}^\varepsilon, H_{\max}^\varepsilon \ll \Delta A/4G$. In this regime, Expansion 1 is always valid at its eigenvalue peak, which is at $\lambda_{\text{avg}} = p\lambda_2 e^{-A_1/4G}$. Expansion 2 is valid at its eigenvalue peak, with $\lambda_{\text{avg}} = (1-p)\lambda_1 e^{-A_1/4G}$, unless p is very small, in which case we need to use the adapted version of Expansion 2. This only has a small effect on the eigenvalue peak. Thus for all parameter values, assuming (A.3), the entropy is given by

$$S(B) = \frac{A_1}{4G} + p \ln\left(\frac{1}{p \ln(\lambda_1)}\right) + (1-p) \ln\left(\frac{1}{(1-p) \ln(\lambda_2)}\right) + \dots, \tag{A.19}$$

where we have suppressed terms that vanish in the limits we've taken. The naïve quantum extremal surface prescription gives the correct answer.

Proof. Consider Expansion 1. The resolvent evaluated at $\mathcal{O}(\lambda_{\text{avg}})$ is approximately

$$\mathcal{O}(R(\lambda_{\text{avg}})) = \mathcal{O}\left(\frac{e^{(A_1+A_2)/4G}}{(1-p)\lambda_2}\right).$$

Therefore the dropped terms at $\lambda = \mathcal{O}(\lambda_{\text{avg}})$ are $\mathcal{O}((1-p)\lambda_2 e^{A_1/4G}/p\lambda_1^2)$. The smallest kept term is $\mathcal{O}(e^{A_1/4G}/\lambda_1)$. Therefore the ratio dropped/kept is $\mathcal{O}((1-p)\lambda_2/p\lambda_1)$, which is small given our choice (A.3).

Now consider Expansion 2. The resolvent at $\lambda = \mathcal{O}(\lambda_{\text{avg}})$ is

$$\mathcal{O}(R(\lambda_{\text{avg}})) = \mathcal{O}\left(\frac{e^{(A_1+A_2)/4G}}{p\lambda_1}\right).$$

The largest dropped term, at $\mathcal{O}(\lambda_{\text{avg}})$, is given by plugging this into the dropped term in (A.4). The smallest kept term is either the second term, with size $\mathcal{O}(e^{A_1/4G}/\lambda_1)$, or the third term with size $\mathcal{O}((1-p)e^{A_1/4G}/p\lambda_1)$. In the latter case, the ratio dropped/kept equals $\mathcal{O}((1-p)\lambda_2/p\lambda_1)$. This is small given (A.3). In the former case, the ratio equals $\mathcal{O}((1-p)^2\lambda_2/p^2\lambda_1)$. This is small, unless p itself is very small.

For small p , we need to be a bit more careful, recognizing that the second term in Expansion 2 only becomes important near the eigenvalue peak where its denominator is small, and also to make use of the adapted version (A.12) of Expansion 2. Using this adapted version, we find that the ratio of the dropped term to the second term is

$$\mathcal{O}\left(\frac{(1-p)^2\lambda_2\lambda_1\left(R - \frac{e^{(A_1+A_2)/4G}}{p\lambda_1}\right)^2}{e^{(2A_1+2A_2)/4G}}\right) = \mathcal{O}\left(\frac{(1-p)^2\lambda_2 e^{(A_1-A_2)/4G}}{p^2\lambda_1^2}\right). \tag{A.20}$$

Going from the left hand side to the right hand side, we used the fact that (A.14) holds near the eigenvalue peak. This ratio is small so long as $(1-p)\lambda_2 \ll p\lambda_1$ and $p\lambda_1 \gg e^{(A_1-A_2)/4G}$.

It is now a simple matter to compute the entropy using the eigenvalues from Expansion 1 and (the adapted) Expansion 2 to get (A.19). \square

Regime 2: $H_{\min}^\varepsilon \ll \Delta A/4G \ll H_{\max}^\varepsilon$. This is the regime in which there are large corrections to the naïve QES prescription. Again, Expansions 1 and 2 are valid at their eigenvalue peak, unless p is small where we need to use the adapted version of Expansion 2. However, Expansion 1 now gives an eigenvalue peak at $\lambda_{\text{avg}} = (1-p)e^{-A_2/4G}$

Thus, assuming (A.3):

$$S(B) = p \frac{A_1}{4G} + p \ln \left(\frac{1}{p \ln(\lambda_1)} \right) + (1-p) \frac{A_2}{4G} + (1-p) \ln \left(\frac{1}{1-p} \right) + \dots, \quad (\text{A.21})$$

and again we have dropped terms that vanish in the limits we've taken.

Proof. Expansion 2 works identically to Regime 1, so we only consider Expansion 1. The largest dropped term for $\lambda = \mathcal{O}(\lambda_{\text{avg}})$ is $\mathcal{O}((1-p)\lambda_2 e^{A_1/4G} / p\lambda_1^2 \sqrt{\lambda_2 e^{(A_2-A_1)/4G}})$. The smallest kept term is $e^{A_1/4G} / \lambda_1$. The ratio dropped/kept is

$$\mathcal{O} \left(\frac{(1-p)\sqrt{\lambda_2}}{p\lambda_1 e^{(A_2-A_1)/4G}} \right).$$

This is small so long as $(1-p)\lambda_2 \ll p\lambda_1$ and $p\lambda_1 \gg e^{(A_1-A_2)/4G}$. \square

Regime 3: $\Delta A/4G \ll H_{\min}^\varepsilon, H_{\max}^\varepsilon$. In this regime, the naïve QES prescription does not receive large corrections. This regime is interesting because it requires $p\lambda_1 e^{(A_2-A_1)/4G} \ll 1$, which can be achieved whether or not $\lambda_1 e^{(A_2-A_1)/4G}$ is greater or less than 1. If $\lambda_1 e^{(A_2-A_1)/4G} \gg 1$, then Expansions 1 and 2 are valid, so long as $p^2 \lambda_1^2 e^{(A_2-A_1)/4G} \gg (1-p)^2 \lambda_2$. However the entropy calculation gives a different answer:

$$S(B) \approx \frac{A_2}{4G} + \dots, \quad (\text{A.22})$$

where “...” represents terms that vanish in the limits we've taken. We get the same answer in a different way if $\lambda_1 e^{(A_2-A_1)/4G} \ll 1$. In this parameter range, Expansion 1 is never valid. Expansion 2 is valid at its eigenvalue peak, and gives (A.22), so long as p is large enough such that $(1-p)^2 \lambda_2 < p^2 \lambda_1$. When p is smaller than that, Expansion 2 cannot be used.

The alternative small R expansion, Expansion 3, is valid for all values of p in this regime and gives (A.22).

Proof. Consider Expansion 1. If $\lambda_1 e^{(A_2-A_1)/4G} \ll 1$, the resolvent is always real and so does not contribute any eigenvalues.²⁷ If $\lambda_1 e^{(A_2-A_1)/4G} \gg 1$, then Expansion 1 works exactly as

²⁷In fact, the solution given by Expansion 1 for large values of R does not actually appear as the resolvent for any value of λ . It gives a different sheet of the solution to the one given by the resolvent.

it did in Regime 2. The smallest kept term is $e^{A_1/4G}/\lambda_1$. The largest dropped term is $\mathcal{O}((1-p)\sqrt{\lambda_2}e^{A_1/4G}/p\lambda_1^2\sqrt{e^{(A_2-A_1)/4G}})$. The ratio dropped/kept is $\mathcal{O}((1-p)\sqrt{\lambda_2}/p\lambda_1\sqrt{e^{(A_2-A_1)/4G}})$, which is small so long as $p^2\lambda_1^2e^{(A_2-A_1)/4G} \gg (1-p)^2\lambda_2$.

Consider Expansion 2. If $\lambda_1e^{(A_2-A_1)/4G} \gg 1$, Expansion 2 works exactly as it did in Regimes 1 and 2. It is therefore valid (when using the adapted version) so long as $p^2\lambda_1^2e^{(A_2-A_1)/4G} \gg (1-p)^2\lambda_2$.

If $\lambda_1e^{(A_2-A_1)/4G} \ll 1$, the largest dropped term in Expansion 2 is $\mathcal{O}((1-p)^2\lambda_2e^{A_2/4G}/p^2\lambda_1)$, while the smallest kept term is either $e^{A_2/4G}$ or $(1-p)R/e^{A_2/4G}=(1-p)\sqrt{e^{(A_2+A_1)/4G}}/p\sqrt{\lambda_1}$. In the former case, the dropped/kept ratio is $\mathcal{O}((1-p)^2\lambda_2/p^2\lambda_1)$, which is small unless $p^2\lambda_1 \lesssim (1-p)^2\lambda_2$. In the latter case, we find that the dropped/kept ratio is $\mathcal{O}((1-p)\lambda_2\sqrt{\lambda_1}e^{(A_2-A_1)/4G}/p\sqrt{\lambda_1})$. This is always small.

What about Expansion 3? The smallest term that we keep is $\mathcal{O}(e^{A_2/4G})$, while the largest term that we drop is $\mathcal{O}(\sqrt{e^{(3A_2-A_1)/4G}}(p^3\lambda_1^2 + (1-p)^3\lambda_2^2)/(p^2\lambda_1 + (1-p)^2\lambda_2)^3)$. The ratio is small so long as $\lambda_1 \ll e^{(A_1-A_2)/4G}$ or $p^2\lambda_1 \ll (1-p)^2\lambda_2$. Between the expansions, we can therefore cover all the possible regimes.

Note that for $p^2\lambda_1 \ll (1-p)^2\lambda_2 \ll p^2\lambda_1^2e^{(A_2-A_1)/4G}$, both Expansions 1 and 2, and Expansion 3 are valid. However, Expansion 3 misses the existence of the second eigenvalue peak that appears in Expansion 2, even though it is a small R expansion and this occurs at smaller R than the main eigenvalue peak. This is because, for these intermediate values of p , the Taylor expansion of the λ_1 term in Expansion 3 was not under control, since $p\lambda_1R \gg e^{(A_1+A_2)/4G}$. We were only able to get away with the expansion anyway because both the true λ_1 term and our approximation of it were only small correction anyway (because p was so small). Near the second eigenvalue peak itself, this isn't true because the true λ_1 term breaks down, and Expansion 3 breaks down. So we do need to use Expansion 2 here.

We compute the entropy as follows. If $\lambda_1e^{(A_2-A_1)/4G} \gg 1$, then for $p^2\lambda_1^2e^{(A_2-A_1)/4G} \gg (1-p)^2\lambda_2$ we can use Expansions 1 and 2 to compute the entropy. For small p , we instead use Expansion 3. If $\lambda_1e^{(A_2-A_1)/4G} \ll 1$, then we can always just use Expansion 3, although we can also use Expansion 2 if p is not too small. In all cases, the entropy is given by (A.22). \square

B Solving the quadratic resolvent

This appendix studies the quadratic resolvent equation,

$$(\lambda - W)R = X + \frac{R}{Y - ZR}, \tag{B.1}$$

where W, X, Y, Z are some fixed real numbers. This equation has the solutions

$$R(\lambda) = \frac{XZ + Y(\lambda - W) - 1}{2Z(\lambda - W)} - \frac{Y}{2Z(\lambda - W)} \sqrt{\lambda - \lambda_+} \sqrt{\lambda - \lambda_-}, \tag{B.2}$$

where

$$\lambda_{\pm} = \frac{(1 \pm \sqrt{XZ})^2}{Y} + W. \tag{B.3}$$

The minus in front of the square root in (B.2) is required by $R(\lambda \rightarrow \infty) = 0$.

Eigenvalues. The density of eigenvalues $D(\lambda)$ is related by the formula

$$D(\lambda) = -\frac{1}{\pi} \lim_{\epsilon \rightarrow 0^+} \text{Im} R(x + i\epsilon). \quad (\text{B.4})$$

So, we need the imaginary part of the resolvent. We can ignore everything not under the square root, because it will not contribute to $D(\lambda)$. Use the handy fact that the square root (with positive real part) of a complex number $a + ib$ can be written as $\sqrt{a + ib} = p + iq$ with

$$p = \frac{1}{\sqrt{2}} \sqrt{\sqrt{a^2 + b^2} + a}, \quad q = \frac{\text{sign}(b)}{\sqrt{2}} \sqrt{\sqrt{a^2 + b^2} - a}. \quad (\text{B.5})$$

The relevant piece of the imaginary part of R gives

$$D(\lambda) = \frac{Y}{2\pi Z(\lambda - W)} \sqrt{\lambda_+ - \lambda} \sqrt{\lambda - \lambda_-}, \quad (\text{B.6})$$

for $\lambda \in [\lambda_-, \lambda_+]$, and $D(\lambda) = 0$ otherwise. To find the number of eigenvalues, we can integrate this using

$$\int_a^b dx \frac{\sqrt{(x-a)(b-x)}}{2\pi x} = \frac{1}{4} (-2\sqrt{ab} + a + b). \quad (\text{B.7})$$

This gives a total number of eigenvalues

$$\text{Number} = \int_{\lambda_-}^{\lambda_+} d\lambda D(\lambda) = \begin{cases} X, & XZ < 1 \\ \frac{1}{Z}, & XZ > 1. \end{cases} \quad (\text{B.8})$$

The average value of λ is

$$\lambda_{\text{avg}} = \frac{1}{\text{Number}} \int_{\lambda_-}^{\lambda_+} d\lambda D(\lambda) \lambda = \begin{cases} \frac{1}{Y} + W, & XZ < 1 \\ \frac{XZ}{Y} + W, & XZ > 1. \end{cases} \quad (\text{B.9})$$

Entropy. In principle we can compute the von Neumann entropy given a density of eigenvalues $D(\lambda)$ with

$$S = - \int_0^1 \lambda \ln(\lambda) D(\lambda) d\lambda. \quad (\text{B.10})$$

The \ln makes this difficult to evaluate in practice. Fortunately, we can obtain a rather good approximation by expanding λ around the average of the eigenvalue distribution. Use

$$\int_{\lambda_-}^{\lambda_+} d\lambda D(\lambda) = \begin{cases} X, & XZ < 1 \\ \frac{1}{Z}, & XZ > 1, \end{cases} \quad (\text{B.11})$$

$$\int_{\lambda_-}^{\lambda_+} d\lambda D(\lambda) (\lambda - \lambda_{\text{avg}}) = 0, \quad (\text{B.12})$$

$$\int_{\lambda_-}^{\lambda_+} d\lambda D(\lambda) (\lambda - \lambda_{\text{avg}})^2 = \begin{cases} \frac{X^2 Z}{Y^2}, & XZ < 1 \\ \frac{X}{Y^2}, & XZ > 1. \end{cases} \quad (\text{B.13})$$

The entropy is

$$\begin{aligned}
 S &\approx \int_{\lambda_-}^{\lambda_+} d\lambda D(\lambda) \left(\lambda_{\text{avg}} \ln \left(\frac{1}{\lambda_{\text{avg}}} \right) + (\lambda - \lambda_{\text{avg}}) \left(\ln \left(\frac{1}{\lambda_{\text{avg}}} \right) - 1 \right) - \frac{1}{2\lambda_{\text{avg}}} (\lambda - \lambda_{\text{avg}})^2 + \dots \right) \\
 &= \begin{cases} -X \left(\frac{1}{Y} + W \right) \ln \left(\frac{1}{Y} + W \right) - \frac{1}{2} \left(\frac{1}{\frac{1}{Y} + W} \right) \left(\frac{X^2 Z}{Y^2} \right), & XZ < 1 \\ - \left(\frac{X}{Y} + \frac{W}{Z} \right) \ln \left(\frac{XZ}{Y} + W \right) - \frac{1}{2} \left(\frac{1}{\frac{XZ}{Y} + W} \right) \left(\frac{X}{Y^2} \right), & XZ > 1. \end{cases} \quad (\text{B.14})
 \end{aligned}$$

Resolvent values. We are sometimes interested in evaluating the resolvent at the location of the average eigenvalue. This is important in determining that the expansions used in section 4 and appendix A provide accurate estimates of the eigenvalues associated to the cubic resolvent (4.30).

Plug (B.9) into (B.2) to get

$$\begin{aligned}
 R(\lambda = \lambda_{\text{avg}}) &= \begin{cases} \frac{Y}{2Z} \left(XZ - \sqrt{XZ(XZ - 4)} \right), & XZ < 1 \\ -\frac{Y}{2XZ^2} \left(1 - 2XZ + \sqrt{1 - 4XZ} \right), & XZ > 1 \end{cases} \quad (\text{B.15}) \\
 &\approx \begin{cases} -i\frac{Y}{Z} \sqrt{XZ} + \frac{XY}{2} + \dots & XZ \ll 1 \\ \frac{Y}{Z} - i\frac{Y}{Z} \frac{1}{\sqrt{XZ}} + \dots & XZ \gg 1. \end{cases}
 \end{aligned}$$

This implies

$$\frac{R(\lambda_{\text{avg}})}{Y - ZR(\lambda_{\text{avg}})} \approx -i\sqrt{\frac{X}{Z}}, \quad (\text{B.16})$$

when either $XZ \ll 1$ or $XZ \gg 1$. Also note

$$\lambda_{\text{avg}} R(\lambda_{\text{avg}}) = \begin{cases} -i\sqrt{\frac{X}{Z}} (1 + WY) + \dots & XZ \ll 1 \\ X + \frac{WY}{Z} + \dots & XZ \gg 1. \end{cases} \quad (\text{B.17})$$

These are useful when comparing dropped terms to kept ones.

Example: bipartite tensor. Apply this to a simple example. Consider a bipartite random tensor, with legs A and B . The resolvent associated to ρ_A satisfies

$$\lambda R = D_A + \frac{R}{D_A - \frac{R}{D_B}}. \quad (\text{B.18})$$

So, $W = 0$, $X = D_A$, $Y = D_A$, and $Z = 1/D_B$. There are $\min(D_A, D_B)$ eigenvalues with average value $\lambda_{\text{avg}} = \max(\frac{1}{D_A}, \frac{1}{D_B})$. The ratio of the width of the peak to λ_{avg} is approximately $\min(D_A, D_B)/\sqrt{D_A D_B}$, and so the peak is very narrow when the dimensions are quite different (width relative to mean like $1/\sqrt{D_{\text{larger}}}$) and widest when the dimensions are equal. The entropy is

$$S_{\text{bip}} = \begin{cases} \ln(D_A) - \frac{1}{2} \frac{D_A}{D_B}, & \frac{D_A}{D_B} < 1 \\ \ln(D_B) - \frac{1}{2} \frac{D_B}{D_A}, & \frac{D_A}{D_B} > 1. \end{cases} \quad (\text{B.19})$$

This is all just as expected.

C Numerics

Here we present numerical evidence supporting the results of section 4. These numerics are of a single tripartite random tensor, with legs B , \bar{B} , b' . As pointed out in [9], computing the entropy of e.g. B is equivalent non-perturbatively to the computation of $S(B)$ in a fixed-area state, like figure 4 with γ_1 fixed to area $\ln D_B$ and γ_2 fixed to area $\ln D_{\bar{B}}$. D_B and $D_{\bar{B}}$ are the dimensions of legs B and \bar{B} respectively. The leg b' is the “bulk” leg, analogous to the state of the bulk fields between γ_1 and γ_2 , and is projected into the state

$$\rho_{b'} = \begin{pmatrix} p\lambda_1 \mathbb{1}_{\frac{1}{\lambda_1}} & 0 \\ 0 & (1-p)\lambda_2 \mathbb{1}_{\frac{1}{\lambda_2}} \end{pmatrix}. \tag{C.1}$$

Figure 9 displays the resulting eigenvalue density $D(\lambda)$ of density matrix ρ_B , in Regimes 1, 2, and 3 from section 4 and also appendix A.

D One-shot decoupling

A great many facts in quantum information theory follow from the same basic principle: the decoupling theorem.

While this powerful theorem was originally proven and used in the independent identically distributed (i.i.d.) setting [39], in which a large number of independent copies of the state are available, more recently a one-shot version has been proven [36], effectively generalizing many key results to the one-shot setting. The chief difference between the two decoupling theorems is the replacement of the von Neumann entropy with the one-shot entropies, the min- and max-entropy.

The setup is as follows. Consider a system $A = A_1 A_2$ (with dimensions $|A_1|$ and $|A_2|$) entangled with a system R . The one-shot decoupling theorem provides a sufficient condition for the average unitary U acting on A to “decouple” A_1 and R . This is often used to provide a sufficient condition for something weaker, the existence of a unitary operator U that decouples A_1 and R .

For our purposes, the theorem says if

$$\ln |A_1| \leq \ln |A_2| + H_{\min}(A|R) - 2 \ln \frac{1}{\varepsilon}, \tag{D.1}$$

then

$$\int dU \left\| \text{tr}_{A_2}(U \rho_{AR} U^\dagger) - \frac{\mathbb{1}_{A_1}}{|A_1|} \otimes \rho_R \right\|_1 \leq \varepsilon, \tag{D.2}$$

where dU is the Haar measure on the group of unitaries acting on \mathcal{H}_A , normalized to $\int dU = 1$.

We present the proof as Theorem 7 below, after some useful definitions and lemmas.

Definition 1. Let X be an operator on Hilbert space \mathcal{H} . The L^2 norm, or *Hilbert-Schmidt norm*, is defined as

$$\|X\|_2 = \sqrt{\text{tr}(X^\dagger X)}. \tag{D.3}$$

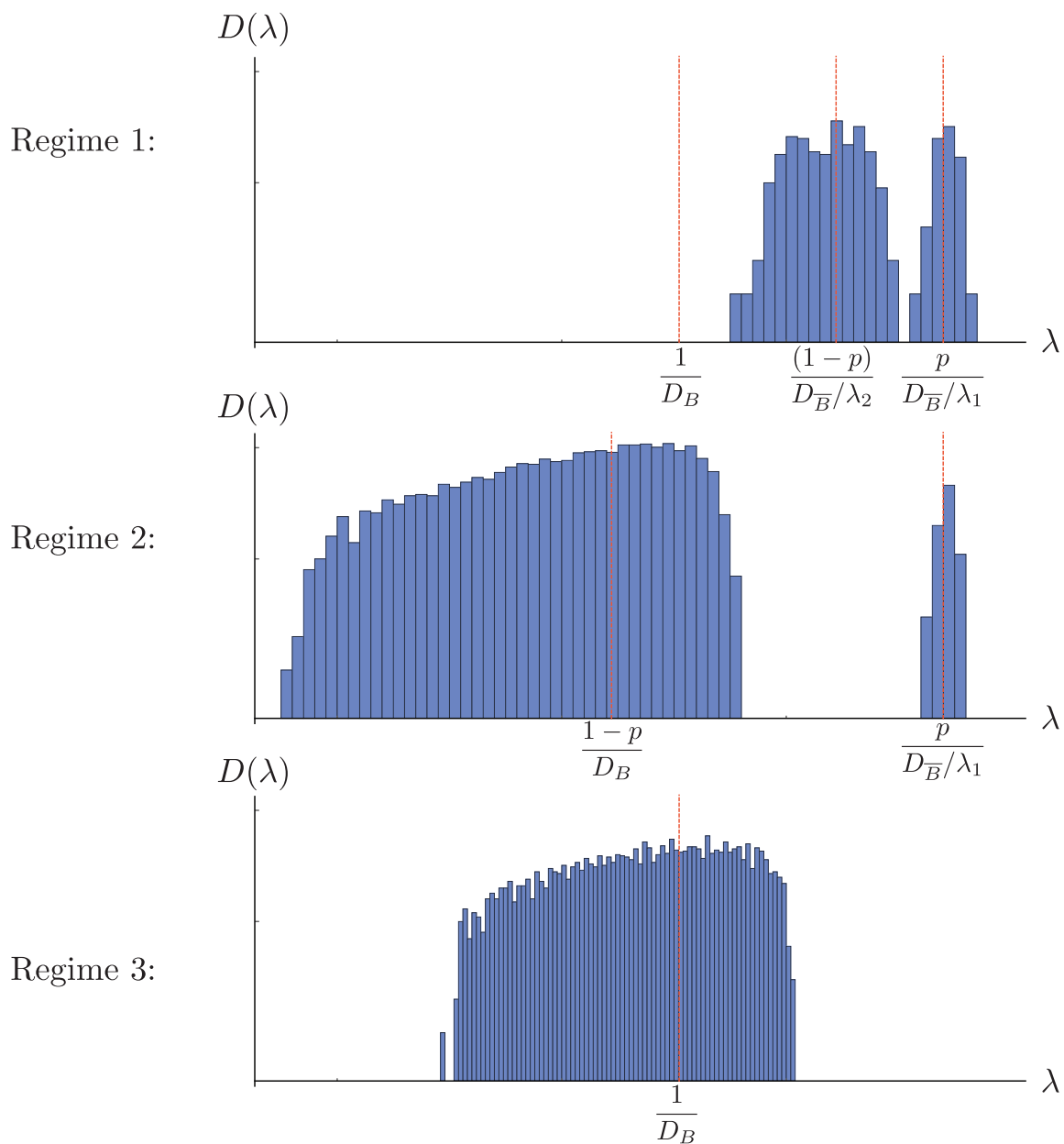


Figure 9. Log-log histogram of the eigenvalue density $D(\lambda)$ of ρ_B , for a random tensor with two “boundary” legs B, \bar{B} and a “bulk” leg b' , for 80 trials. All plots are made with boundary leg dimensions $D_B = 30$, $D_{\bar{B}} = 1$ and bulk leg state (C.1), with $p = 1/2$. Regime 1 plot made with $\lambda_1 = 1$, $\lambda_2 = 1/3$; Regime 2 plot $\lambda_1 = 1$, $\lambda_2 = 1/45$; Regime 3 plot $\lambda_1 = 1/20$, $\lambda_2 = 1/45$. Note the agreement with figure 5.

This upperbounds the L^1 norm ($\|X\|_1 = \text{tr} \sqrt{X^\dagger X}$), as $\|X\|_1 \leq \sqrt{d} \|X\|_2$, where d is the dimension of \mathcal{H} . This bound is involved in the i.i.d. proof of decoupling [39], but the one-shot version we are interested in requires a stronger bound.

Lemma 2 (Lemma 5.1.3 of [7]). Let S be a Hermitian operator on Hilbert space \mathcal{H} , and σ be a nonnegative operator on \mathcal{H} . Then

$$\|S\|_1 \leq \sqrt{\text{tr} \sigma} \left\| \sigma^{-1/4} S \sigma^{-1/4} \right\|_2. \tag{D.4}$$

Proof. We first note that (D.4) can be rewritten as

$$\text{tr} \left| \sqrt{\sigma'} S' \sqrt{\sigma'} \right| \leq \sqrt{\text{tr}(S'^2) \text{tr}(\sigma'^2)}, \tag{D.5}$$

where $\sigma' = \sqrt{\sigma}$ and $S' = \sigma^{-1/4} S \sigma^{-1/4}$. Let $|v\rangle$ denote an orthonormal eigenbasis of $S = \sigma'^{1/2} S' \sigma'^{1/2}$, and let $S' = \sum_y \alpha_y |y\rangle \langle y|$ be a spectral decomposition of S' . Then

$$\text{tr} \left| \sqrt{\sigma'} S' \sqrt{\sigma'} \right| = \sum_v \left| \langle v | \sqrt{\sigma'} S' \sqrt{\sigma'} | v \rangle \right| \tag{D.6}$$

$$= \sum_v \left| \sum_y \alpha_y \langle v | \sqrt{\sigma'} | y \rangle \langle y | \sqrt{\sigma'} | v \rangle \right| \tag{D.7}$$

$$\leq \sum_v \sum_y |\alpha_y| \langle v | \sqrt{\sigma'} | y \rangle \langle y | \sqrt{\sigma'} | v \rangle \tag{D.8}$$

$$= \sum_v \langle v | \sqrt{\sigma'} | S' | \sqrt{\sigma'} | v \rangle \tag{D.9}$$

$$= \text{tr}(\sqrt{\sigma'} | S' | \sqrt{\sigma'}) = \text{tr}(|S'| \sigma') \tag{D.10}$$

$$\leq \sqrt{\text{tr}(|S'|^2) \text{tr}(\sigma'^2)} = \sqrt{\text{tr}(S'^2) \text{tr}(\sigma'^2)}. \tag{D.11}$$

The second inequality is the Cauchy-Schwarz inequality, applied to the Hilbert space $\text{End}(\mathcal{H})$ of operators on \mathcal{H} , with the inner product $\langle A|B \rangle = \text{tr}(A^\dagger B)$. \square

Having bounded the L^1 norm by this particular L^2 norm, we will later bound the relevant L^2 norm by something else. First, we need to define the conditional collision entropy $H_C(A|B)$, which, as we'll prove, bounds the conditional min-entropy $H_{\min}(A|B)$.

Definition 3. The *quantum conditional collision entropy* for density matrix ρ_{AB} on Hilbert space $\mathcal{H}_{AB} = \mathcal{H}_A \otimes \mathcal{H}_B$ is defined as

$$H_C(A|B) = - \inf_{\sigma_B} \ln \text{tr} \left[\left((\mathbb{1}_A \otimes \sigma_B^{-1/4}) \rho_{AB} (\mathbb{1}_A \otimes \sigma_B^{-1/4}) \right)^2 \right] \tag{D.12}$$

where the infimum is taken over all density matrices σ_B on Hilbert space \mathcal{H}_B . Note that σ_B^{-1} is the ‘generalized inverse’ of σ_B , defined as the inverse on its support. That is, $\sigma_B \sigma_B^{-1} = \sigma_B^{-1} \sigma_B = \sigma_B^0 = (\sigma_B^{-1})^0$.

Lemma 4 (Lemma B.3 of [36]). Let ρ_{AB} be a nonnegative Hermitian matrix on Hilbert space \mathcal{H}_{AB} , with trace less than or equal to 1. Then

$$H_{\min}(A|B) \leq H_C(A|B). \tag{D.13}$$

Proof. By the definition of $H_{\min}(A|B)$ there exists some σ_B such that $H_{\min}(A|B) = -D_{\infty}(\rho_{AB}||\rho_A \otimes \sigma_B)$. By the definition of $D_{\infty}(\rho_{AB}||\rho_A \otimes \sigma_B)$, we have

$$e^{\lambda} \mathbb{1}_A \otimes \sigma_B \geq \rho_{AB}. \quad (\text{D.14})$$

if and only if $\lambda \geq -H_{\min}(A|B)$. Hence the largest eigenvalue of $\sigma_B^{-1/2} \rho_{AB} \sigma_B^{-1/2}$ is $e^{-H_{\min}(A|B)}$. We can rewrite this as

$$H_{\min}(A|B) = -\ln \max_{\omega_{AB}} \text{tr} \left[\omega_{AB} (\mathbb{1}_A \otimes \sigma_B^{-1/2}) \rho_{AB} (\mathbb{1}_A \otimes \sigma_B^{-1/2}) \right], \quad (\text{D.15})$$

where the maximization is over density matrices ω_{AB} and is achieved when ω_{AB} is a projector onto the largest eigenvalue of $\sigma_B^{-1/2} \rho_{AB} \sigma_B^{-1/2}$. For κ_B and ω_{AB} arbitrary density matrices on \mathcal{H}_B and \mathcal{H}_{AB} respectively,

$$\begin{aligned} H_C(A|B) &= -\ln \min_{\kappa_B} \text{tr} \left[\rho_{AB} (\mathbb{1}_A \otimes \kappa_B^{-1/2}) \rho_{AB} (\mathbb{1}_A \otimes \kappa_B^{-1/2}) \right] \\ &\geq -\ln \text{tr} \left[\rho_{AB} (\mathbb{1}_A \otimes \sigma_B^{-1/2}) \rho_{AB} (\mathbb{1}_A \otimes \sigma_B^{-1/2}) \right] \\ &\geq -\ln \max_{\omega_{AB}} \text{tr} \left[\omega_{AB} (\mathbb{1}_A \otimes \sigma_B^{-1/2}) \rho_{AB} (\mathbb{1}_A \otimes \sigma_B^{-1/2}) \right] \\ &= H_{\min}(A|B). \end{aligned} \quad (\text{D.16})$$

□

Lemma 5 (Lemma C.1 of [36]). Let F_{AB} denote the swap operator of $\mathcal{H}_A \otimes \mathcal{H}_B$. Let $A = A_1 A_2$. Then

$$\int dU (U \otimes U)^\dagger (\mathbb{1}_{A_2 A'_2} \otimes F_{A_1 A'_1}) (U \otimes U) \leq \frac{1}{|A_1|} \mathbb{1}_{AA'} + \frac{1}{|A_2|} F_{AA'}, \quad (\text{D.17})$$

where dU is the Haar measure on the space of unitaries acting on \mathcal{H}_A , normalized to $\int dU = 1$.

Proof. For any Hermitian X , it follows from Schur's lemma (see e.g. [35]) that

$$\int dU (U \otimes U)^\dagger X (U \otimes U) = a_+(X) \Pi_A^+ + a_-(X) \Pi_A^-, \quad (\text{D.18})$$

where we have defined

$$\begin{aligned} \Pi_A^\pm &\equiv \frac{1}{2} (\mathbb{1}_{AA'} \pm F_{AA'}), \\ a_\pm(X) &\equiv \frac{1}{\text{rank}(\Pi_A^\pm)} \text{tr}(X \Pi_A^\pm). \end{aligned} \quad (\text{D.19})$$

Plug in $X = (\mathbb{1}_{A_2 A'_2} \otimes F_{A_1 A'_1})$ and find

$$\text{tr} (\Pi_A^\pm (\mathbb{1}_{A_2 A'_2} \otimes F_{A_1 A'_1})) = \frac{1}{2} |A_1| \cdot |A_2|^2 \pm |A_1|^2 \cdot |A_2|. \quad (\text{D.20})$$

Using $\text{rank}(\Pi_A^\pm) = \frac{1}{2} |A| (|A| \pm 1)$, we get

$$a_\pm (\mathbb{1}_{A_2 A'_2} \otimes F_{A_1 A'_1}) = \frac{|A_2| \pm |A_1|}{|A_1| \cdot |A_2| \pm 1}. \quad (\text{D.21})$$

Plugging all of this into (D.18) gives

$$\begin{aligned} \int dU (U \otimes U)^\dagger (\mathbb{1}_{A_2 A_2'} \otimes F_{A_1 A_1'}) (U \otimes U) &= \frac{a_+ + a_-}{2} \mathbb{1}_{AA'} + \frac{a_+ - a_-}{2} F_{AA'} \\ &\leq \frac{1}{|A_1|} \mathbb{1}_{AA'} + \frac{1}{|A_2|} F_{AA'}, \end{aligned} \quad (\text{D.22})$$

which is what we wanted to show. \square

Lemma 6 (Lemma C.2 of [36]). Let ρ_{AR} be a density matrix on \mathcal{H}_{AR} , $A = A_1 A_2$, and $\sigma_{A_1 R}(U) = \text{tr}_{A_2} ((U \otimes \mathbb{1}_R) \rho_{AR} (U \otimes \mathbb{1}_R)^\dagger)$. Then

$$\int dU \text{tr} (\sigma_{A_1 R}(U)^2) \leq \frac{1}{|A_1|} \text{tr}(\rho_R^2) + \frac{1}{|A_2|} \text{tr}(\rho_{AR}^2), \quad (\text{D.23})$$

where dU is the Haar measure on the space of unitaries acting on \mathcal{H}_A , normalized to $\int dU = 1$.

Proof. Use Lemma 5 to get

$$\begin{aligned} &\int dU \text{tr} (\sigma_{A_1 R}(U)^2) \\ &= \int dU \text{tr} ((U_A \otimes U_{A'} \otimes \mathbb{1}_{RR'}) \rho_{A_1 A_2 R} \otimes \rho_{A_1' A_2' R'} (U_A^\dagger \otimes U_{A'}^\dagger \otimes \mathbb{1}_{RR'}) (F_{A_1 A_1'} \otimes \mathbb{1}_{A_2 A_2'} \otimes F_{RR'})) \\ &= \text{tr} \left((\rho_{A_1 A_2 R} \otimes \rho_{A_1' A_2' R'}) \int dU (U \otimes U)^\dagger (\mathbb{1}_{A_2 A_2'} \otimes F_{A_1 A_1'}) (U \otimes U) \otimes F_{RR'} \right) \\ &\leq \text{tr} \left((\rho_{A_1 A_2 R} \otimes \rho_{A_1' A_2' R'}) \left(\frac{1}{|A_1|} \mathbb{1}_{AA'} + \frac{1}{|A_2|} F_{AA'} \right) \otimes F_{RR'} \right) \\ &= \frac{1}{|A_1|} \text{tr}(\rho_R^2) + \frac{1}{|A_2|} \text{tr}(\rho_{AR}^2), \end{aligned} \quad (\text{D.24})$$

which is what we wanted to show. \square

We can finally combine these to prove the one-shot decoupling theorem.

Theorem 7 (Theorem III.1 of [36]). Consider a state ρ_{AR} on Hilbert space $\mathcal{H}_A = \mathcal{H}_{A_1} \otimes \mathcal{H}_{A_2}$, with factors of dimensions $|A_1|$ and $|A_2|$ respectively, and Hilbert space \mathcal{H}_R . Then

$$\ln |A_1| \leq \ln |A_2| + H_{\min}(A|R) - 2 \ln \frac{1}{\varepsilon} \quad (\text{D.25})$$

implies

$$\int dU \left\| \text{tr}_{A_2} (U \rho_{AR} U) - \frac{\mathbb{1}_{A_1}}{|A_1|} \otimes \rho_R \right\|_1 \leq \varepsilon, \quad (\text{D.26})$$

where dU is the Haar measure on the group of unitaries acting on \mathcal{H}_A , normalized to $\int dU = 1$.

Proof. Let $\sigma_{A_1 R}(U) = \text{tr}_{A_2} ((U \otimes \mathbb{1}_R) \rho_{AR} (U \otimes \mathbb{1}_R)^\dagger)$. Because $H_C(A|R) \geq H_{\min}(A|R)$ by Lemma 4, and $\|S\|_1 \leq \sqrt{\text{tr} \sigma} \left\| \sigma^{-1/4} S \sigma^{-1/4} \right\|_2$ by Lemma 2, it suffices to show that

$$\ln |A_1| \leq \ln |A_2| + H_C(A|R) - 2 \ln \frac{1}{\varepsilon} \quad (\text{D.27})$$

implies

$$\int dU \left\| (\mathbb{1}_{A_1} \otimes \omega_R^{-1/4}) \left(\sigma_{A_1 R}(U) - \frac{\mathbb{1}_{A_1}}{|A_1|} \otimes \rho_R \right) (\mathbb{1}_{A_1} \otimes \omega_R^{-1/4}) \right\|_2^2 \leq \frac{\varepsilon^2}{|A_1|}, \quad (\text{D.28})$$

where ω_R is some density matrix on \mathcal{H}_R . Define

$$\begin{aligned} \tilde{\rho}_{AR} &\equiv \left(\mathbb{1}_A \otimes \omega_R^{-1/4} \right) \rho_{AR} \left(\mathbb{1}_A \otimes \omega_R^{-1/4} \right), \\ \tilde{\sigma}_{A_1 R}(U) &\equiv \text{tr}_{A_2} \left((U \otimes \mathbb{1}_R) \tilde{\rho}_{AR} (U \otimes \mathbb{1}_R)^\dagger \right). \end{aligned} \quad (\text{D.29})$$

The left hand side of (D.28) then equals

$$\begin{aligned} \int dU \left\| \tilde{\sigma}_{A_1 R}(U) - \frac{\mathbb{1}_{A_1}}{|A_1|} \otimes \tilde{\rho}_R \right\|_2^2 &= \int dU \text{tr} \left(\left(\tilde{\sigma}_{A_1 R}(U) - \frac{\mathbb{1}_{A_1}}{|A_1|} \otimes \tilde{\rho}_R \right)^2 \right) \\ &= \int dU \left[\text{tr} \left(\tilde{\sigma}_{A_1 R}^2(U) \right) - 2 \text{tr} \left(\tilde{\sigma}_{A_1 R}(U) \frac{\mathbb{1}_{A_1}}{|A_1|} \otimes \tilde{\rho}_R \right) + \text{tr} \left(\frac{\mathbb{1}_{A_1}}{|A_1|^2} \otimes \tilde{\rho}_R^2 \right) \right] \\ &= \int dU \left[\text{tr} \left(\tilde{\sigma}_{A_1 R}^2(U) \right) - \text{tr} \left(\frac{\mathbb{1}_{A_1}}{|A_1|^2} \otimes \tilde{\rho}_R^2 \right) \right] \\ &= \int dU \text{tr} \left(\tilde{\sigma}_{A_1 R}^2(U) \right) - \frac{1}{|A_1|} \text{tr} \left(\tilde{\rho}_R^2 \right) \\ &\leq \frac{1}{|A_2|} \text{tr} \left(\tilde{\rho}_{AR}^2 \right) \leq \frac{\varepsilon^2}{|A_1|}, \end{aligned} \quad (\text{D.30})$$

where in the third line we have used $\frac{\mathbb{1}_{A_1}}{|A_1|} \otimes \tilde{\rho}_R = \int dU \tilde{\sigma}_{A_1 R}(U)$, in the first inequality we have used Lemma 6, and in the final inequality we have used (D.27). \square

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] N. Engelhardt and A.C. Wall, *Quantum Extremal Surfaces: Holographic Entanglement Entropy beyond the Classical Regime*, *JHEP* **01** (2015) 073 [[arXiv:1408.3203](https://arxiv.org/abs/1408.3203)] [[INSPIRE](https://inspirehep.net/literature/1247083)].
- [2] G. Penington, *Entanglement Wedge Reconstruction and the Information Paradox*, *JHEP* **09** (2020) 002 [[arXiv:1905.08255](https://arxiv.org/abs/1905.08255)] [[INSPIRE](https://inspirehep.net/literature/1713491)].
- [3] 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](https://arxiv.org/abs/1905.08762)] [[INSPIRE](https://inspirehep.net/literature/1713491)].
- [4] C. Akers, S. Leichenauer and A. Levine, *Large Breakdowns of Entanglement Wedge Reconstruction*, *Phys. Rev. D* **100** (2019) 126006 [[arXiv:1908.03975](https://arxiv.org/abs/1908.03975)] [[INSPIRE](https://inspirehep.net/literature/1744103)].
- [5] M.A. Nielsen and I.L. Chuang, *Quantum Computation and Quantum Information: 10th Anniversary Edition*, Cambridge University Press, U.S.A., 10th ed., (2011).
- [6] R. Renner and R. Koenig, *Universally composable privacy amplification against quantum adversaries*, [quant-ph/0403133](https://arxiv.org/abs/quant-ph/0403133).

- [7] R. Renner, *Security of quantum key distribution*, [quant-ph/0512258](#).
- [8] P. Hayden and G. Penington, *Learning the Alpha-bits of Black Holes*, *JHEP* **12** (2019) 007 [[arXiv:1807.06041](#)] [[INSPIRE](#)].
- [9] G. Penington, S.H. Shenker, D. Stanford and Z. Yang, *Replica wormholes and the black hole interior*, [arXiv:1911.11977](#) [[INSPIRE](#)].
- [10] B. Czech, J.L. Karczmarek, F. Nogueira and M. Van Raamsdonk, *The Gravity Dual of a Density Matrix*, *Class. Quant. Grav.* **29** (2012) 155009 [[arXiv:1204.1330](#)] [[INSPIRE](#)].
- [11] A. Almheiri, X. Dong and D. Harlow, *Bulk Locality and Quantum Error Correction in AdS/CFT*, *JHEP* **04** (2015) 163 [[arXiv:1411.7041](#)] [[INSPIRE](#)].
- [12] X. Dong, D. Harlow and A.C. Wall, *Reconstruction of Bulk Operators within the Entanglement Wedge in Gauge-Gravity Duality*, *Phys. Rev. Lett.* **117** (2016) 021601 [[arXiv:1601.05416](#)] [[INSPIRE](#)].
- [13] X. Dong and H. Wang, *Enhanced corrections near holographic entanglement transitions: a chaotic case study*, *JHEP* **11** (2020) 007 [[arXiv:2006.10051](#)] [[INSPIRE](#)].
- [14] 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](#)].
- [15] J.M. Maldacena, *Eternal black holes in anti-de Sitter*, *JHEP* **04** (2003) 021 [[hep-th/0106112](#)] [[INSPIRE](#)].
- [16] A. Almheiri, X. Dong and B. Swingle, *Linearity of Holographic Entanglement Entropy*, *JHEP* **02** (2017) 074 [[arXiv:1606.04537](#)] [[INSPIRE](#)].
- [17] D. Harlow, *The Ryu-Takayanagi Formula from Quantum Error Correction*, *Commun. Math. Phys.* **354** (2017) 865 [[arXiv:1607.03901](#)] [[INSPIRE](#)].
- [18] P. Boes, J. Eisert, R. Gallego, M.P. Müller and H. Wilming, *Von Neumann entropy from unitarity*, *Phys. Rev. Lett.* **122** (2019) 210402 [[arXiv:1807.08773](#)] [[INSPIRE](#)].
- [19] R. Renner and S. Wolf, *Smooth Rényi entropy and applications*, in *International Symposium on Information Theory, 2004. ISIT 2004. Proceedings*, Chicago, IL, U.S.A., pp. 233, 27 June–2 July 2004 [[DOI](#)].
- [20] R. König, R. Renner and C. Schaffner, *The operational meaning of min-and max-entropy*, *IEEE Trans. Inform. Theory* **55** (2009) 4337.
- [21] M. Horodecki, J. Oppenheim and A. Winter, *Partial quantum information*, *Nature* **436** (2005) 673.
- [22] P. Calabrese and J.L. Cardy, *Entanglement entropy and quantum field theory*, *J. Stat. Mech.* **0406** (2004) P06002 [[hep-th/0405152](#)] [[INSPIRE](#)].
- [23] A. Lewkowycz and J. Maldacena, *Generalized gravitational entropy*, *JHEP* **08** (2013) 090 [[arXiv:1304.4926](#)] [[INSPIRE](#)].
- [24] T. Faulkner, A. Lewkowycz and J. Maldacena, *Quantum corrections to holographic entanglement entropy*, *JHEP* **11** (2013) 074 [[arXiv:1307.2892](#)] [[INSPIRE](#)].
- [25] X. Dong and A. Lewkowycz, *Entropy, Extremality, Euclidean Variations, and the Equations of Motion*, *JHEP* **01** (2018) 081 [[arXiv:1705.08453](#)] [[INSPIRE](#)].
- [26] A. Almheiri, T. Hartman, J. Maldacena, E. Shaghoulian and A. Tajdini, *Replica Wormholes and the Entropy of Hawking Radiation*, *JHEP* **05** (2020) 013 [[arXiv:1911.12333](#)] [[INSPIRE](#)].

- [27] C. Akers and P. Rath, *Holographic Renyi Entropy from Quantum Error Correction*, *JHEP* **05** (2019) 052 [[arXiv:1811.05171](#)] [[INSPIRE](#)].
- [28] 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](#)].
- [29] 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](#)].
- [30] N. Bao, G. Penington, J. Sorce and A.C. Wall, *Holographic Tensor Networks in Full AdS/CFT*, [arXiv:1902.10157](#) [[INSPIRE](#)].
- [31] 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](#)].
- [32] X. Dong, *The Gravity Dual of Renyi Entropy*, *Nature Commun.* **7** (2016) 12472 [[arXiv:1601.06788](#)] [[INSPIRE](#)].
- [33] M. Fannes, *A continuity property of the entropy density for spin lattice systems*, *Commun. Math. Phys.* **31** (1973) 291.
- [34] P. Hayden, S. Nezami, X.-L. Qi, N. Thomas, M. Walter and Z. Yang, *Holographic duality from random tensor networks*, *JHEP* **11** (2016) 009 [[arXiv:1601.01694](#)] [[INSPIRE](#)].
- [35] A.W. Harrow, *The church of the symmetric subspace*, [arXiv:1308.6595](#).
- [36] M. Berta, M. Christandl and R. Renner, *The quantum reverse shannon theorem based on one-shot information theory*, *Commun. Math. Phys.* **306** (2011) 579.
- [37] F. Dupuis, M. Berta, J. Wullschleger and R. Renner, *One-shot decoupling*, *Commun. Math. Phys.* **328** (2014) 251.
- [38] C.-F. Chen, G. Penington and G. Salton, *Entanglement Wedge Reconstruction using the Petz Map*, *JHEP* **01** (2020) 168 [[arXiv:1902.02844](#)] [[INSPIRE](#)].
- [39] P. Hayden, M. Horodecki, A. Winter and J. Yard, *A decoupling approach to the quantum capacity*, *Open Syst. Info. Dyn.* **15** (2008) 7.
- [40] F. Dupuis, *The decoupling approach to quantum information theory*, [arXiv:1004.1641](#).
- [41] H. Barnum and E. Knill, *Reversing quantum dynamics with near-optimal quantum and classical fidelity*, *J. Math. Phys.* **43** (2002) 2097.
- [42] M. Ohya and D. Petz, *Quantum entropy and its use*, Springer Science & Business Media (2004).
- [43] 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](#)].
- [44] H.F. Jia and M. Rangamani, *Petz reconstruction in random tensor networks*, *JHEP* **10** (2020) 050 [[arXiv:2006.12601](#)] [[INSPIRE](#)].
- [45] M. Horodecki, J. Oppenheim and A. Winter, *Quantum State Merging and Negative Information*, *Commun. Math. Phys.* **269** (2006) 107 [[quant-ph/0512247](#)] [[INSPIRE](#)].
- [46] M. Berta, *Single-shot quantum state merging*, [arXiv:0912.4495](#).
- [47] N. Bao and H. Ooguri, *Distinguishability of black hole microstates*, *Phys. Rev. D* **96** (2017) 066017 [[arXiv:1705.07943](#)] [[INSPIRE](#)].

- [48] P. Hayden and G. Penington, *Approximate Quantum Error Correction Revisited: Introducing the Alpha-Bit*, *Commun. Math. Phys.* **374** (2020) 369 [[arXiv:1706.09434](#)] [[INSPIRE](#)].
- [49] A. Abeyesinghe, I. Devetak, P. Hayden and A. Winter, *The mother of all protocols: Restructuring quantum information's family tree*, *Proc. Roy. Soc. Lond. A* **465** (2009) 2537 [[quant-ph/0606225](#)].
- [50] N. Datta and M.-H. Hsieh, *The apex of the family tree of protocols: optimal rates and resource inequalities*, *New J. Phys.* **13** (2011) 093042.
- [51] V.E. Hubeny, M. Rangamani and T. Takayanagi, *A Covariant holographic entanglement entropy proposal*, *JHEP* **07** (2007) 062 [[arXiv:0705.0016](#)] [[INSPIRE](#)].
- [52] S. Ryu and T. Takayanagi, *Holographic derivation of entanglement entropy from AdS/CFT*, *Phys. Rev. Lett.* **96** (2006) 181602 [[hep-th/0603001](#)] [[INSPIRE](#)].
- [53] 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](#)].
- [54] C. Akers, N. Engelhardt, G. Penington and M. Usatyuk, *Quantum Maximin Surfaces*, *JHEP* **08** (2020) 140 [[arXiv:1912.02799](#)] [[INSPIRE](#)].
- [55] A. Vitanov, F. Dupuis, M. Tomamichel and R. Renner, *Chain rules for smooth min- and max-entropies*, *IEEE Trans. Inform. Theory* **59** (2013) 2603.
- [56] C. Akers and G. Penington, *Quantum minimal surfaces from quantum error correction*, in progress.
- [57] R.M. Wald, *Black hole entropy is the Noether charge*, *Phys. Rev. D* **48** (1993) R3427 [[gr-qc/9307038](#)] [[INSPIRE](#)].
- [58] X. Dong, *Holographic Entanglement Entropy for General Higher Derivative Gravity*, *JHEP* **01** (2014) 044 [[arXiv:1310.5713](#)] [[INSPIRE](#)].
- [59] A.C. Wall, *A proof of the generalized second law for rapidly changing fields and arbitrary horizon slices*, *Phys. Rev. D* **85** (2012) 104049 [Erratum *ibid.* **87** (2013) 069904] [[arXiv:1105.3445](#)] [[INSPIRE](#)].
- [60] R. Bousso, Z. Fisher, S. Leichenauer and A.C. Wall, *Quantum focusing conjecture*, *Phys. Rev. D* **93** (2016) 064044 [[arXiv:1506.02669](#)] [[INSPIRE](#)].
- [61] R. Bousso, V. Chandrasekaran and A. Shahbazi-Moghaddam, *From black hole entropy to energy-minimizing states in QFT*, *Phys. Rev. D* **101** (2020) 046001 [[arXiv:1906.05299](#)] [[INSPIRE](#)].
- [62] P. Hayden, B. Swingle and M. Walter, *One-shot Information Theory in Quantum Field Theory*, unpublished.
- [63] M. Freedman and M. Headrick, *Bit threads and holographic entanglement*, *Commun. Math. Phys.* **352** (2017) 407 [[arXiv:1604.00354](#)] [[INSPIRE](#)].