THE UNIVERSITY OF CHICAGO

ON THE INFORMATION CONTENT OF BLACK HOLES AND THERMALIZED STATES

A DISSERTATION SUBMITTED TO
THE FACULTY OF THE DIVISION OF THE PHYSICAL SCIENCES
IN CANDIDACY FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

DEPARTMENT OF PHYSICS

BY
JONAH KUDLER-FLAM

CHICAGO, ILLINOIS
AUGUST 2022

To my parents

"You need something to open up a new door, to show you something you seen before but overlooked a hundred times or more."

– Bob Dylan; Last Thoughts On Woody Guthrie

TABLE OF CONTENTS

LIST OF FIGURES

ACKNOWLEDGMENTS

First and foremost, I would like to thank my advisor, Shinsei Ryu. Shinsei's excitement and unending curiosity in all things physics is unmatched and has served as a remarkable inspiration for me. He has taught me how to search not only for answers, but for questions in places where I would never have thought to look. It is no surprise that some of his most beautiful work has involved making connections between seemingly disparate fields of physics. The breadth and depth of his knowledge has been a tremendous resource for me. I know that I will continue to learn from this time under his guidance for many years to come.

I was very fortunate to also work closely with Hong Liu, another great physicist and dedicated mentor, during the year that I visited MIT. While I never step foot in the CTP due to COVID, Hong welcomed me with kindness and enthusiasm. He pushed me to think and work in new ways, and supported me in my pursuits, for which I am deeply grateful.

One of the most enjoyable parts of doing physics is getting to know and working with so many wonderful people. I would like to thank Meseret Asrat, Chang-Han Chen, Yuya Kusuki, Shang Liu, Yuhan Liu, Ian MacCormack, Vladimir Narovlansky, Laimei Nie, Masahiro Nozaki, Raman Sohal, Mao Tian Tan, Shreya Vardhan, Ashvin Vishwanath for many enjoyable and fruitful collaborations. I would especially like to thank Hassan Shapourian who was not only a collaborator but also a mentor who helped guide me through my Ph.D. from my first semester to my last.

Grandparents are not frequently acknowledged in physics theses, but it just wouldn't feel right without them here. My Nana taught me algebra at the dinner table, a story she never let me forget. My Papa, who had a passion for engineering but was made to go into patent law, always took great interest in my intellectual pursuits. I feel that we have been on this journey together. My Bubbie, a school teacher, always impressed on me the value of education and of being a good teacher. I know she is proud of me despite becoming the wrong kind of doctor. If my Zaydie was still with us now, there is no doubt in my mind

that he would print out this 200 page thesis, laminate it, and put it in his binder full of my papers.

I thank my parents, to whom this thesis is dedicated and could not exist without. My parents have given me continuous love and support throughout my life, and instilled in me the desire to always be learning and growing. I thank my sister Miriam for always bringing humor and joy to my life. Finally, I thank Joni for her constant support, companionship, and love.

ABSTRACT

The theory of quantum information has emerged as an indispensable tool in the study of many-body, high-energy, and gravitational physics, vastly over-delivering on its initial promises in quantum communications and computation. Significant insight has come from the analysis and quantification of quantum entanglement, a phenomenon that leaves an indelible imprint on the structure of a quantum state. In this thesis, I move beyond entanglement theory to incorporate the theory of distinguishability in many-body and gravitational physics. The notion of distinguishability provides a natural language and unique perspective on the physics of black hole evaporation and thermalization of isolated quantum many-body systems. I first develop tools to evaluate measures of distinguishability in random states and evaluate "Page curves" for various relative entropies and fidelities. I then relate these computations and generalize them to gravitational systems, characterizing when and how different black hole microstates in a gravitational model become distinguishable using measurements on their radiation and in what sense information can be recovered, a fine-grained resolution to a version of Hawking's information paradox. I furthermore discuss the role of relative entropy in AdS/CFT and the random tensor networks and quantum error-correcting codes that it is modeled by. Moving away from gravity, I discuss how distinguishability measures characterize the physics of quantum thermalization both in out-of-equilibrium processes and in high-energy eigenstates of non-integrable Hamiltonian systems. Under an ansatz for these high-energy states, I derive the eigenstate thermalization hypothesis in its strongest form. Along with unpublished results, this thesis incorporates lightly edited material from Refs. [106, 108, 112, 183].

CHAPTER 1

INTRODUCTION: PARADOXES

"People disagreeing everywhere you look. Makes you wanna stop and read a book."

— Watching the River Flow

As people like to say that Lord Kelvin said before the discovery of quantum mechanics and general relativity, "There is nothing new to be discovered in physics now. All that remains is more and more precise measurement." While Lord Kelvin never actually said this,[1] it succinctly captures the driving force behind theoretical progress, inconsistencies, or more strongly, paradoxes. By the late nineteenth century, classical physics was well-understood and boasted remarkable achievements, such as Newtonian gravity and Maxwell's theory of electromagnetism, both seemingly consistent with the physics of the natural world. Without apparent inconsistencies, a scientist is hard-pressed to find meaningful direction and it is not unreasonable, though dangerous, to become complacent. Fortunately, paradoxes arose, such as the ultraviolet (Rayleigh–Jeans) catastrophe, that could only be resolved by the introduction of the genuinely new physical structures of quantum theory.

In this thesis, I will discuss two more modern and interrelated (apparent) paradoxes that have arisen from quantum theory and my work to resolve and elucidate their nature using the framework of quantum information theory. The first is Hawking's black hole information paradox, the breakdown of the predictability of quantum field theory in the presence of black holes. The second is the problem of emergence of thermal physics in isolated quantum systems where thermalization seems a priori impossible.

A unifying idea spanning quantum information theory, quantum chaos and thermalization, and black hole physics is that of (in)distinguishability of quantum states. In quantum

1. A version of this misquote is better attributed to Albert A. Michelson, the namesake of the building this thesis will be defended in, during an 1894 address at the dedication of the Ryerson Physical Laboratory at the University of Chicago.

information theory, we would like to understand what the space of quantum states are. In particular, how can we characterize which states are close or far away and endow the Hilbert space with a geometry? This notion of distinguishability is critical for storing and processing quantum information.

Quantum chaos and thermalization is all about distinguishibility. A natural *definition* of quantum thermalization is that the state is indistinguishable (up to a certain error) from a completely thermal e.g. Gibbs state. It is then important to characterize which systems thermalize and the mechanism for thermalization to occur. We can begin with two states that are easily distinguishable e.g. the "all spin up" and "all spin down" states of a quantum spin chain. If we evolve these states with a thermalizing Hamiltonian, the states will become indistinguishable using "simple" measurements.

Similarly, the black hole information problem is most naturally framed in terms of thermalization and indistinguishability. Black holes can be formed in many different ways. Moreover, they have an extraordinary number of microstates [22, 76]. Even so, using semiclassical calculations, Hawking showed that all black holes with identical thermodynamic quantities (mass, charge, and angular momentum) will radiate *thermal* radiation [76]. This means that at late times, after the black hole has evaporated, all of these microstates are completely indistinguishable, which is in sharp tension with the unitarity of quantum mechanics. To resolve this apparent paradox, different black hole microstates must be made distinguishable directly from measurements on the radiation.

Remarkably, all three of these broad problems may be addressed using random matrix theory calculations of distinguishability measures. In this thesis, this statement precise is made precise and I elucidate the intricate and surprising connections.

1.1 Organization

The remainder of this thesis will be organized as follows: In Chapter 2, I provide background, describing relevant quantities and techniques from quantum information theory. Particular emphasis is placed on the operational interpretations of various distinguishability measures that are used throughout the work. In Chapter 3, I introduce random quantum states and their relation to black holes and thermalization. I introduce a diagrammatic technique to compute moments of matrices drawn from the the Wishart ensemble and subsequently evaluate many distinct distinguishability measures in the limit of large Hilbert space dimension. These analytical results are cross-checked with finite size numerical simulations. Alternative derivations of the results are also provided using the techniques of free probability theory, for which I give some general background. In Chapter 4, I relate these computations to the physics of black holes and extend them accordingly. Namely, I study the distinguishablity between different states with the same bulk geometry (same "code subspace") in the AdS/CFT correspondence. Of particular interest is the distinguishability of different microstates of a black hole. In a two-dimensional model of black hole evaporation using Jackiw-Teitelboim gravity, I study how information leaks out of a black hole via the Hawking modes by characterizing how the individual microstates can be distinguished using measurements solely on the radiation. This is made more explicit using general arguments for how and when information thrown into a black hole can be recovered from the state of the radiation. I then discuss further developments in AdS/CFT by classifying various corrections and violations to the holographic relative entropy (JLMS) formula. In Chapter 5, I change gears, focusing on non-gravitational, many-body quantum systems. Here, I discuss two approaches to thermalization, that which emerges at late times after non-equilibrium dynamics and that which is present in individual energy eigenstates. Under certain assumptions this thermalization can be understood by generalizing the techniques from random matrix theory explored previously. In Chapter 6, I discuss ongoing and future lines of research that I hope to be

answered. The focus of this section is on near-term tractable goals, with minimal emphasis placed on more fanciful long-term goals.

CHAPTER 2

QUANTUM INFORMATION THEORY AND ITS USE

"The truth was obscure, too profound and too pure."

— Where Are You Tonight? (Journey Through Dark Heat)

2.1 Entanglement

While distinguishability measures are the main focus of this thesis, we will first provide a brief review of some notions of entanglement theory because these will arise at various points and they play a crucial role in setting context.

In quantum mechanics, a pure state is an element of a Hilbert space, $\mathcal{H}$. A state need not be pure. For example, our information about a quantum state may consist of a classical probability distribution over pure states. In this more general context, we must consider density matrices which are positive semi-definite linear maps on $\mathcal{H}$ with unit trace, $w(\mathcal{H})$. Pure states become projectors. For a given $\rho \in w(\mathcal{H})$, we can define the von Neumann entropy

$$S_{vN}(\rho) := -\mathrm{Tr}\left[\rho \log\left[\rho\right]\right], \tag{2.1}$$

which reduces to the classical Shannon entropy for diagonal ρ. For pure states, the entropy is always zero because all microscopic information is known about the state. In general, Hilbert spaces can factorize into subsystems as $\mathcal{H} = \mathcal{H}_A \otimes \mathcal{H}_B$. Given a state, $\rho \in w(\mathcal{H})$, one can define a reduced state on subsystem A by taking the partial trace over $\mathcal{H}_B$

$$\rho_A := \mathrm{Tr}_B\left[\rho\right]. \tag{2.2}$$

If ρ is a pure state, then the von Neumann entropy of ρ_A is faithful measure of entanglement

between A and B and may then rightfully be called the entanglement entropy. This entropy has many operational interpretations, such as the number of Bell pairs shared between A and B that can be (asymptotically) distilled from the state ρ using only local operations and classical communications (LOCC) between the systems.

With all this talk of entanglement, we should really have a definition. To do this, we first define a separable state, which is any state that can be written in the following form

$$\rho = \sum_k p_k \rho_A^{(k)} \otimes \rho_B^{(k)}, \quad \rho_A^{(k)} \in w(\mathcal{H}_A), \quad \rho_B^{(k)} \in w(\mathcal{H}_B), \tag{2.3}$$

where p_k is a classical probability distribution. This constitutes the set of states that can be generated using LOCC. All states that are not separable are called entangled. It is clear that the von Neumann entropy of ρ_A for a generic separable state is non-zero, so it is not a measure of entanglement.

A frequently used measure for mixed states is the mutual information, which may be defined as a linear combination of von Neumann entropies

$$I(A,B) = S_{vN}(\rho_A) + S_{vN}(\rho_B) - S_{vN}(\rho_{AB}). \tag{2.4}$$

The mutual information is positive semi-definite. It is only zero for product states, which are of the form $\rho = \rho_A \otimes \rho_B$. Because it is non-zero for other separable states, it is not a faithful measure of entanglement, though it does measure total correlations, including classical correlations and has been shown to bound connected correlation functions between the two subsystems [193]

$$I(A,B) \geq \frac{\left(\langle \mathcal{O}_A \mathcal{O}_B \rangle - \langle \mathcal{O}_A \rangle \langle \mathcal{O}_B \rangle\right)^2}{2\left|\mathcal{O}_A\right|_1^2 \left|\mathcal{O}_B\right|_1^2}. \tag{2.5}$$

In the denominator, the norms are one-norms. The mutual information is also monotonic

under inclusions for arbitrary C

$$I(A, B \cup C) \geq I(A, B), \tag{2.6}$$

so it makes sense as a correlation measure.

In general, genuine measures of entanglement are very difficult to compute in practice. This is because they are frequently defined using optimizations over large spaces of quantum states. One exception to these difficulties is the logarithmic negativity, which is tractable to compute and requires no optimizations. The negativity is based on the positive partial transpose (PPT) criterion [160, 91]. The partial transpose (say with respect to B[1]) of ρ in the orthonormal basis composed of $|i\rangle_A$ of $\mathcal{H}_A$ and $|j\rangle_B$ of $\mathcal{H}_B$ is defined to have matrix elements

$$\rho^{T_B}_{i_A j_B, k_A l_B} = \rho_{i_A l_B, k_A j_B}, \tag{2.7}$$

which is a Hermitian matrix with unit trace. While the eigenvalues of ρ are non-negative, this is not necessarily the case for its partial transpose. Negativity is based on this property and quantifies the amount of entanglement by the negative eigenvalues of ρ^{T_B}. Denoting the eigenvalues of ρ^{T_B} by λ_i, the *negativity* is defined by [184, 54, 165]

$$\mathcal{N} := \frac{|\rho^{T_B}|_1 - 1}{2} = \sum_{\lambda_i < 0} |\lambda_i|, \tag{2.8}$$

and similarly the *logarithmic negativity* is

$$\mathcal{E} = \log|\rho^{T_B}|_1. \tag{2.9}$$

Indeed, for separable states, $\rho^{T_B} = \sum_a p_a \rho_A^{(k)} \otimes \left(\rho_B^{(k)}\right)^T$, where $\left(\rho_B^{(k)}\right)^T$ is also a density

1. These definitions of negativity do not depend on the choice of bases for $\mathcal{H}_A, \mathcal{H}_B$, and are invariant under interchanging A and B.

matrix, so ρ^{T_B} has no negative eigenvalues. This shows that a necessary condition for a state to be unentangled is to have vanishing negativity.[2] Crucially, negativity is monotonically decreasing under LOCC. Note that it is known that determining if a density matrix is separable is generally NP hard [73, 71].

2.2 Distinguishability

I will now review various distinguishability measures commonly used in quantum information theory. Each measure has an operational meaning via quantum hypothesis testing and there are various relations between the measures.

Relative entropy We begin with the quantum relative entropy which is arguably the most important quantity in quantum information theory as many of the deepest results in the field are directly derivable from its fundamental properties. The classical relative entropy or Kullback-Leibler divergence is defined as

$$D_{KL}(P||Q) := \log\left[\sum_{x\in X} P(x)\log\left(\frac{P(x)}{Q(x)}\right)\right], \tag{2.10}$$

where P and Q are classical probability distributions over a set X. The quantum relative entropy is the noncommutative analog defined for two density matrices, ρ and σ as

$$D(\rho||\sigma) := \mathrm{Tr}\left[\rho\log\left[\rho\right] - \rho\log\left[\sigma\right]\right]. \tag{2.11}$$

This is only well-defined when the support of ρ is contained within the support of σ. Otherwise, the relative entropy is infinite.

The relative entropy acts as a distinguishability measure as can be seen from its basic

2. It is not a sufficient condition. However, the entanglement that goes undetected by the negativity may not be useful as negativity places an upper bound on the distillable entanglement [184].

properties. The first is positivity, $D(\rho||\sigma) \geq 0$, with the inequality saturated if and only if $\rho = \sigma$. The second is referred to as the data processing inequality or monotonicity of relative entropy[3] which states that the relative entropy is non-increasing under completely-positive trace-preserving (CPTP) quantum channels, $\mathcal{N}$ [124]

$$D(\mathcal{N}(\rho)||\mathcal{N}(\sigma)) \leq D(\rho||\sigma). \tag{2.12}$$

This property is crucial for a distinguishability measure because it asserts that if you are given two quantum states, after performing operations on them, they can never become easier to distinguish.

A particularly important quantum channel is the partial trace operation on a bipartite Hilbert space

$$\mathcal{N} : \mathcal{H}_A \otimes \mathcal{H}_B \to \mathcal{H}_A \tag{2.13}$$

$$\rho \mapsto \mathrm{Tr}_B [\rho] := \rho_A. \tag{2.14}$$

Under the partial trace, we lose all information about region B, making ρ harder to distinguish from other states that look similar on A. The partial trace will play a central role throughout the rest of the thesis because we are generally interested in how to distinguish states when only having access to a subregion.

While the relative entropy characterizes the structure of the space of quantum states, importantly, it is not a metric. This is most obviously seen from the definition which is not symmetric under exchange of ρ and σ. This is a feature and not a bug as can be seen by its operational meaning that we will soon describe.

The relative entropy is a parent quantity to many other central information-theoretic

3. Positivity can actually be derived from monotonicity, though we chose to separate these conditions for added clarity.

quantities, such as the von Neumann entropy

$$S_{vN}(\rho_A) = \log\left[d_A\right] - D\left(\rho_A || \frac{\mathbb{1}}{d_A}\right), \tag{2.15}$$

where d_A is the Hilbert space dimension, the mutual information

$$I(A,B) = D(\rho_{AB}||\rho_A \otimes \rho_B), \tag{2.16}$$

and conditional entropy

$$S(B|A) = \log\left[d_B\right] - D\left(\rho_{AB}||\rho_A \otimes \frac{\mathbb{1}}{d_B}\right). \tag{2.17}$$

In these terms, the strong subadditivity of von Neumann entropy

$$S_{vN}(\rho_B) + S_{vN}(\rho_{ABC}) \leq S_{vN}(\rho_{AB}) + S_{vN}(\rho_{BC}) \tag{2.18}$$

is a straightforward consequence of the data processing inequality

$$D(\mathrm{Tr}_C\left[\rho_{ABC}\right] || \mathrm{Tr}_C\left[\rho_A \otimes \rho_{BC}\right]) \leq D(\rho_{ABC}||\rho_A \otimes \rho_{BC}). \tag{2.19}$$

Rényi relative entropies Like the Kullback-Leibler divergence, the relative entropy can be generalized into Rényi relative entropies. However, because of the noncommutativity of density matrices, there are many inequivalent ways to generalize the relative entropy such that it reduces to the classical α-Rényi divergences, the unique set of quantities satisfying the five axioms of a generalized divergence [170]

$$D_{KL,\alpha}(P||Q) := \frac{1}{\alpha - 1} \log\left[\sum_{x \in X} P(x)^\alpha Q(x)^{1-\alpha}\right], \tag{2.20}$$

where α is a positive semi-definite real variable. We will study two complementary families which have served the most uses in quantum information theory.

The first is the most obvious quantum analog of (2.20) and is referred to as the Petz Rényi relative entropy (PRRE) [161]

$$D_\alpha(\rho||\sigma) := \frac{1}{\alpha - 1} \log \left[\mathrm{Tr} \left[\rho^\alpha \sigma^{1-\alpha} \right] \right]. \tag{2.21}$$

The PRRE satisfies various nice properties, such as reduction to the von Neumann relative entropy when $\alpha \to 1$. For $\alpha \in [0, 1)$, the PRRE is finite even when the support of ρ is larger than the support of σ. Most importantly, the PRRE satisfies the data processing inequality when $\alpha \in [0, 2]$ [122, 181, 161]. One particularly useful case is at $\alpha = 1/2$, which defines what has been called Holevo's "just-as-good fidelity" [190] or affinity [101]

$$F_H(\rho||\sigma) := \left(\mathrm{Tr} \left[\sqrt{\rho}\sqrt{\sigma} \right] \right)^2 = e^{-D_{1/2}(\rho||\sigma)}, \tag{2.22}$$

which, for most purposes is just as (if not more) useful as the more widely used Uhlmann fidelity

$$F(\rho||\sigma) := \left(\mathrm{Tr} \left[\sqrt{\sqrt{\sigma}\rho\sqrt{\sigma}} \right] \right)^2. \tag{2.23}$$

Both satisfy all of Jozsa's axioms for distinguishability measures [97] and define metrics on the space of quantum states

$$\begin{aligned} D_B(\rho||\sigma) &:= \sqrt{2\left(1 - \sqrt{F(\rho||\sigma)}\right)}, \\ D_A(\rho||\sigma) &:= \arccos\left[\sqrt{F(\rho||\sigma)}\right], \\ D_H(\rho||\sigma) &:= 2\left(1 - \sqrt{F_H(\rho||\sigma)}\right), \end{aligned} \tag{2.24}$$

called the Bures distance, Bures angle, and Hellinger distance repectively.

The other quantum generalization of (2.20) we will study is the sandwiched Rényi relative entropy[4] (SRRE) [140, 191]

$$\tilde{D}_\alpha(\rho||\sigma) := \frac{1}{\alpha-1}\log\left[\mathrm{Tr}\left[\left(\sigma^{\frac{1-\alpha}{2\alpha}}\rho\sigma^{\frac{1-\alpha}{2\alpha}}\right)^\alpha\right]\right]. \tag{2.26}$$

It is clear that this is equivalent to the PRRE when ρ and σ commute and reduces to the Uhlmann fidelity at $\alpha = 1/2$. Like the PRRE, the SRRE reduces to the von Neumann relative entropy in the $\alpha \to 1$ limit and is only finite if either $\alpha \in [0, 1)$ or the support of ρ is contained within the support of σ. The most important property of SRRE is that it satisfies the data-processing inequality for $\alpha \in [1/2, \infty)$. In this way, it is complementary to the PRRE. Similar formulas for Rényi analogs of entropy, mutual information, and conditional entropies can be written in terms of the Rényi relative entropies.

Trace distance The final distinguishability measure that we will study is the trace distance, defined as

$$T(\rho||\sigma) := \frac{1}{2}\left|\rho - \sigma\right|_1, \tag{2.27}$$

The trace distance defines a metric on the space of quantum states and takes values between zero and one. However, unlike Holevo's just-as-good and Uhlmann fidelities, it does not descend from a relative entropy. The trace distance is monotonically decreasing under quantum operations. It will play a central role in our discussion of eigenstate thermalization in Chapter 5.

4. Both PRRE and SRRE can be described as specific cases of the α-z-relative entropies defined as [18]

$$D_{\alpha,z}(\rho||\sigma) := \frac{1}{\alpha-1}\log\left[\mathrm{Tr}\left[\left(\sigma^{\frac{1-\alpha}{2z}}\rho^{\frac{\alpha}{z}}\sigma^{\frac{1-\alpha}{2z}}\right)^z\right]\right], \tag{2.25}$$

though we will not discuss this more general quantity.

There are various useful relations between the above distinguishability measures. Firstly, both PRRE and SRRE are monotonic in α

$$D_{\alpha_1}(\rho||\sigma) \leq D_{\alpha_2}(\rho||\sigma), \quad \tilde{D}_{\alpha_1}(\rho||\sigma) \leq \tilde{D}_{\alpha_2}(\rho||\sigma), \quad \alpha_1 < \alpha_2, \tag{2.28}$$

while the SRRE lower bounds the PRRE

$$\tilde{D}_{\alpha}(\rho||\sigma) \leq D_{\alpha}(\rho||\sigma), \quad \alpha \geq 1. \tag{2.29}$$

By Pinsker's inequality, the von Neumann relative entropy upper bounds the trace distance [153]

$$\frac{1}{2}T(\rho||\sigma)^2 \leq D(\rho||\sigma), \tag{2.30}$$

while the Fuchs-van de Graaf inequalities assert that both fidelities place upper and lower bound the trace distance [67]

$$1 - \sqrt{F(\rho||\sigma)} \leq 1 - \sqrt{F_H(\rho||\sigma)} \leq T(\rho||\sigma) \leq \sqrt{1 - F(\rho||\sigma)} \leq \sqrt{1 - F_H(\rho||\sigma)}. \tag{2.31}$$

These are strong results that will be useful due to the difficulty in directly computing the trace distance. They are also important, nontrivial consistency checks of our results.

2.2.1 Operational interpretations in hypothesis testing

The most fundamental information processing processes are quantum state discrimination (QSD) and hypothesis testing (QHT). It should then be no surprise that this is where the most fundamental quantities, relative entropy and trace distance, find their operational meanings. In this section, we make precise what it means for states to be distinguishable by first introducing QSD and QHT, then stating what the distinguishability measures say about

our ability to perform these tasks. For more details, we refer the reader to the literature e.g. Refs. [78, 100].

The general set-up is that we are given a state on $\mathcal{H}$ that is either ρ or σ and we wish to determine which state we were given. We are allowed to use any positive operator-valued measure (POVM) which is a collection of positive semi-definite operators, $\{M_i\}$, that sum to the identity operator on $\mathcal{H}$. Each subscript, i, corresponds to a measurement outcome. Because we are looking for a binary outcome (is our state ρ or σ?), we can consolidate the M_i's into just two elements. For outcomes $i \in \mathcal{A}$, we conclude the state is ρ, while for outcomes $i \notin \mathcal{A}$, we conclude the state is σ. Our POVM is then $\{A, \mathbb{1} - A\}$ where $A := \sum_{i \in \mathcal{A}} M_i$. There are many choices for A and we want to optimize this choice as to have the least error in our conclusions. There are two types of errors. The probability of mistakenly concluding that we have σ when we were really given ρ is given by

$$\alpha(A) := \mathrm{Tr}\left[(\mathbb{1} - A)\rho\right], \tag{2.32}$$

while the probability of mistakenly concluding that we have ρ when we were really given σ is given by

$$\beta(A) := \mathrm{Tr}\left[A\sigma\right]. \tag{2.33}$$

These are referred to as the error probabilities of the first and second kind respectively (or type I and II).

There are various ways of optimizing these errors[5]. The symmetric way is called state discrimination. The smallest combined error is given by the trace distance between the states

5. Recently, an interpolation QSD and QHT has bee introduced [174], which will be commented on in Section 3.5.

[85, 86]

$$\min_A \left[\alpha(A) + \beta(A)\right] = 1 - T(\rho||\sigma), \tag{2.34}$$

where the optimization is taken over all POVM. If the trace distance is very large (close to one), we are able to choose a POVM that has very small error probabilities. If the trace distance is small (close to zero), then the combined error is close to one, the maximal optimized error which can be saturated by taking $A = \mathbb{1}$. Likewise, the probability that we correctly discriminate, P_+, is also given by the trace distance

$$P_+(A) := \frac{1}{2} \max_A \left[\mathrm{Tr}\left[A\rho\right] + \mathrm{Tr}\left[(\mathbb{1} - A)\sigma\right]\right] = \frac{1}{2}\left(1 + T(\rho||\sigma)\right). \tag{2.35}$$

State discrimination can be made easier if instead of given one copy of the state, we are given multiple, n, copies. This is the topic of asymptotic state discrimination. With these n copies, we can ask what is the optimal POVM on $\mathcal{H}^{\otimes n}$. The error probabilities are generalized in the obvious way

$$\alpha_n(A) := \mathrm{Tr}\left[(\mathbb{1} - A)\rho^{\otimes n}\right], \quad \beta_n(A) := \mathrm{Tr}\left[A\sigma^{\otimes n}\right]. \tag{2.36}$$

The sum of the errors can be shown to be bounded above by Holevo's just-as-good fidelity

$$\min_A \left[\alpha_n(A) + \beta_n(A)\right] \leq F_H(\rho||\sigma)^{\frac{n}{2}}. \tag{2.37}$$

Unless the states are identical ($F_H = 1$), the error rate exponentially decays to zero as we are given a large number of copies. If the fidelity is small, we may only need one (or very few) copies to confidently discriminate the states. Asymptotically ($n \to \infty$), this is strengthened

to an equality by the *quantum Chernoff bound* [16, 148]

$$\lim_{n\to\infty} -\frac{\log\left[\min_A\left[\alpha_n(A)+\beta_n(A)\right]\right]}{n} = \max_{\alpha\in(0,1)}(1-\alpha)D_\alpha(\rho||\sigma) := \xi(\rho||\sigma). \tag{2.38}$$

The quantity on the right-hand side of this equation is called the *quantum Chernoff distance*.

We progress to the asymmetric treatment of this problem, quantum hypothesis testing. The asymmetric optimization is the task of minimizing one of the errors while keeping the other error below some fixed, finite threshold ϵ. We define

$$\alpha_n^*(\epsilon) := \min_{\beta_n(A)\leq\epsilon}\left[\alpha_n(A)\right], \quad \beta_n^*(\epsilon) := \min_{\alpha_n(A)\leq\epsilon}\left[\beta_n(A)\right]. \tag{2.39}$$

Quantum Stein's Lemma [87, 152] asserts that for *any* $\epsilon\in(0,1)$, the type II error decreases exponentially with the rate given by the relative entropy

$$\lim_{n\to\infty} -\frac{\log\left[\beta_n^*(\epsilon)\right]}{n} = D(\rho||\sigma). \tag{2.40}$$

Quantum Stein's Lemma can be further refined to optimize the error of the first kind assuming the error of the second kind decays exponentially. Defining

$$\tilde{\alpha}_{n,r} := \min_{\beta_n(A)\leq e^{-nr}}\left[\alpha_n(A)\right], \tag{2.41}$$

the PRRE determines this error rate if $r < D(\rho||\sigma)$ [79, 142, 138]

$$\lim_{n\to\infty} -\frac{\log\left[\tilde{\alpha}_{n,r}\right]}{n} = \max_{\alpha\in(0,1)}\left[\frac{\alpha-1}{\alpha}\left(r - D_\alpha(\rho||\sigma)\right)\right], \tag{2.42}$$

while SRRE determines this error rate if $r > D(\rho||\sigma)$ [139]

$$\lim_{n\to\infty} -\frac{\log\left[1-\tilde{\alpha}_{n,r}\right]}{n} = \max_{\alpha\in(1,\infty)}\left[\frac{\alpha-1}{\alpha}\left(r - \tilde{D}_\alpha(\rho||\sigma)\right)\right]. \tag{2.43}$$

With the above review, we have a thorough understanding of how to quantify the ability to discriminate between *two* quantum states. It would be desirable to generalize this to an arbitrary, finite number of states $\{\rho_i\}$. This is particularly important for the black hole information problem where there are many states with the same thermodynamic parameters. In the case that we are discriminating many states, we no longer consolidate the M_i's into A and $\mathbb{1} - A$. Rather, each measurement outcome i can lead us to conclude that we have state ρ_i. If we are given the state ρ_i with probability p_i, the error probability is given by

$$P_{err}(\rho_i, M_i) := \sum_{i=1} \mathrm{Tr}\left[p_i \rho_i (\mathbb{1} - M_i)\right], \tag{2.44}$$

whose optimized value we define as

$$P^*_{err}(\rho_i) := \min_{\{M_i\}} \left[P_{err}(\rho_i, M_i)\right]. \tag{2.45}$$

Rather remarkably, building on the work of Refs. [149, 151, 150, 19], the quantum Chernoff bound was generalized in the multiple state case, referred to as the *multiple quantum Chernoff bound*[6] [120]

$$\lim_{n\to\infty} -\frac{\log\left[P^*_{err}(\rho_i)\right]}{n} = \min_{i\neq j}\left[\max_{\alpha\in(0,1)} (1-\alpha) D_\alpha(\rho_i||\rho_j)\right]. \tag{2.46}$$

The value on the right hand side is referred to as the *multiple quantum Chernoff distance*. When comparing to (2.38), it is surprising that when discriminating between arbitrarily many more states, all one needs to do is apply a global minimum.

In the one-shot case, bounds can be placed on $P^*_{err}(\rho_i)$, though, to our knowledge, an equality is not known. If we take the spectral decompositions of our POVM as $M_i :=$

6. The quantum Sanov's lemma provides the analogous asymmetric multiple state hypothesis testing result [77, 24]. There are also intriguing new multiple state divergences obeying the data processing inequality whose operational meaning is not yet fully understood [68].

$\sum_i^{T_i} \lambda_{ik} Q_{ik}$, an upper and bound is given by [120]

$$\frac{\sum_{i<j} \sum_{k,l} \min\left[\lambda_{ik}, \lambda_{jl}\right] \mathrm{Tr}\left[Q_{ik} Q_{jl}\right]}{2(r-1)} \leq P^*_{err}(\rho_i) \leq 10(r-1)^2 T^2 \sum_{i<j} \sum_{k,l} \min\left[\lambda_{ik}, \lambda_{jl}\right] \mathrm{Tr}\left[Q_{ik} Q_{jl}\right], \tag{2.47}$$

where r is the total number of states and $T := \max[T_i]$. The upper bound can be made more intuitive, though generally weaker, by noting

$$\sum_{k,l} \min\left[\lambda_{ik}, \lambda_{jl}\right] \mathrm{Tr}\left[Q_{ik} Q_{jl}\right] \leq \min_{\alpha \in (0,1)} (1-\alpha) D_\alpha(\rho_i || \rho_j), \tag{2.48}$$

leading to

$$\begin{aligned} P^*_{err}(\rho_i) &\leq 10(r-1)^2 T^2 \sum_{i<j} \min_{\alpha \in (0,1)} (1-\alpha) D_\alpha(\rho_i || \rho_j) \\ &\leq 5(r-1)^3 r T^2 \max_{i \neq j} \left[\min_{\alpha \in (0,1)} (1-\alpha) D_\alpha(\rho_i || \rho_j) \right], \end{aligned} \tag{2.49}$$

where in the second line, the remaining sum has been removed to mimic the form of (2.46) even though this formula is strictly weaker when setting $r = 2$. It is worth noting that determining P^*_{err} is a computation can be formulated as a semi-definite program [194, 100], which means that it may be efficiently evaluated.

2.3 Recovery

An important question concerning a quantum channel $\mathcal{N}$ is whether or not it is reversible i.e. whether there exists a recovery channel $\mathcal{R}$ such that for any state ρ, $\mathcal{R} \circ \mathcal{N}(\rho) = \rho$. This question plays a central role, for example, in the theory of quantum error correction as well

as quantum thermalization. If the quantum channel is unitary, the initial state is perfectly recovered by acting on the output with the adjoint of the unitary. In the other extreme, a quantum channel could replace all states with the maximally mixed state, in which case an initial state is unrecoverable as the information about it is completely lost. A recovery map that is independent of the initial state ρ is called *universal*.

It follows from a theorem by Petz [164, 163, 162] that a quantum channel $\mathcal{N}$ is reversible if and only if the data processing inequality is saturated

$$D(\rho||\sigma) = D(\mathcal{N}(\rho)||\mathcal{N}(\sigma)) \tag{2.50}$$

for all ρ, σ. Here, $D(\rho||\sigma)$ is the relative entropy. Furthermore, suppose we fix some reference state σ. Then, for any state X in the support of $\mathcal{N}(\sigma)$, there exists a recovery channel $\mathcal{R}$. This channel is given explicitly by the Petz map,

$$\mathcal{P}_{\sigma,\mathcal{N}}(X) = \sigma^{\frac{1}{2}}\mathcal{N}^{\dagger}\left((\mathcal{N}(\sigma))^{-\frac{1}{2}}X(\mathcal{N}(\sigma))^{-\frac{1}{2}}\right)\sigma^{\frac{1}{2}}, \tag{2.51}$$

where $\mathcal{N}^{\dagger}$ is the adjoint map of $\mathcal{N}$. As a basic check, relative entropy is invariant under unitary channels $\rho \to U\rho U^{\dagger}$, and we find from (2.51), $\mathcal{P}_{\sigma,\mathcal{N}}(X) = U^{\dagger}XU$.

Interestingly, it has recently been understood that the change of relative entropy under quantum channels places strict bounds on how well a state can be recovered. The basic idea is intuitive; if two states that were initially easily distinguishable become nearly indistinguishable under a channel, then it should be impossible to identify what the initial states were using only information about the indistinguishable output states. In particular, it can

be shown [98] that there exists a recovery map $\mathcal{R}_{\sigma,\mathcal{N}}$ with $\mathcal{R}_{\sigma,\mathcal{N}} \circ \mathcal{N}(\sigma) = \sigma$ satisfying[7],

$$F(\rho, [\mathcal{R}_{\sigma,\mathcal{N}} \circ \mathcal{N}](\rho)) \geq \exp(D(\mathcal{N}(\rho)||\mathcal{N}(\sigma)) - D(\rho||\sigma)), \tag{2.52}$$

where F is the Uhlmann fidelity. For example, the bound (2.52) holds for an explicit but complicated recovery channel called the twirled Petz map [98].

7. This was proven for Type I von Neumann algebras but was recently generalized to the Type III algebras relevant to quantum field theory [63]. This technicality will not play an important role for us.

CHAPTER 3

RANDOM STATES

"Half of the people can be part right all of the time

Some of the people can be all right part of the time

But all of the people can't be all right all of the time"

— Talkin' World War III Blues

The techniques of random matrix theory have become ubiquitous across far-ranging fields of physics. Originally used to characterize the spectra of heavy nuclei [189], random matrix theory has flourished in its applications in quantum information theory [36], quantum chaos and thermalization [42], and black hole physics [155, 84]. What's more is that these fields are now understood to be deeply related to one another and, to some extent, inseparable.

In this chapter, we consider Haar random states. This ensemble of states can be described in several ways. Perhaps the simplest is to consider an arbitrary reference state $|0\rangle \in \mathcal{H}$ and act with a random unitary matrix drawn from the Haar measure, the unique left-right invariant measure over $U(\dim \mathcal{H}) : |\Psi\rangle = U|0\rangle$. This ensemble is particularly nice because the averages over α copies of Haar random states are sums of permutations, τ, of the α copies

$$\overline{|\Psi\rangle\langle\Psi|^{\otimes\alpha}} = \frac{\sum_{\tau\in S_\alpha} g_\tau}{\sum_{\tau\in S_\alpha} \mathrm{Tr}\left[g_\tau\right]}, \tag{3.1}$$

where g_τ is the matrix representation of τ and the denominator ensures that the state has unit norm. We are generally interested in ensembles for mixed states that are induced from taking a partial trace over a sub-Hilbert space. If $\mathcal{H} = \mathcal{H}_A \otimes \mathcal{H}_B$, the reduced density matrix on A is given by

$$\rho_A^{\otimes\alpha} := \mathrm{Tr}_B\left[\overline{|\Psi\rangle\langle\Psi|^{\otimes\alpha}}\right] = \frac{\sum_{\tau\in S_\alpha} g_{\tau_A} \mathrm{Tr}\left[g_{\tau_B}\right]}{\sum_{\tau\in S_\alpha} \mathrm{Tr}\left[g_\tau\right]}, \tag{3.2}$$

where the subscript on the permutation elements mean that they only permute within a sub-Hilbert space. The trace of a permutation element is straightforward to work out, equaling the dimension of the Hilbert space, $d_A d_B$, to the number of cycles in the permutation, $C(\tau)$. The denominator can then be written as

$$\sum_{\tau \in S_\alpha} \mathrm{Tr}\left[g_\tau\right] = \sum_{\tau \in S_\alpha} \left(d_A d_B\right)^{C(\tau)}, \tag{3.3}$$

which can be summed exactly because we know that the number of permutations of α elements with k cycles is given by the Stirling number of the first kind. However, we can easily avoid this technical point because we will be interested in the regime where the Hilbert space dimension is large. Therefore, only the permutations that maximize $C(\tau)$ will contribute at leading order. The unique permutation that maximizes $C(\tau)$ is the identity permutation which has α cycles, so from here forward, we will approximate the denominator as $(d_A d_B)^\alpha$.

There is an alternative description of the same induced ensemble of density matrices that will be useful for us when generalizing to tensor networks. Rather than starting with fiducial state $|0\rangle$, we begin with complete bases, $|i\rangle_A$ and $|J\rangle_B$, on $\mathcal{H}_A$ and $\mathcal{H}_B$ respectively. The Haar random state is then represented as

$$|\Psi\rangle = \mathcal{N} \sum_{iJ} X_{iJ} \, |i\rangle_A \otimes |J\rangle_B \,, \tag{3.4}$$

where X_{iJ} are complex Gaussian independent and identically distributed matrix elements of $d_A \times d_B$ matrix X with (unnormalized) joint probability distribution

$$P(\{X_{iJ}\}) \propto \exp\left[-d_A d_B \mathrm{Tr}\left(X X^\dagger\right)\right], \tag{3.5}$$

and $\mathcal{N}$ is a normalization constant. The reduced density matrix on $\mathcal{H}_A$ is then [198, 178, 197]

$$\rho_A = \frac{XX^\dagger}{\mathrm{Tr}\left[XX^\dagger\right]}. \tag{3.6}$$

Ensemble averages over α copies are given by the same formula in terms of permutations, so at large dimensions, $\rho_A \simeq XX^\dagger$. This is the famous Wishart-Laguerre ensemble and is equivalent to the previously introduced Haar random states [145, 35]. The advantage of working with random Gaussian states instead of Haar random states is due to "Wick calculus" being simpler than "Weingarten calculus." The difference will appear for random tensor networks although the ensembles will still be equivalent at large Hilbert space dimension [37]. Moreover, the class of random tensor networks used for holography involving projected Haar random states [81] precisely correspond to the states we study even at finite Hilbert space dimension, as explained in Section 3.9.1.

We now introduce a diagrammatic approach for computations of certain moments of the Wishart ensemble involving multiple states, building on Refs. [27, 99, 176]. This will prove invaluable in the following calculations.

We represent the elements of the random global pure state as two vertical lines

$$|\Psi\rangle_{iJ} = X_{iJ} := \quad \overset{iJ}{\vdots} \quad , \tag{3.7}$$

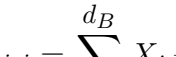
where the solid line represents $\mathcal{H}_A$ and the dashed line $\mathcal{H}_B$. To form the density matrix, we take the outer product

$$\left[|\Psi\rangle\langle\Psi|\right]_{iJ,jK} = X_{iJ}X^*_{jK} = \quad \overset{iJ}{\vdots} \qquad \overset{Kj}{\vdots} \quad . \tag{3.8}$$

We will usually drop the index labeling of the lines to avoid cumbersome notation. All matrix manipulations are done on the lower ends of the lines. For example, we can take a

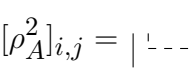
partial trace over $\mathcal{H}_B$ by connecting the dashed line

$$[\rho_A]_{i,j} = \sum_{J=1}^{d_B} X_{iJ} X^*_{jJ} := \quad , \tag{3.9}$$



square the matrix by taking two copies and connecting the *bra* of the first matrix with the *ket* of the second

$$[\rho_A^2]_{i,j} = \quad , \tag{3.10}$$



then take a trace by connecting the remaining solid lines to determine the purity

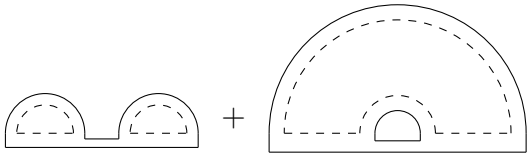
$$\mathrm{Tr}\left[\rho_A^2\right] = \quad . \tag{3.11}$$

For every insertion of the density matrix, we include a factor of $(d_A d_B)^{-1}$. This will give the normalization factor that we computed from (3.3).

The ensemble averaging of the states are done on the upper ends of the lines. The rule here is that we must add up all diagrams contracting any *bra* with any *ket*. For α insertions of the density matrix, there will be $\alpha!$ diagrams, corresponding to the $\alpha!$ allowed permutations. Within each diagram, we count the number of loops with each loop giving a factor of the Hilbert space dimension. One can see that this diagrammatic sum is precisely the numerator of (3.2).

We can now practice by taking the ensemble averaged purity. There are two (2!) diagrams descending from (3.11)

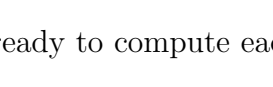
$$\overline{\mathrm{Tr}\left[\rho_A^2\right]} = \quad + \quad , \tag{3.12}$$



immediately leading to $d_A^{-1} + d_B^{-1}$.

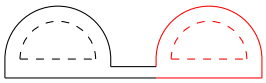
Because we are interested in distinguishing density matrices that are *independently* sampled from the ensemble, we must extend the diagrammatic technique. We do this by introducing different colors for different density matrices. When ensemble averaging, *bra*'s of one color can only contract with *ket*'s of the same color. For example, the overlap between independent induced states ρ_A (black) and σ_A (red) looks similar to the purity

$$\mathrm{Tr}\left[\rho_A \sigma_A\right] = \quad , \tag{3.13}$$



but the ensemble averaging will only include a single diagram

$$\overline{\mathrm{Tr}\left[\rho_A \sigma_A\right]} = \quad = \frac{1}{d_A} \tag{3.14}$$



because the second diagram would have connected the black and red indices which is disallowed. With this formalism, we are now ready to compute each distinguishability measures using a replica trick.

3.1 Relative entropy

We begin with the von Neumann relative entropy, both because it is the most fundamental quantity and the simplest to compute using our techniques. This will illustrate our strategy that will be used throughout. The relative entropy may be computed using a replica trick. That is, we first compute a certain series of moments of the ensemble and then analytically continue to arrive at the desired quantity. The replica trick for the relative entropy is given by [116]

$$D(\rho_A || \sigma_A) = \lim_{\alpha \to 1} \frac{1}{\alpha - 1} \left(\log \left[\mathrm{Tr}\left[\rho_A^\alpha\right] \right] - \log \left[\mathrm{Tr}\left[\rho_A \sigma_A^{\alpha - 1} \right] \right] \right). \tag{3.15}$$

We compute the ensemble average of the two terms separately. The first term is the Rényi entropy and, as a diagram, looks like

$$\mathrm{Tr}\left[\rho_A^\alpha\right] =$$  (3.16)

While we make the dimensions of the sub-Hilbert spaces large, $d_A, d_B \to \infty$, we keep their relative sizes, d_A/d_B, finite. The leading diagrams maximize the total number of loops. These are the planar diagrams as this double line notation corresponds the standard large-N topological expansion. Planar diagrams correspond to the *non-crossing permutations*, NC_α, a well-studied object in enumerative combinatorics and probability theory [105, 177]. The ensemble averaged Rényi purity is then given by

$$\overline{\mathrm{Tr}\left[\rho_A^\alpha\right]} = \frac{1}{(d_A d_B)^\alpha} \sum_{\tau \in NC_\alpha} d_A^{C(\eta^{-1}\circ\tau)} d_B^{C(\tau)}, \tag{3.17}$$

where η is the cyclic permutation, spawning from the matrix multiplication and trace in (3.16). The non-crossing permutations maximize the total exponent as $C(\eta^{-1}\circ\tau) + C(\tau) = \alpha + 1$ [105, 177]. A more refined statement is that the number of non-crossing permutations with $C(\eta^{-1}\circ\tau) = k$ (and therefore $C(\tau) = \alpha + 1 - k$) is given by the Narayana number [105, 177]

$$N_{\alpha,k} := \frac{1}{\alpha}\binom{\alpha}{k}\binom{\alpha}{k-1}. \tag{3.18}$$

With this information, we can reorganize (3.17) as a sum over k instead of a sum over permutations

$$\overline{\mathrm{Tr}\left[\rho_A^\alpha\right]} = \frac{1}{(d_A d_B)^\alpha} \sum_{k=1}^{\alpha} N_{\alpha,k} d_A^k d_B^{\alpha+1-k}, \tag{3.19}$$

which can be rewritten again as a hypergeometric function[1]

$$\overline{\mathrm{Tr}\left[\rho_A^\alpha\right]} = \begin{cases} d_A^{1-\alpha}\, {}_2F_1\left(1-\alpha,-\alpha;2;\frac{d_A}{d_B}\right), & d_A < d_B \\ d_B^{1-\alpha}\, {}_2F_1\left(1-\alpha,-\alpha;2;\frac{d_B}{d_A}\right), & d_A > d_B \end{cases}. \tag{3.20}$$

The $A \leftrightarrow B$ symmetry of Rényi entropies of bipartite pure states is manifest.

Taking the logarithm and analytically continuing to $\alpha = 1$, we obtain Page's formula [154]

$$\lim_{\alpha\to 1} \frac{1}{1-\alpha} \overline{\log\left[\mathrm{Tr}\left[\rho_A^\alpha\right]\right]} = \begin{cases} \log\left[d_A\right] - \frac{d_A}{2d_B}, & d_A < d_B \\ \log\left[d_B\right] - \frac{d_B}{2d_A}, & d_B < d_A \end{cases}. \tag{3.21}$$

In writing this formula, we have assumed that logarithm and ensemble average commute, which we justify in Section 3.8.

The second term in (3.15) involves both ρ_A and σ_A[2]

$$\mathrm{Tr}\left[\rho_A \sigma_A^{1-\alpha}\right] =$$  . (3.22)

Because there is only a single copy of ρ_A, when ensemble averaging, we must contract the first density matrix with itself. There are no constraints on how to contract the red lines.

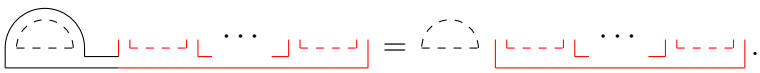
1. The two elements of the piecewise function are equivalent on the integers. The reason why we write it as a piecewise function is for ease of analytic continuation to non-integer values because the hypergeometric functions are entire when the argument is less than one.

2. This can be generalized such that the auxiliary systems for σ and ρ are of different sizes d_{B_1} and d_{B_2}. In the diagrammatics, this corresponds to assigning different weights to the black and red dashed lines. The resulting generalized sums are still tractable, though, we do not currently have use for these calculations in our applications to black holes because d_B corresponds to the size of the black hole, which is simple to measure by an outside observer, rendering σ_A and ρ_A easily distinguishable when $d_{B_1} \neq d_{B_2}$. Some exact results for this set up in the Wishart ensemble can be found in Refs. [113, 114, 115].

This means that the S_α permutations are broken down to $\mathbb{1} \times S_{\alpha-1}$

$$\overline{\mathrm{Tr}\left[\rho_A \sigma_A^{1-\alpha}\right]} = \frac{1}{(d_A d_B)^\alpha} \sum_{\tau \in \mathbb{1} \times S_{\alpha-1}} d_A^{C(\eta^{-1}\circ\tau)} d_B^{C(\tau)}. \tag{3.23}$$

We still need to maximize the exponent by choosing non-crossing permutations, though many such permutations are disallowed by the identity factor on the first matrix. The diagrams are topological, so we have



(3.24)

From this diagram, it is clear that the cardinality of the intersection of NC_α and $\mathbb{1} \times S_{\alpha-1}$ is given by the cardinality of $NC_{\alpha-1}$ and the number of such non-crossing permutations with $C(\eta^{-1} \circ \tau) = k$ is given by Narayana number $N_{\alpha-1,k}$

$$\overline{\mathrm{Tr}\left[\rho_A \sigma_A^{1-\alpha}\right]} = \frac{1}{(d_A d_B)^\alpha} \sum_{k=1}^{\alpha} N_{\alpha-1,k} d_A^k d_B^{\alpha+1-k}, \tag{3.25}$$

which can also be represented by a hypergeometric function

$$\overline{\mathrm{Tr}\left[\rho_A \sigma_A^{\alpha-1}\right]} = \begin{cases} d_A^{1-\alpha}\, {}_2F_1\left(1-\alpha, 2-\alpha; 2; \frac{d_A}{d_B}\right), & d_A < d_B \\ d_B^{2-\alpha} d_A^{-1}\, {}_2F_1\left(1-\alpha, 2-\alpha; 2; \frac{d_B}{d_A}\right), & d_A > d_B \end{cases}. \tag{3.26}$$

Taking the $\alpha \to 1$ limit, we have

$$\lim_{\alpha\to 1} \frac{1}{1-\alpha} \overline{\log\left[\mathrm{Tr}\left[\rho_A \sigma_A^{\alpha-1}\right]\right]} = \begin{cases} \log\left[d_A\right] + 1 + \left(\frac{d_B}{d_A} - 1\right) \log\left[1 - \frac{d_A}{d_B}\right], & d_A < d_B \\ \infty, & d_B < d_A \end{cases}. \tag{3.27}$$

Therefore, we find the ensemble average of the relative entropy to be

$$\overline{D(\rho_A||\sigma_A)} = \begin{cases} 1 + \frac{d_A}{2d_B} + \left(\frac{d_B}{d_A} - 1\right) \log\left[1 - \frac{d_A}{d_B}\right], & d_A < d_B \\ \infty, & d_A > d_B \end{cases}. \tag{3.28}$$

This is a satisfying, simple answer. For small d_A/d_B, the relative entropy is given by d_A/d_B. If we think in terms of "number of qubits," N_A and N_B, this is exponentially small in the difference $(N_B - N_A)$, meaning that the states will be very difficult to distinguish whenever we have access to a few qubits less than half the system; the asymptotic error rate, $\beta_n^*(\epsilon)$, is very small, meaning we will need exponentially (in N_A) many copies of the state to identify it with confidence. (3.28) is also monotonically increasing in d_A/d_B, a consequence of the data processing inequality when we take the partial trace as the quantum channel. When $d_A \to d_B$, the relative entropy approaches the curious value of 3/2. This value of 3/2 was also determined in Ref. [167] using very different techniques which serves as an additional consistency check of our results.

When $d_A > d_B$, every reduced state on $\mathcal{H}_A$ in the ensemble will be rank deficient with $d_A - d_B$ zero eigenvalues. This is because the Wishart ensemble has rank at most $\min(d_A, d_B)$. It is therefore overwhelmingly unlikely that two independent states, ρ_A and σ_A, will have the same support. In particular, the support of ρ_A will not be contained within the support of σ_A. This is the reason why the relative entropy becomes infinite in this regime; there will be a measurement we can choose that easily distinguishes ρ_A and σ_A.

3.2 Petz Rényi relative entropy and Holevo's just-as-good fidelity

To understand more sophisticated structures in Haar random states, we progress to the computation of the PRRE. The PRRE has a tricky $1 - \alpha$ exponent for σ_A, so we use a

replica trick with two replica parameters, α and m

$$D_\alpha(\rho_A||\sigma_A) = \lim_{m\to 1-\alpha} \frac{1}{\alpha-1} \log\left[\mathrm{Tr}\left[\rho_A^\alpha \sigma_A^m\right]\right]. \tag{3.29}$$

We will compute this for $\alpha, m \in \mathbb{Z}^+$, only taking the limit to $\alpha, m \in \mathbb{R}$ at the end of the calculation. The positive integer moments in diagrammatic form are

$\mathrm{Tr}\left[\rho_A^\alpha \sigma_A^m\right] =$  , (3.30)

where there are α black density matrices and m red density matrices. When ensemble averaging, we are only able to contract using the subgroup $S_\alpha \times S_m \subset S_{\alpha+m}$, leading to the sum over permutations

$$\overline{\mathrm{Tr}\left[\rho_A^\alpha \sigma_A^m\right]} = \frac{1}{(d_A d_B)^{\alpha+m}} \sum_{\tau \in S_\alpha \times S_m} d_A^{C(\eta^{-1}\circ\tau)} d_B^{C(\tau)}. \tag{3.31}$$

As can be seen by the diagram, even with the restricted sum, there are many ways to contract the lines that are non-crossing, hence maximizing the exponents. These are precisely the non-crossing permutations acting independently on the black and red indices, so the combinatorial factor will be given by the product of two Narayana numbers

$$\overline{\mathrm{Tr}\left[\rho_A^\alpha \sigma_A^m\right]} = \frac{1}{(d_A d_B)^{\alpha+m}} \sum_{k=1}^{\alpha} \sum_{j=1}^{m} N_{\alpha,k} N_{m,j} d_A^{k+j-1} d_B^{2+\alpha+m-k-j}. \tag{3.32}$$

The reason why there is an additional "-1" in the exponent of d_A is that the black and red lines are connected at the bottom of the diagram due to the matrix multiplication. Note that this expression is a generalization of the replica trick used in the previous section for the relative entropy, (3.25), if we set $\alpha = 1$ and $m = \alpha - 1$. As before, the double sum can

be expressed in terms of hypergeometric functions

$$\overline{\mathrm{Tr}\left[\rho_A^\alpha \sigma_A^m\right]} = \begin{cases} d_A^{-m-\alpha+1}\, {}_2F_1\left(1-m,-m;2;\frac{d_A}{d_B}\right)\, {}_2F_1\left(1-\alpha,-\alpha;2;\frac{d_A}{d_B}\right), & d_A < d_B \\ \frac{d_B^{-m-\alpha+2}}{d_A}\, {}_2F_1\left(1-m,-m;2;\frac{d_B}{d_A}\right)\, {}_2F_1\left(1-\alpha,-\alpha;2;\frac{d_B}{d_A}\right), & d_A > d_B \end{cases}. \tag{3.33}$$

Now that the sum that required m to be an integer is complete, it is safe to take the $m \to 1-\alpha$ limit

$$\overline{\mathrm{Tr}\left[\rho_A^\alpha \sigma_A^{1-\alpha}\right]} = \begin{cases} {}_2F_1\left(1-\alpha,-\alpha;2;\frac{d_A}{d_B}\right)\, {}_2F_1\left(\alpha-1,\alpha;2;\frac{d_A}{d_B}\right), & d_A < d_B \\ \frac{d_B}{d_A}\, {}_2F_1\left(1-\alpha,-\alpha;2;\frac{d_B}{d_A}\right)\, {}_2F_1\left(\alpha-1,\alpha;2;\frac{d_B}{d_A}\right), & d_A > d_B \end{cases}. \tag{3.34}$$

When $d_A = d_B$, this precisely agrees with a formula from Ref. [167]. Taking the logarithm leads to an exact closed-form expression for the PRRE in the large Hilbert space dimension limit

$$\overline{D_\alpha(\rho_A||\sigma_A)} = \frac{1}{\alpha-1} \begin{cases} \log\left[{}_2F_1\left(1-\alpha,-\alpha;2;\frac{d_A}{d_B}\right)\, {}_2F_1\left(\alpha-1,\alpha;2;\frac{d_A}{d_B}\right)\right], & d_A < d_B \\ \log\left[\frac{d_B}{d_A}\, {}_2F_1\left(1-\alpha,-\alpha;2;\frac{d_B}{d_A}\right)\, {}_2F_1\left(\alpha-1,\alpha;2;\frac{d_B}{d_A}\right)\right], & d_A > d_B \end{cases}. \tag{3.35}$$

This is a rare instance where we have an exact closed-form solution for relative entropies and can be thought of as the "Page formula" for PRRE. Importantly, this equation contains much more information about random quantum states than (3.28). A highlight is the finiteness of (3.35) for $\alpha < 1$ in the $d_A > d_B$ regime. This explains the approach of random quantum states to complete distinguishability. There are a few consistency checks that we can readily verify. Namely, we note that (3.35) reduces to (3.28) if we send $\alpha \to 1$, (3.35) is monotonically increasing in d_A/d_B (data processing inequality), and monotonically increasing in α.

An additional desirable property of (3.35) is that it is simple enough that we can perform

the optimization needed to compute the quantum Chernoff distance

$$\overline{\xi(\rho_A||\sigma_A)} = \begin{cases} -2\log\left[{}_2F_1\left(\frac{1}{2},-\frac{1}{2};2;\frac{d_A}{d_B}\right)\right], & d_A < d_B \\ -\log\left[\frac{d_B}{d_A}\,{}_2F_1\left(\frac{1}{2},-\frac{1}{2};2;\frac{d_B}{d_A}\right)^2\right], & d_A > d_B \end{cases}, \tag{3.36}$$

where the optimal value of α in (2.38) is found to be $1/2$. This definitively establishes the error rate in quantum state discrimination for a measure one set of quantum states. Because $\alpha = 1/2$ is the optimal value, this adds to the usefulness of Holevo's just-as-good fidelity, which is given by

$$\overline{F_H(\rho_A||\sigma_A)} = \begin{cases} {}_2F_1\left(\frac{1}{2},-\frac{1}{2};2;\frac{d_A}{d_B}\right)^4, & d_A < d_B \\ \frac{d_B^2}{d_A^2}\,{}_2F_1\left(\frac{1}{2},-\frac{1}{2};2;\frac{d_B}{d_A}\right)^4, & d_A > d_B \end{cases}. \tag{3.37}$$

In order to evaluate the quantum multiple Chernoff distance, we need to characterize the fluctuations in the PRRE. To compute the variance, we must compute

$$\mathrm{Tr}\left[\rho_A^\alpha \sigma_A^m\right]^2 = \quad . \tag{3.38}$$

Only the diagrams that connect the two blocks will contribute to the variance because the disconnected diagrams are subtracted. For small d_A/d_B, these contributions will be $O(d_A^{1-2\alpha-2m}d_B^{-1})$ or $O(d_A^{-1}d_B^{-1})$ after taking the relevant limit. We may use a Taylor expansion of the logarithm to determine that the variance of the PRRE, σ^2, will be the same order. The higher degree central moments, and therefore higher cumulants, will be subleading because in general, the n^{th} central moment will be $O(d_A^{1-n\alpha-nm}d_B^{-1})$. Therefore the PRRE will follow a normal distribution at subleading order.

For a normal distribution, the probability of random variable X being r standard devia-

tions, σ, below the mean, μ, is

$$\Pr(X - \mu \leq -r\sigma) = \int_{-\infty}^{-k\sigma} dx \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{x^2}{2\sigma^2}} = \frac{1}{2}\text{erfc}\left(\frac{r}{\sqrt{2}}\right). \tag{3.39}$$

Therefore, if we have W independent samplings of ρ_A and σ_A, the probability that the minimum relative entropy will be at most r standard deviations from the mean is

$$\Pr\left(\min\left[D_\alpha(\rho_A||\sigma_A)\right] \geq \mu - r\sigma\right) = \left(1 - \frac{1}{2}\text{erfc}\left(\frac{r}{\sqrt{2}}\right)\right)^W \tag{3.40}$$

If we are discriminating between W states, the quantum multiple Chernoff distance will be

$$\xi_W(\rho_A||\sigma_A) \geq \frac{d_A}{4d_B} - \sqrt{2}\text{erfc}^{-1}\left(2 - 2(1-\epsilon_1)^{\frac{1}{W}}\right)\sigma \tag{3.41}$$

with probability $1 - \epsilon_1$. In order to be confident in the state discrimination ($P_{err}^* < \epsilon_2$), we need

$$n \simeq \frac{\log\left[\epsilon_2^{-1}\right]}{\frac{d_A}{4d_B} - \sqrt{2}\text{erfc}^{-1}\left(2 - 2(1-\epsilon_1)^{\frac{1}{W}}\right)\sigma} \tag{3.42}$$

copies of the state. Due to σ being suppressed in the total Hilbert space dimension, this formula only mildly depends on W even when W is of order the Hilbert space dimension. Thus, the multiple Chernoff bound is essentially just as tight as the two-state Chernoff bound.

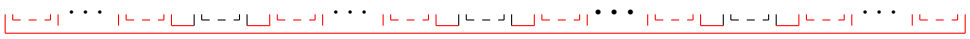
3.3 Sandwiched Rényi relative entropy and Uhlmann fidelity

Continuing our progression in difficulty, we now compute the SRRE using a new replica trick requiring two replica indices

$$\tilde{D}_\alpha(\rho_A||\sigma_A) := \lim_{m\to\frac{1-\alpha}{2\alpha}} \frac{1}{\alpha-1}\log\left[\mathrm{Tr}\left[\left(\sigma_A^m\rho_A\sigma_A^m\right)^\alpha\right]\right]. \tag{3.43}$$

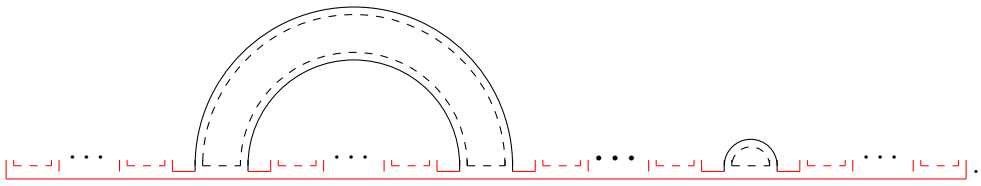
The associated diagrams are more complicated because there the red and black lines are not cleanly partitioned

$$\mathrm{Tr}\left[\left(\sigma_A^m\rho_A\sigma_A^m\right)^\alpha\right] =$$



. (3.44)

There are still many ways to contract the above diagram without crossing lines. We take two steps. First, we need to have the black lines contract with themselves in a non-crossing manner. For example, we may have



. (3.45)

This gives a factor of $d_B^{C(\tau)}$ where τ is the non-crossing permutation of the α black lines. We can see from this diagram that depending on how the black lines are contracted, this restricts the allowed permutations for the red lines. This is why this computation is more complicated than for the PREE where the black and red permutations simply factorized as $NC_\alpha \times NC_m$. In order for the global permutation to be non-crossing, the red permutations must be non-crossing within each block partitioned off by the black permutations. In the above example diagram, the black "rainbow" restricts the red permutation to be of the form

$NC_{2m} \times \ldots$. The identity permutation on the black density matrix on the right places no additional restrictions. In terms of equations, the diagrams may be summed as

$$\overline{\mathrm{Tr}\left[\left(\sigma_A^m \rho_A \sigma_A^m\right)^\alpha\right]} = \frac{1}{(d_A d_B)^{\alpha(2m+1)}} \sum_{\tau \in NC_\alpha} d_B^{C(\tau)} \prod_{\zeta \in cyc(\eta^{-1}\circ\tau)} \sum_{\gamma \in NC_{2m|\zeta|}} d_A^{C(\eta^{-1}\circ\gamma)} d_B^{C(\gamma)} \tag{3.46}$$

where the product is over the cycles of $\eta^{-1} \circ \tau$ and $|\cdot|$ represents the length of the cycle.

We first focus on the $d_A < d_B$ regime where the inner sum may be computed as before

$$\begin{aligned}\overline{\mathrm{Tr}\left[\left(\sigma_A^m \rho_A \sigma_A^m\right)^\alpha\right]} &= \frac{\sum_{\tau \in NC_\alpha} d_B^{C(\tau)} \prod_{\zeta \in cyc(\eta^{-1}\circ\tau)} d_A d_B^{2|\zeta|m}\, {}_2F_1\left(-2|\zeta|m, 1-2|\zeta|m; 2; \frac{d_A}{d_B}\right)}{(d_A d_B)^{\alpha(2m+1)}} \\ &= \frac{\sum_{\tau \in NC_\alpha} d_A^{\alpha+1-C(\tau)} d_B^{C(\tau)+2m\alpha} \prod_{\zeta \in cyc(\eta^{-1}\circ\tau)} {}_2F_1\left(-2|\zeta|m, 1-2|\zeta|m; 2; \frac{d_A}{d_B}\right)}{(d_A d_B)^{\alpha(2m+1)}},\end{aligned} \tag{3.47}$$

where in the second line, we have pulled out the factors of d_A and d_B from the product by enforcing the global permutation to be noncrossing. This formula does not need m to be an integer, so it is now safe to take the $m \to \frac{1-\alpha}{2\alpha}$ limit

$$\begin{aligned}&\overline{\mathrm{Tr}\left[\left(\sigma_A^{\frac{1-\alpha}{2\alpha}} \rho_A \sigma_A^{\frac{1-\alpha}{2\alpha}}\right)^\alpha\right]} \\ &\quad = \sum_{\tau \in NC_\alpha} \left(\frac{d_A}{d_B}\right)^{\alpha - C(\tau)} \prod_{\zeta \in cyc(\eta^{-1}\circ\tau)} {}_2F_1\left(\frac{|\zeta|(\alpha-1)}{\alpha}, 1+|\zeta|-\frac{|\zeta|}{\alpha}; 2; \frac{d_A}{d_B}\right).\end{aligned} \tag{3.48}$$

The product over cycle structures makes this formula still very difficult. Fortunately, Kreweras solved exactly this combinatorial problem about cycle structure in his landmark paper on non-crossing partitions [105]. He found that the number of non-crossing permutations of $\{1, 2, \ldots, \alpha\}$ with cycle structure[3] $\{m_i\}$ is given by, what we will call, the Kreweras number

3. This notation means that there are m_i cycles of length i.

[105, 177]

$$K_{\{m_i\}} := \frac{\alpha!}{(\alpha - b + 1)! m_1! \dots m_\alpha!}, \quad b := \sum_i m_i. \tag{3.49}$$

Therefore, we can reorganize the sum such that there are no more references to permutations, only natural numbers

$$\begin{aligned} &\overline{\mathrm{Tr}\left[\left(\sigma_A^{\frac{1-\alpha}{2\alpha}} \rho_A \sigma_A^{\frac{1-\alpha}{2\alpha}}\right)^\alpha\right]} \\ &\quad = \sum_{m_1,\dots,m_\alpha=0}^{\alpha} K_{\{m_i\}} \left(\frac{d_A}{d_B}\right)^{\sum_i m_i - 1} \prod_{m_i \neq 0} {}_2F_1\left(\frac{i(\alpha-1)}{\alpha}, 1 + i - \frac{i}{\alpha}; 2; \frac{d_A}{d_B}\right)^{m_i}. \end{aligned} \tag{3.50}$$

This formula still presents a daunting task to evaluate in terms of elementary functions for generic α, though it provides a tractable, controlled expansion in d_A/d_B. This is because, for small d_A/d_B, the hypergeometric function is close to one. We then must consider the smallest values of $\sum_i m_i$. First, we take only the leading term with $\sum_i m_i = 1$ (cyclic permutation)

$$\overline{\mathrm{Tr}\left[\left(\sigma_A^{\frac{1-\alpha}{2\alpha}} \rho_A \sigma_A^{\frac{1-\alpha}{2\alpha}}\right)^\alpha\right]} = {}_2F_1\left(\alpha - 1, \alpha; 2; \frac{d_A}{d_B}\right) + O\left(\frac{d_A}{d_B}\right). \tag{3.51}$$

This is not terribly useful because, as explained above, to this order, the RHS is exactly one, which would lead to the SRRE being identically zero. To find a nontrivial result, we need the next term where $\sum_i m_i = 2$ which can be achieved in many ways. These are the all the

ways to sum to integers between 1 and $\alpha - 1$ to α

$$\overline{\mathrm{Tr}\left[\left(\sigma_A^{\frac{1-\alpha}{2\alpha}}\rho_A\sigma_A^{\frac{1-\alpha}{2\alpha}}\right)^\alpha\right]} = \frac{d_A}{d_B}\sum_{j=1}^{\lfloor\frac{\alpha}{2}\rfloor} K_{\{m_j,m_{\alpha-j}=1\}}\,{}_2F_1\left(\frac{j(\alpha-1)}{\alpha}, 1+j-\frac{j}{\alpha}; 2; \frac{d_A}{d_B}\right)$$
$$\times\ {}_2F_1\left(\frac{(\alpha-j)(\alpha-1)}{\alpha}, 1+(\alpha-j)-\frac{\alpha-j}{\alpha}; 2; \frac{d_A}{d_B}\right) +\ {}_2F_1\left(\alpha-1, \alpha; 2; \frac{d_A}{d_B}\right)$$
$$+ O\left(\frac{d_A}{d_B}\right)^2, \tag{3.52}$$

where the floor function in the sum ensures that we do not double count. The Kreweras number is

$$K_{\{m_j,m_{\alpha-j}=1\}} = \begin{cases} \alpha, & j \neq \frac{\alpha}{2} \\ \frac{\alpha}{2}, & j = \frac{\alpha}{2} \end{cases}. \tag{3.53}$$

$j = \frac{\alpha}{2}$ will only occur when α is even. The exact form of the hypergeometric functions in the sum are not important at this order because for small d_A/d_B, they are all close to one. Therefore, only the Kreweras number is important. We can easily compute the sum at this order for any integer α and find that this parity effect disappears,

$$\overline{\mathrm{Tr}\left[\left(\sigma_A^{\frac{1-\alpha}{2\alpha}}\rho_A\sigma_A^{\frac{1-\alpha}{2\alpha}}\right)^\alpha\right]} = 1 + \alpha(\alpha-1)\frac{d_A}{d_B} + O\left(\frac{d_A}{d_B}\right)^2, \tag{3.54}$$

leading to an SRRE of

$$\overline{\tilde{D}_\alpha(\rho_A||\sigma_A)} = \alpha\frac{d_A}{d_B} + O\left(\frac{d_A}{d_B}\right)^2. \tag{3.55}$$

Note that this agrees with the previously derived von Neumann relative entropy in the relevant $\alpha \to 1$ limit. Moreover, it obeys the data processing inequality for all positive α if we take the quantum channel to be the partial trace.

The Uhlmann fidelity is found[4] by setting $\alpha = 1/2$

$$\overline{F(\rho_A||\sigma_A)} = 1 - \frac{d_A}{2d_B} + O\left(\frac{d_A}{d_B}\right)^2. \tag{3.56}$$

At this order, the Uhlmann fidelity is identical to Holevo's just-as-good fidelity (3.37). The value of the fidelity exactly at $d_A = d_B$ was found in Ref. [167] to be $\frac{9}{16}$ and additional results may be found in Ref. [199].

We can also evaluate the SRRE exactly for any integer moment using (3.50). Here, we work out the least tedious case of $\alpha = 2$ which is also known as the *collision relative entropy* [169]. In this case, (3.52) is actually exact and does not contain $O\left(\frac{d_A}{d_B}\right)^2$ corrections. We only sum over $j = 1$, so

$$\overline{\text{Tr}\left[\left(\sigma_A^{-\frac{1}{4}}\rho_A\sigma_A^{-\frac{1}{4}}\right)^2\right]} = \frac{d_A}{d_B}\,{}_2F_1\left(\frac{1}{2},\frac{3}{2};2;\frac{d_A}{d_B}\right)^2 + \frac{1}{1-\frac{d_A}{d_B}}, \tag{3.57}$$

leading to an SRRE of

$$\overline{\tilde{D}_2(\rho_A||\sigma_A)} = \begin{cases} \log\left[\frac{d_A}{d_B}\,{}_2F_1\left(\frac{1}{2},\frac{3}{2};2;\frac{d_A}{d_B}\right)^2 + \frac{1}{1-\frac{d_A}{d_B}}\right], & d_A < d_B \\ \infty, & d_A > d_B \end{cases}, \tag{3.58}$$

where we have set the SRRE to infinity when $d_A > d_B$ because the von Neumann relative entropy is infinite in this regime and the SRRE's are monotonically increasing with α. It is straightforward to evaluate the higher integer SRRE's if desired.

It is equally important to investigate the opposite regime where d_A/d_B is large. In this

4. We note that an exact expression was recently found for the fidelity of two random density matrices, consistent with our large-N results [115]. Our result is complementary as the exact expression is very complicated and not tractable at large-N.

case, the inner sum in (3.46) gives

$$\overline{\mathrm{Tr}\left[\left(\sigma_A^m \rho_A \sigma_A^m\right)^\alpha\right]} = \frac{\sum_{\tau \in NC_\alpha} d_B^{C(\tau)} \prod_{\zeta \in cyc(\eta^{-1}\circ\tau)} d_A^{2|\zeta|m} d_B \, {}_2F_1\left(-2|\zeta|m, 1-2|\zeta|m; 2; \frac{d_B}{d_A}\right)}{(d_A d_B)^{\alpha(2m+1)}}$$
$$= \frac{\sum_{\tau \in NC_\alpha} \left(\frac{d_B}{d_A}\right)^{C(\tau)+C(\eta^{-1}\circ\tau)} d_A^{\alpha(2m+1)+1} \prod_{\zeta \in cyc(\eta^{-1}\circ\tau)} {}_2F_1\left(-2|\zeta|m, 1-2|\zeta|m; 2; \frac{d_B}{d_A}\right)}{(d_A d_B)^{\alpha(2m+1)}}. \tag{3.59}$$

Now that we have done the sums over the m permutations, we can safely take $m \to \frac{1-\alpha}{2\alpha}$ and rewrite the sum in terms of Kreweras numbers

$$\overline{\mathrm{Tr}\left[\left(\sigma_A^{\frac{1-\alpha}{2\alpha}} \rho_A \sigma_A^{\frac{1-\alpha}{2\alpha}}\right)^\alpha\right]}$$
$$= \left(\frac{d_B}{d_A}\right)^\alpha \sum_{m_1,\ldots,m_\alpha=0}^{\alpha} K_{\{m_i\}} \prod_{m_i \neq 0} {}_2F_1\left(\frac{i(\alpha-1)}{\alpha}, 1+\frac{i(\alpha-1)}{\alpha}; 2; \frac{d_B}{d_A}\right)^{m_i}. \tag{3.60}$$

This is an exact formula, but is difficult to evaluate away from limits. For large d_A/d_B, all of the hypergeometric functions are close to one so all that matters is the total number of noncrossing permutations, which is given by the Catalan number

$$C_\alpha := \frac{1}{\alpha+1}\binom{2\alpha}{\alpha}. \tag{3.61}$$

Therefore, at leading order, we have

$$\overline{\mathrm{Tr}\left[\left(\sigma_A^{\frac{1-\alpha}{2\alpha}} \rho_A \sigma_A^{\frac{1-\alpha}{2\alpha}}\right)^\alpha\right]} = C_\alpha \left(\frac{d_B}{d_A}\right)^\alpha + O\left(\frac{d_B}{d_A}\right)^{\alpha+1}. \tag{3.62}$$

Unlike for small d_A/d_B, there are no additional terms at leading order. The SRRE is thus

$$\overline{\tilde{D}_\alpha(\rho_A||\sigma_A)} = \frac{\alpha}{\alpha-1}\log\left[\frac{d_B}{d_A}\right] + \frac{1}{\alpha-1}\log\left[C_\alpha\right] + O\left(\frac{d_B}{d_A}\right). \tag{3.63}$$

Important to note is that this is only well-defined for $\alpha < 1$. This is to be expected because of rank deficiency. In the well-defined regime, the SRRE is monotonic in α and manifestly obeys the data processing inequality.

We can evaluate the asymptotic expression at $\alpha = 1/2$ to find the Uhlmann fidelity

$$\overline{F(\rho_A||\sigma_A)} = \frac{64 d_B}{9\pi^2 d_A} + O\left(\frac{d_B}{d_A}\right)^{3/2}. \tag{3.64}$$

The prefactor comes from the Catalan number which is nonintegral for noninteger α. We see that the fidelity is inversely proportional to d_A/d_B, decaying to zero when subsystem A occupies most of the Hilbert space. The full spectrum of $\sigma_A^{\frac{1-\alpha}{2\alpha}} \rho_A \sigma_A^{\frac{1-\alpha}{2\alpha}}$, and hence the fidelity, may be evaluated using techniques of free probability theory. This is completed in Section 3.6. The answer is the free multiplicative convolution of two Marchenko-Pastur distributions.

3.4 Trace distance

The final distinguishability measure we discuss is the trace distance. This is the ideal measure when discussing one-shot state discrimination (2.34). More general than (2.27), we can define an α-norm version of the trace distance

$$T_\alpha(\rho_A||\sigma_A) := \frac{1}{2^{1/\alpha}} |\rho_A - \sigma_A|_\alpha, \tag{3.65}$$

where the α-norm of an operator, A, is defined as

$$|A|_\alpha := \left(\mathrm{Tr}\left[\sqrt{A^\dagger A}^\alpha\right]\right)^{1/\alpha}. \tag{3.66}$$

For even α and Hermitian A, we can dispose of the square root

$$|A|_{\tilde{\alpha}} = \left(\mathrm{Tr}\left[A^{2\tilde{\alpha}}\right]\right)^{1/2\tilde{\alpha}}, \quad 2\tilde{\alpha} := \alpha. \tag{3.67}$$

The trace norm is then the $\tilde{\alpha} \to 1/2$ limit of this expression.

The replica trick, while only requiring a single replica parameter, is quite difficult as we must compute all even powers of $\rho_A - \sigma_A$ which involves arbitrary mixing of ρ_A and σ_A [195]

$$\mathrm{Tr}\left[(\rho_A - \sigma_A)^\alpha\right] = \sum_{\mathcal{S}\in\mathcal{P}(\{1,2,\ldots,\alpha\})} (-1)^{|\mathcal{S}|}\mathrm{Tr}(\tau_{\mathcal{S}_1}\ldots\tau_{\mathcal{S}_\alpha}), \tag{3.68}$$

where the sum runs over the power set of $\{1, 2, \ldots, \alpha\}$ and $|\cdot|$ is the cardinality of the subset. We have $\tau_{\mathcal{S}_i} = \rho_A$ if $i \in \mathcal{S}$ and $\tau_{\mathcal{S}_i} = \sigma_A$ if $i \notin \mathcal{S}$. Each term in the sum can be expressed as an appropriate summation over the symmetric group, though this is far from straightforward.

Consider the small d_A/d_B limit. In this case, the terms that maximize $C(\tau)$ will dominate the sum. This is when τ is the identity. This permutation is always present, regardless of $\mathcal{S}$ and universally contributes as $d_A^{1-\alpha}$ which in the $\alpha \to 1$ limit contributes at $O(1)$. To see if this contributes to the overall sum, we need to understand the cardinalities. The number of subsets with cardinality k is given by the binomial coefficient, so the identity contributes as

$$\overline{\mathrm{Tr}\left[(\rho_A - \sigma_A)^\alpha\right]} \supset \sum_{k=0}^{\alpha} \binom{\alpha}{k} (-1)^k d_A^{1-\alpha} = 0. \tag{3.69}$$

To get a nontrivial answer, we must therefore move beyond the identity permutation. This is to be expected because the trace distance will be small for small d_A/d_B and should not be $O(1)$. The next leading term is when $C(\tau) = \alpha - 1$ which corresponds to the identity on all sites except for two which are swapped; this is always non-crossing. This contributes

universally as $d_A^{2-\alpha} d_B^{-1}$, which will lead to the $O(d_A/d_B)$ contribution. The combinatorics are slightly more complicated. If $|\mathcal{S}| = k$, then there are $\binom{k}{2} + \binom{\alpha-k}{2}$ ways to have a single pairing because we can only choose pairings within the block of k ρ_A's or $n-k$ σ_A's. Therefore, the contribution at this order is

$$\overline{\mathrm{Tr}\left[(\rho_A - \sigma_A)^\alpha\right]} \supset \sum_{k=0}^{\alpha} \binom{\alpha}{k} \left(\binom{k}{2} + \binom{\alpha-k}{2} \right) (-1)^k d_A^{2-\alpha} d_B^{-1} = 2\delta_{\alpha,2} d_A^{2-\alpha} d_B^{-1}. \quad (3.70)$$

This is not an analytic function, so the $\tilde{\alpha} \to 1/2$ limit is quite ambiguous. We are free to work to higher orders, though we will argue that this will not help our cause.

In the small d_B/d_A regime, the expansion is more involved. The leading terms come from maximizing $C(\eta^{-1} \circ \tau)$. We can only have $C(\eta^{-1} \circ \tau) = \alpha$ in the case that $\mathcal{S} = \{1, 2, \ldots, \alpha\}$ or is empty because otherwise, $\tau = \eta$ will not be an allowed permutation. At this order, we therefore have

$$\overline{\mathrm{Tr}\left[(\rho_A - \sigma_A)^\alpha\right]} = \begin{cases} 2d_B^{1-\alpha}, & \alpha \in 2\mathbb{Z} \\ 0 & \alpha \in 2\mathbb{Z}+1 \end{cases}. \quad (3.71)$$

The parity effect arises from the exponent of the sign in the sum. Analytically continuing the even integers to one, we find

$$\overline{T(\rho_A || \sigma_A)} = 1 + O\left(\frac{d_B}{d_A}\right), \quad (3.72)$$

meaning that the states are nearly maximally distant. To understand how the trace distance approaches one, we need to work at the next order. The noncrossing permutations that give $C(\eta^{-1} \circ \tau) = \alpha - 1$ are those that are of the form $\eta_{\alpha_1} \times \eta_{\alpha_2}$. This means that the ρ_A's and σ_A's must be in disjoint blocks i.e. $\mathcal{S}$ is a set only containing consecutive integers. There are $(\alpha - 1)$ ways to partition α into nonzero integers α_1 and α_2 if we define the tuple (α_1, α_2)

to be distinct from (α_2, α_1). There is an additional factor of α coming from the rotations of $\mathcal{S}$ to $\mathcal{S}+1$, leading to

$$\overline{\mathrm{Tr}\left[(\rho_A-\sigma_A)^\alpha\right]} \supset \sum_{k=1}^{\alpha-1} \alpha(-1)^k d_A^{-1} d_B^{2-\alpha} = \begin{cases} -\alpha d_A^{-1} d_B^{2-\alpha}, & \alpha \in 2\mathbb{Z} \\ 0 & \alpha \in 2\mathbb{Z}+1 \end{cases}. \tag{3.73}$$

Finally, when $\mathcal{S} = \{1, 2, \ldots, \alpha\}$ or $\mathcal{S} = \emptyset$, we again can partition the elements into α_1 and α_2 size blocks but this time (α_1, α_2) and (α_2, α_1) are indistinguishable. Therefore, for even α, there are $\alpha/2$ possibilities while for odd α, there are only $(\alpha-1)/2$ possibilities plus the rotation factors[5], leading to

$$\overline{\mathrm{Tr}\left[(\rho_A-\sigma_A)^\alpha\right]} \supset \begin{cases} \alpha(\alpha-1) d_A^{-1} d_B^{2-\alpha}, & \alpha \in 2\mathbb{Z} \\ 0 & \alpha \in 2\mathbb{Z}+1 \end{cases}, \tag{3.74}$$

where the odd terms are trivial because the $\mathcal{S} = \{1, 2, \ldots, \alpha\}$ and $\mathcal{S} = \emptyset$ terms exactly cancel in the sum due to the power of the sign. Taking the $\alpha \to 1$ limit of even α, we find the trace norm at this order to be

$$\overline{T(\rho_A||\sigma_A)} = 1 - \frac{d_B}{2d_A} + O\left(\frac{d_B}{d_A}\right)^2. \tag{3.75}$$

The trace distance may be evaluated away from limits using free probability techniques as we review in Section 3.6 [130]

$$\overline{T(\rho_A||\sigma_A)} = \begin{cases} \frac{\sqrt{d_A(2d_B-d_A)}(d_A+d_B)+d_B(4d_A-2d_B)\sin^{-1}\left(\sqrt{\frac{d_A}{2d_B}}\right)}{2\pi d_A d_B}, & d_A < 2d_B \\ 1-\frac{d_B}{2d_A}, & d_A > 2d_B \end{cases}. \tag{3.76}$$

5. The rotation factor for even α when $\alpha_1 = \alpha_2$ is only $\alpha/2$ because of indistinguishability.

Interestingly, our asymptotic formula, (3.75), was exact. The trace distance at $d_A = d_B$ was found in Ref. [167] to be $\frac{1}{2} + \frac{1}{\pi}$.

Without free probability, we can still use bounds from Section 2.2 to place strong constraints on the trace distance. In particular, this is helpful for the small d_A/d_B where we were unable to find an analytic answer. First, we use Pinsker's inequality (2.30), using the relative entropy (3.28) as an upper bound for $d_A < d_B$

$$\overline{T(\rho_A||\sigma_A)} \leq \sqrt{2 + \frac{d_A}{d_B} + 2\left(\frac{d_B}{d_A} - 1\right)\log\left(1 - \frac{d_A}{d_B}\right)}. \tag{3.77}$$

This is only a useful bound when the RHS is less than one. Recall that when $d_A = d_B$, the RHS will be $\sqrt{3} > 1$. For small d_A/d_B, we have

$$\overline{T(\rho_A||\sigma_A)} \leq \sqrt{\frac{2d_A}{d_B}} + O\left(\frac{d_A}{d_B}\right)^{3/2}. \tag{3.78}$$

This means that the trace distance is very small, though without a lower bound, we cannot yet say that we could not find the leading order trace distance from the above expansion. We can determine a lower bound using Holevo's just-as-good fidelity which is, in general, stronger than the lower bound from Uhlmann fidelity

$$1 - {}_2F_1\left(\frac{1}{2}, -\frac{1}{2}; 2; \frac{d_A}{d_B}\right)^2 \leq \overline{T(\rho_A||\sigma_A)} \leq \sqrt{1 - {}_2F_1\left(\frac{1}{2}, -\frac{1}{2}; 2; \frac{d_A}{d_B}\right)^4}, \tag{3.79}$$

where we have also included the upper bound that is, in general, weaker than the upper bound from Uhlmann fidelity. For small d_A/d_B, this gives

$$\frac{d_A}{4d_B} + O\left(\frac{d_A}{d_B}\right)^2 \leq \overline{T(\rho_A||\sigma_A)} \leq \sqrt{\frac{d_A}{2d_B}} + O\left(\frac{d_A}{d_B}\right)^{3/2}. \tag{3.80}$$

This is a stronger upper bound than from Pinsker's inequality, though the scaling is still

not nailed down, only constrained to between linear and square root with d_A/d_B. Because the scaling is at most linear, we cannot hope to find the leading order behavior from the expansion at the beginning of this subsection because the linear term was zero if we are to trust the "continuation." The Uhlmann fidelity for small d_A/d_B does not strengthen the upper bound at leading order.

We also want to characterize the $d_A > d_B$ regime. While Pinsker's inequality does not help here because the relative entropy is infinite, Holevo's just-as-good fidelity places nontrivial upper and lower bounds

$$1 - \frac{d_B}{d_A}\,{}_2F_1\left(\frac{1}{2}, -\frac{1}{2}; 2; \frac{d_B}{d_A}\right)^2 \leq \overline{T(\rho_A||\sigma_A)} \leq \sqrt{1 - \frac{d_B^2}{d_A^2}\,{}_2F_1\left(\frac{1}{2}, -\frac{1}{2}; 2; \frac{d_B}{d_A}\right)^4}. \tag{3.81}$$

For small d_B/d_A, this is

$$1 - \frac{d_B}{d_A} + O\left(\frac{d_B}{d_A}\right)^2 \leq \overline{T(\rho_A||\sigma_A)} \leq 1 - \frac{d_B^2}{2d_A^2} + O\left(\frac{d_B}{d_A}\right)^3, \tag{3.82}$$

meaning the states are almost as far away from each other as possible, approaching one exponentially in $N_A - N_B$. The upper bound can be improved by the Uhlmann fidelity at leading order such that the scaling behavior is completely fixed

$$1 - \frac{d_B}{d_A} + O\left(\frac{d_B}{d_A}\right)^2 \leq \overline{T(\rho_A||\sigma_A)} \leq 1 - \frac{32 d_B}{9\pi^2 d_A} + O\left(\frac{d_B}{d_A}\right)^2. \tag{3.83}$$

These bounds are consistent with the analytic expressions (3.75) and (3.76).

3.5 Interpolating between QSD and QHT

We have, so far, characterized the asymptotic error rates in distinguishing states in the totally symmetric (QSD) and totally asymmetric (QHT) cases with the quantum Chernoff distance and relative entropy respectively. It is natural to ask if the there is a way to

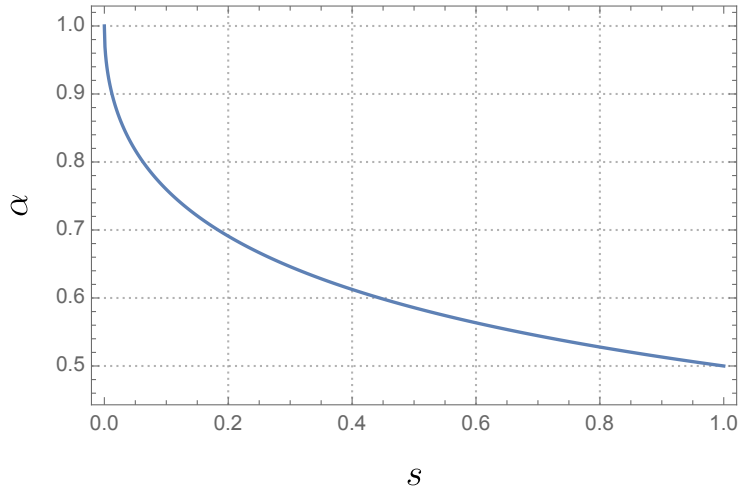


Figure 3.1: The interpolation between QHT ($s = 0$) and QSD ($s = 1$) is shown for $d_A < d_B$.

interpolate between the two. This was addressed in Ref. [174] where the type II error $\beta(A)$ was optimized given the constraint that $\alpha(A) \leq \varepsilon^{1-s}\beta(A)^s$ for $s \geq 0$ and $\varepsilon > 0$. Note that this coincides with QHT for $s = 0$ and QSD for $s = 1$. This is referred to as s-hypothesis testing and the error rate was proven to be given by the s-quantum divergence defined as

$$\xi_s(\rho||\sigma) := \max_{0\leq\alpha\leq 1} \frac{\log\left[\mathrm{Tr}\left[\rho^\alpha \sigma^{1-\alpha}\right]\right]}{\alpha(1-s)-1}. \tag{3.84}$$

It is instructive to examine the two familiar limits. When $s = 0$, the quantity being maximized is the PRRE. We know that the PRRE is monotonically increasing with α so $\xi_0(\rho||\sigma)$ is given by the relative entropy in accordance with our expectation. When $s = 1$, the RHS becomes the definition of the quantum Chernoff distance. Using (3.35), we can evaluate the s-quantum divergence for random states. We plot the value of α that maximizes the RHS as a function of s in Fig. 3.1. This function monotonically decreases from one at $s = 0$ to zero at $s = \infty$, passing through $\alpha = 1/2$ at $s = 1$.

3.6 Distinguishability from free probability theory

We have tried to avoid the use of free probability theory because while it is immensely powerful, it is very technical and unintuitive. Moreover, it lacks the clear connections to gravitational and general chaotic systems that the replica trick has. Alas, for exact calculations of the trace distance and Uhlmann fidelity, we were unable to push our diagrammatic techniques far enough so we address distinguishability measures using free probability now.

We begin a minimal introduction of the tools we will use from applications of free probability theory as applied to random matrices. The interested reader is encouraged to consult Refs. [146, 134]. First, we define a unital linear form, τ, acting on a vector space. For $N \times N$ matrices, X, this can be taken to be the normalized trace $\tau(X) := \frac{1}{N}\mathrm{Tr}\,[X]$. Random variables X and Y are *free* if they satisfy

$$\tau\left(P_1(X)Q_1(Y)\ldots P_k(X)Q_k(Y)\right) = 0, \tag{3.85}$$

for any set of polynomials $\{P, Q\}$ when $\tau\left(P_i(X)\right) = \tau\left(Q_i(Y)\right) = 0$. This, in turn, implies a factorization property

$$\tau\left(P(X)Q(Y)\right) = \tau\left(P(X)\right)\tau\left(Q(Y)\right), \tag{3.86}$$

which can be seen by taking $P_1(X) = P(X) - \tau(P(X))$ and $Q_1(X) = Q(X) - \tau(Q(X))$. There are several useful ways to package the moments of a free random variable, $m_n := \tau(X^n)$. These enable the evaluation of empirical spectral measures for random matrices. For a free random variable X, we define the moment function as a formal power series

$$M_X(z) := \sum_{m=0}^{\infty} \tau(X^n) z^m, \quad z \in \mathbb{C}. \tag{3.87}$$

This is related to the Green's function or *Cauchy transform* as

$$G_X(z) := \frac{1 + M_X(z)}{z}. \tag{3.88}$$

From the Cauchy transform, one may extract the spectral measure using a Stieltjes transformation

$$\mu_X(\lambda) = -\frac{1}{\pi} \lim_{\epsilon \to 0} \mathrm{Im}\left[G_X(\lambda + i\epsilon)\right]. \tag{3.89}$$

Sometimes, we know that random variables X and Y have spectral measures ρ_X and ρ_Y but want to know the spectral measure of their sum, $X + Y$, or product, XY. The spectral measure of their sum is defined as the free convolution $\mu_{X+Y} := \mu_X \boxplus \mu_Y$ while the spectral measure of their product is defined as the free multiplicative product $\mu_{XY} := \mu_X \boxtimes \mu_Y$. In order to obtain the free convolution, it is convenient to introduce the *R-transform*

$$R_X(G_X(z)) + \frac{1}{G_X(z)} := z. \tag{3.90}$$

The R-transform of a sum of free random variables is given by the sum of their individual R-transforms

$$R_{X+Y}(z) = R_X(z) + R_Y(z). \tag{3.91}$$

For the free multiplicative product, it is convenient to introduce the *S-tranform*

$$zM_X(z)S_X(M_X(z)) := 1 + M_X(z). \tag{3.92}$$

The S-tranform of a product of free random variables is given by the product of their indi-

vidual S-transforms

$$S_{XY}(z) = S_X(z)S_Y(z). \tag{3.93}$$

The key to the usefulness of free probability for us is that Wishart random matrices are free random variables asymptotically, as the dimensions become large. Their empirical spectral measure is given by the Marchenko-Pastur distribution

$$\mu_{MP}^{(c)}(x) = \max\left[1 - \frac{1}{c}, 0\right]\delta(x) + \frac{\sqrt{(x - c(1 - c^{-1/2})^2)(c(1 + c^{-1/2})^2 - x)}}{2\pi cx}, \tag{3.94}$$

where $c := d_A/d_B$ is the rectangular parameter and $x := d_A\lambda$.

Petz Rényi relative entropy For the PRRE, we need to evaluate $\overline{\mathrm{Tr}\left[\rho_A^\alpha \sigma_A^{1-\alpha}\right]}$. Because ρ_A and σ_A are asymptotically free random variables with Marchenko-Pastur distributions, $\mu_{MP}^{(c)}$, this factorizes according to (3.86)

$$\overline{\mathrm{Tr}\left[\rho_A^\alpha \sigma_A^{1-\alpha}\right]} = \frac{\overline{\mathrm{Tr}\left[\rho_A^\alpha\right]\mathrm{Tr}\left[\sigma_A^{1-\alpha}\right]}}{d_A} = \int \mu_{MP}^{(c)}(x)x^\alpha \int \mu_{MP}^{(c)}(x)x^{1-\alpha}, \tag{3.95}$$

These integrals may be evaluated to reproduce (3.34), the result from the replica trick.

Sandwiched Rényi relative entropy For the SRRE, we need to evaluate the averaged moments of $\sigma_A^{\frac{1-\alpha}{2\alpha}}\rho_A\sigma_A^{\frac{1-\alpha}{2\alpha}}$. Using the replica trick, we succeeded for integer α, but were unable to analytically continue to real-valued α in order to evaluate the Uhlmann fidelity at $\alpha = \frac{1}{2}$. Here, we evaluate the spectrum of $\sigma_A^{\frac{1}{2}}\rho_A\sigma_A^{\frac{1}{2}}$ (equivalently $\rho_A\sigma_A$) to accomplish this goal. This is the free multiplicative convolution of Marchenko-Pastur laws $\mu_{MP}^{(c)} \boxtimes \mu_{MP}^{(c)}$ and has been called a generalized Fuss-Catalan distribution [136]. The Fuss-Catalan distributions themselves were first derived in Ref. [159]. The S-tranform for the Marchenko-

Pastur distribution is given by

$$S_{\rho_A}(z) = \frac{1}{1 + cM_{\rho_A}(z)}. \tag{3.96}$$

Therefore, the S-transform for $\rho_A \sigma_A$ is

$$S_{\rho_A \sigma_A}(z) = \frac{1}{\left(1 + cM_{\rho_A}(z)\right)^2}. \tag{3.97}$$

Plugging this into (3.92), we find

$$zM_{\rho_A \sigma_A}(z) = (1 + M_{\rho_A \sigma_A}(z))(1 + cM_{\rho_A \sigma_A}(z))^2. \tag{3.98}$$

Taking the correct root of this cubic equation and taking the Stieltjes transformation, we find the spectral measure

$$\begin{aligned} \mu_{MP}^{(c)} \boxtimes \mu_{MP}^{(c)}(x) = \max\left[0, \frac{c-1}{c}\right] \delta(x) \\ - \frac{1}{\pi} \begin{cases} \frac{1}{\sqrt{3} \cdot 2^{2/3} x} \frac{3x + (c-1)^2}{\left(\sqrt{P} + A\right)^{1/3}} - \frac{\sqrt{3} \cdot 2^{2/3}}{12 c^2 x} \left(\sqrt{P} + A\right)^{1/3}, & (x_1, x_2) \\ 0, & \text{otherwise} \end{cases}, \end{aligned} \tag{3.99}$$

where we have defined

$$\begin{aligned} A &= c^3 \left[9(c+2)x + 2(c-1)^3\right] \\ P &= 27 c^6 x \left[\left(8 - (c-20)c\right) x + 4(c-1)^3 - 4x^2\right] \end{aligned} \tag{3.100}$$

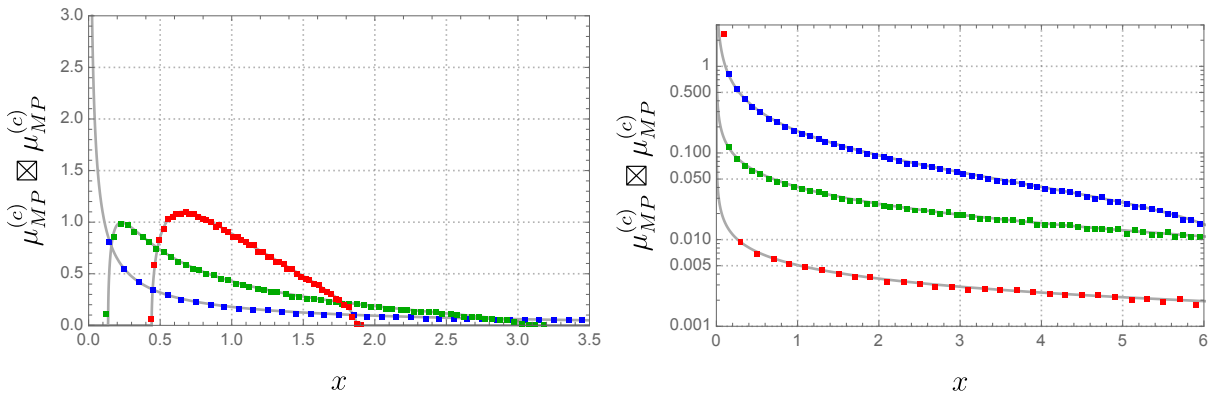


Figure 3.2: The probability density function for the generalized Fuss-Catalan distribution is shown (grey lines) with comparison to numerics. On the left, the blue, green and red dots correspond to $c = 2^0, 2^{-2}, 2^{-4}$ respectively. On the right, the blue, green and red dots correspond to $c = 2^0, 2^2, 2^4$ respectively. The total system size is 2^{16} and we disorder average over 10^3 realizations.

and end points

$$x_1 = \max\left[0, \frac{8 + 20c - c^2 - \sqrt{c}(8+c)^{3/2}}{8}\right]. \qquad (3.101)$$
$$x_2 = \frac{8 + 20c - c^2 + \sqrt{c}(8+c)^{3/2}}{8}$$

The Uhlmann fidelity is then given by the following integral

$$F(\rho_A||\sigma_A) = \left(\int_{x_1}^{x_2} \mu_{MP}^{(c)} \boxtimes \mu_{MP}^{(c)}(x)\sqrt{x}\right)^2. \qquad (3.102)$$

We compare the derived distribution with numerics in Fig. 3.2.

Trace distance The trace distance is defined using the trace norm of the difference of ρ_A and σ_A. This is tailor-made for a computation in free probability because of free convolution. The R-transform for the Marchenko-Pastur distribution is

$$\mathcal{R}_{\rho_A}(z) = \frac{1}{1 - cz}. \qquad (3.103)$$

Because we are taking the difference, we must rescale the second R-transform

$$\mathcal{R}_{-\sigma_A}(z) = -\mathcal{R}_{\sigma_A}(-z) = -\frac{1}{1+cz}. \tag{3.104}$$

Therefore, the R-transform of the difference is

$$\mathcal{R}_{\rho_A-\sigma_A}(z) = \frac{2cz}{1-c^2z^2}. \tag{3.105}$$

Inverting the R-transform, one finds the Cauchy transform

$$\frac{2cG(z)}{1-c^2G(z)^2} + \frac{1}{G(z)} = z. \tag{3.106}$$

Once again, one can take the correct root of the cubic equation and take the Stieltjes transformation to find the spectral measure [167, 130]

$$\mu_{MP}^{(c)} \boxplus \mathcal{D}_{-1}\left[\mu_{MP}^{(c)}\right](x) = \begin{cases} \frac{\sqrt{(2-c)^2+3x^2}}{\sqrt{3}\pi c|x|} \sinh\left[\frac{\log\left[\eta(x)+\sqrt{\eta^2(x)-1}\right]}{3}\right], & |x| \in (x_-, x_+) \\ \max\left[0, 1-\frac{2}{c}\right]\delta(x), & \text{otherwise} \end{cases}, \tag{3.107}$$

where $\mathcal{D}_{-1}$ represents the rescaling, the function $\eta(x)$ is given by

$$\eta(x) := \frac{9(c+1)^2x^2 + (2-c)^2}{((2-c)^2+3x^2)^{3/2}}, \tag{3.108}$$

and the endpoints of the spectrum are

$$x_\pm = \max\left[0, \frac{1}{4}\left(\sqrt{4c+1}\pm 3\right)^{3/2}\left(\sqrt{4c+1}\mp 1\right)^{1/2}\right]. \tag{3.109}$$

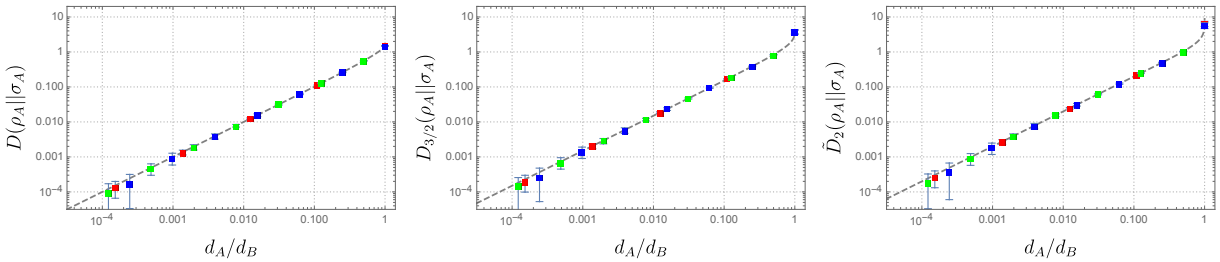


Figure 3.3: Three representative examples of relative entropies are shown, all of which with $\alpha \geq 1$, so they are infinite for $d_A/d_B > 1$. Left: von Neumann relative entropy with dashed line given by (3.28). Center: PRRE with $\alpha = 3/2$ with dashed line given by (3.35). Right SRRE with $\alpha = 2$ with dashed line given by (3.58). The data are given for total Hilbert space dimensions of 2^{14} (blue), 2^{15} (green), and 3^{10} (red). The error bars represent the statistical fluctuations in the 10^3 disorder realizations which decay for large Hilbert spaces.

One can then evaluate the trace distance from the integral

$$T(\rho_A||\sigma_A) = \frac{1}{2}\int \mu_{MP}^{(c)} \boxplus \mathcal{D}_{-1}\left[\mu_{MP}^{(c)}\right](x)|x|, \tag{3.110}$$

leading to (3.76).

3.7 Small-N numerics

All of our computations thus far have been in the limit where both d_A and d_B are large. It is important to ask whether these asymptotic results are accurate when d_A and d_B are finite. One motivation is if these predictions can be observed in experiments and Noisy Intermediate-Scale Quantum (NISQ) technology [166]. Of course, the Hilbert space dimensions are exponentially large in the number of qubits, so there is hope that our results are predictive for small-scale experiments. In this section, we numerically compute the various distance measures and compare to the asymptotic formulas. This serves as a further consistency check of our results, which we find to be extraordinarily accurate.

In Fig. 3.3, we plot the von Neumann relative entropy, $D_{3/2}$, and $\tilde{D}_2$. All of these quantities are infinite for $d_A > d_B$ due to the rank deficiencies in the reduced density

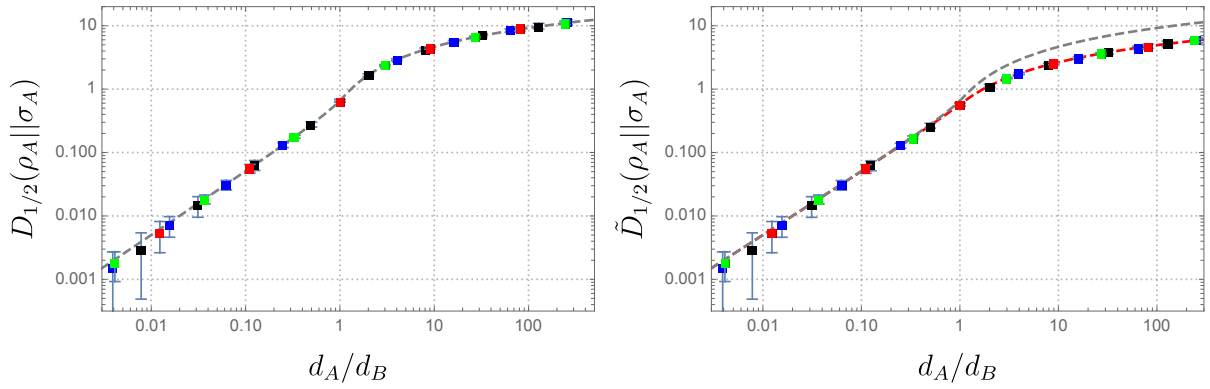


Figure 3.4: Representative examples of relative entropies are shown for $\alpha < 1$ such that there are no divergences. Left: PRRE for $\alpha = 1/2$ (related to Holevo's just-as-good fidelity) with dashed line given by (3.35). Right: SRRE for $\alpha = 1/2$ (related to Uhlmann fidelity) with dashed red lines given by the answer from the spectrum (3.99). The grey line is the upper bound from the PRRE. The data are given for total Hilbert space dimensions 2^9 (black), 2^{10} (blue), 3^6 (red), and 3^7 (green). The error bars represent the statistical fluctuations in the 10^3 disorder realizations which decay for large Hilbert spaces.

matrix. For this reason, we are able to sample very large Hilbert space dimensions because the bottleneck on classical computers is d_A and not the total system size. We find very accurate agreement between the exact large-N predictions and the small-N numerics. The fluctuations in the entropies are noticeably larger for small d_A because of the subleading corrections that we have thus far ignored.

In Fig. 3.4, we investigate the other regime by plotting $D_{1/2}$ and $\tilde{D}_{1/2}$. These quantities are related to Holevo's just-as-good and Uhlmann fidelities respectively and are therefore well-defined in the $d_A > d_B$ regime. This limits the Hilbert space sizes we can probe, though we still find very accurate agreement with the large-N analysis.

Finally, in Fig. 3.5, we plot the trace distance, examining both the small d_A/d_B and large d_A/d_B regimes. The large-N expressions precisely agree with numerics and are bounded within the fidelities.

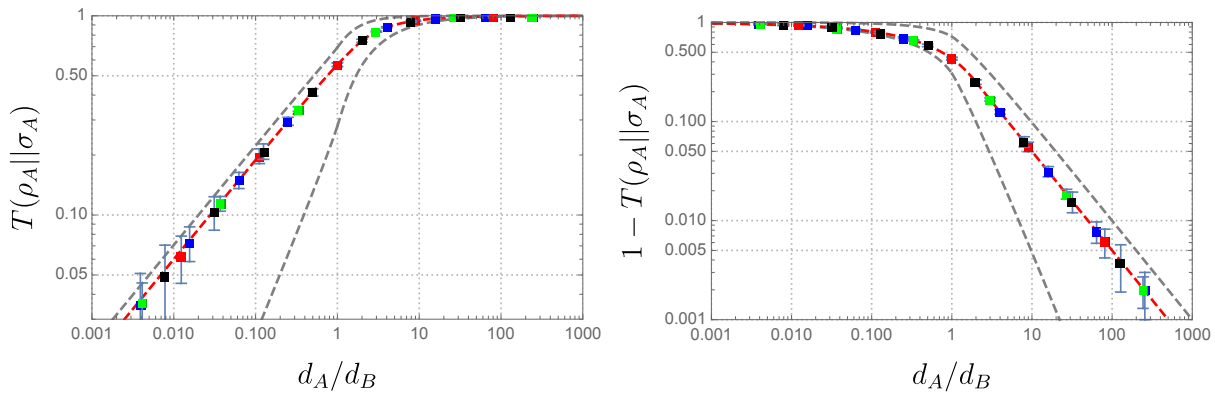


Figure 3.5: Left: The trace distance is shown with dashed red line given by (3.76) and grey lines given by the bounds from the fidelity (3.79) and (3.81). Right: One minus the trace distance is shown to display the approach to one. The data are given for total Hilbert space dimensions 2^9 (black), 2^{10} (blue), 3^6 (red), and 3^7 (green). The error bars represent the statistical fluctuations in the 10^3 disorder realizations which decay for large Hilbert spaces.

3.8 Commutation of ensemble average and logarithm

We have been using the replica trick to compute relative entropies. This has involved evaluating ensemble averages of traces of powers of density matrices and then taking a logarithm. In general, the ensemble average and logarithm do not commute. We now show that the two operations approximately commute in the large-N limit. To properly take the average of a logarithm, we need an additional replica trick

$$\overline{\log\left[\mathrm{Tr}\left[\rho_A^\alpha \sigma_A^m\right]\right]} = \lim_{q\to 0} \frac{\overline{\left(\mathrm{Tr}\left[\rho_A^\alpha \sigma_A^m\right]\right)^q} - 1}{q}. \tag{3.111}$$

For illustration, we work with the PRRE though the argument is the same for all other quantities. In diagrams, the necessary moments are

$\overline{\left(\mathrm{Tr}\left[\rho_A^\alpha \sigma_A^m\right]\right)^q} =$  $\ldots$, (3.112)

where there are q total blocks. As a sum over permutations, this is

$$\overline{\left(\mathrm{Tr}\left[\rho_A^\alpha \sigma_A^m\right]\right)^q} = \frac{1}{(d_A d_B)^{q(\alpha+m)}} \sum_{\tau \in (S_\alpha \times S_m)^{\times q}} d_A^{C((\eta^{-1})^{\times q} \circ \tau)} d_B^{C(\tau)}, \tag{3.113}$$

where in cycle notation

$$(\eta^{-1})^{\times q} = \prod_{i=0}^{q-1} (1+q, 2+q, \ldots, \alpha+m+q). \tag{3.114}$$

The leading terms come from noncrossing contractions within each block independently, leading to $C((\eta^{-1})^{\times q} \circ \tau) + C(\tau) = q(\alpha+m+1)$. Contractions that connect the blocks will be subleading with $C((\eta^{-1})^{\times q} \circ \tau) + C(\tau) \leq q(\alpha+m+1) - 2$. Therefore, the ensemble average factorizes at leading order

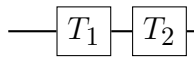
$$\overline{\left(\mathrm{Tr}\left[\rho_A^\alpha \sigma_A^m\right]\right)^q} \simeq \overline{\left(\mathrm{Tr}\left[\rho_A^\alpha \sigma_A^m\right]\right)}^q. \tag{3.115}$$

Using (3.111), this implies that the ensemble average and logarithm approximately commute at large-N.

3.9 Tensor Networks

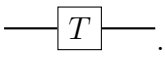
Tensor networks represent a generalization of the states we have considered thus far, adding in the ingredient of locality. As such, tensor networks have been particularly useful as toy models of holographic duality [156, 81]. They are also independently interesting as presenting new classes of ensembles of random states with novel spectral properties [37]. In this section, we generalize our computations to generic random tensor networks, finding qualitatively new phenomena. A specific application of these results is for the random tensor networks used for modeling holography. We clarify which random states faithfully represent holographic states and which do not.

We begin with the warm-up example of a random tensor network with two tensors, T_1 and T_2, contracted together



(3.116)

This is the simplest generalization of the single-tensor network i.e. Haar random state

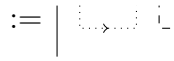


(3.117)

The two-tensor network has one additional degree of freedom, the dimension of the internal bond, d_b. For T_1 and T_2 independently Gaussian, it is straightforward to generalize the diagrammatic approach. The state is now

$$|\Psi\rangle :=$$, (3.118)

where the dotted line is for d_b and is always contracted. The arrow indicates that the dotted lines must be connected in a way that has all arrows with the same orientation. The reduced density matrix is

$$\rho_A :=$$  . (3.119)

Note the directions of the arrows. We can see that the normalization associated with each density matrix is $(d_A d_B d_b)^{-1}$. When taking the average of the moments, we now have a double sum over the permutation group, corresponding to the two random tensors. For example, the purity moments will be

$$\overline{\mathrm{Tr}\left[\rho_A^\alpha\right]} = \frac{1}{(d_A d_B d_b)^\alpha} \sum_{\tau_1,\tau_2 \in S_\alpha} d_A^{C(\eta^{-1}\circ\tau_1)} d_B^{C(\tau_2)} d_b^{C(\tau_1^{-1}\circ\tau_2)}. \tag{3.120}$$

To solve this equation at leading order, we need to maximize the exponents. That is, we must find the set of permutations, $\{\tau_1, \tau_2\}$ that maximize $C(\eta^{-1} \circ \tau_1) + C(\tau_2) + C(\tau_1^{-1} \circ \tau_2)$. This is already a significantly harder problem than the single-tensor network where the answer is that τ must be a non-crossing permutation.

Interestingly, this maximization may be rephrased as a classical network flow problem [37]. We attach a "source" to B and a "sink" to A and determine the maximal flow, $w_{\text{max-flow}}$, of the network where each edge has a weight corresponding to the logarithm of the Hilbert space dimension. We apply the Ford-Fulkerson method in which, one at a time, we take a path from the source to the sink through the tensors, subtracting the weight of the edges by one as we go along the path [65]. Each one of these paths is called an *augmenting path*. We repeat this process until there are no more paths from the source to the sink such that we are left with a *residual network*. The rules for each permutation are that

1. All τ_i's are non-crossing.

2. τ_i's are non-decreasing along each augmenting path in the network i.e. each permutation is contained within all permutations further along the path.

3. All τ_i's in the connected component of the source in the residual network are set to the identity.

4. All τ_i's in the connected component of the sink in the residual network are set to η.

5. All τ_i's in the same connected component are identical.

At leading order, the moments will then be

$$\overline{\text{Tr}\left[\rho_A^\alpha\right]} = N^{-(\alpha-1)w_{\text{max-flow}}} \sum_{\{\tau_1,\tau_2\}} \tilde{d}_A^{C(\eta^{-1}\circ\tau_1)-\alpha} \tilde{d}_B^{C(\tau_2)-\alpha} \tilde{d}_b^{C(\tau_1^{-1}\circ\tau_2)-\alpha}, \tag{3.121}$$

where $\{\tau_1, \tau_2\}$ is the set of permutations obeying the constraints and the dimensions with tildes are $O(1)$ due to multiples of N, a large parameter, being pulled out. For example, if

$d_A = O(N^2)$, then $\tilde{d}_A := d_A N^{-2}$. In the special case that all $\tilde{d}_i$'s are one, we have

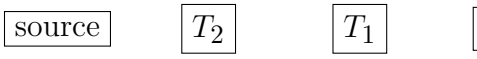
$$\overline{\mathrm{Tr}\left[\rho_A^\alpha\right]} = F_\alpha N^{-(\alpha-1)w_{\text{max-flow}}}, \tag{3.122}$$

where F_α represents the number of paths satisfying the constraints.

For example, consider the case where $d_A = d_B = d_b = N$. There will be a single augmenting path ($w_{\text{max-flow}} = 1$)

(3.123)

such that the resulting network will consist of disconnected tensors with the constraint that $\tau_1 \leq \tau_2 \in NC_\alpha$



. (3.124)

The number of such permutations is given by the second Fuss-Catalan number

$$FC_\alpha^{(2)} := \frac{1}{2\alpha+1}\binom{3\alpha}{\alpha}, \tag{3.125}$$

so the moments will be given by

$$\overline{\mathrm{Tr}\left[\rho_A^\alpha\right]} = \frac{1}{2\alpha+1}\binom{3\alpha}{\alpha} N^{1-\alpha}. \tag{3.126}$$

The associated von Neumann entropy is

$$S_{vN}(A) = \log[N] - \frac{5}{6}. \tag{3.127}$$

This generalizes Page's formula.

Had we instead taken, for example, $\sqrt{d_A} = d_B = d_b = N$, the same augmenting path would have led to the following residual network

(3.128)

Because T_1 is still connected to the sink τ_1 will be set to η while τ_2 can be any non-crossing permutation of which there are a Catalan number's worth, leading to

$$S_{vN}(A) = \log[N] - \frac{1}{2}. \tag{3.129}$$

More generally, we can have a tensor network with n tensors, $\{T_1, T_2, \ldots, T_n\}$. A set of indices of these tensors will be contracted. We refer to the dimensions of these indices by the tensors they connect e.g. d_{ij}. There is also a set of uncontracted indices which correspond to systems A and B. We refer to the dimensions of these indices as d_{Ai} and d_{Bi} which label the subsystem they belong to and the tensors that they are indices of. The purity moments can then be expressed as a sum over n permutation elements

$$\overline{\mathrm{Tr}\left[\rho_A^\alpha\right]} = \frac{1}{(\prod_i d_{Ai} d_{Bi} \prod_j d_{ij})^\alpha} \sum_{\tau_1,\ldots,\tau_n \in S_\alpha} \prod_{i=1}^{n} d_{Ai}^{C(\eta^{-1}\circ\tau_i)} d_{Bi}^{C(\tau_i)} \prod_{j=1}^{n} d_{ij}^{C(\tau_i^{-1}\circ\tau_j)}. \tag{3.130}$$

Here, we must maximize the more complicated exponent which can also be formulated as a network flow problem. (3.121) is generalized to

$$\overline{\mathrm{Tr}\left[\rho_A^\alpha\right]} = N^{-(\alpha-1)w_{\text{max-flow}}} \sum_{\{\tau_i\}} \prod_{i}^{n} \tilde{d}_{Ai}^{C(\eta^{-1}\circ\tau_i)-\alpha} \tilde{d}_{Bi}^{C(\tau_i)-\alpha} \prod_{j=1}^{n} \tilde{d}_{ij}^{C(\tau_i^{-1}\circ\tau_j)-\alpha}, \tag{3.131}$$

where $\{\tau_i\}$ is the set of permutations obeying the updated rules. (3.122) still applies, though the combinatorics may become significantly more difficult. If we are not concerned with the $O(1)$ constant, we only need to determine the maximal flow. By the max-flow min-cut

theorem, the maximal flow from the source to the sink will always be equal to the minimal cut, γ_A, in the network needed to separate the source and sink into disconnected components [65, 56]. The von Neumann entropy is then

$$S_{vN}(A) = \gamma_A \log[N] + O(1). \tag{3.132}$$

We can now generalize this, as before, to relative entropy. We will explicitly compute the PRRE. This only changes the permutation allowed in the sum

$$\overline{\mathrm{Tr}\left[\rho_A^\alpha \sigma_A^m\right]} = \frac{1}{(\prod_i d_{Ai} d_{Bi} \prod_j d_{ij})^{\alpha+m}} \sum_{\tau_1,\ldots,\tau_n \in S_\alpha \times S_m} \prod_{i=1}^n d_{Ai}^{C(\eta^{-1}\circ\tau_i)} d_{Bi}^{C(\tau_i)} \prod_{j=1}^n d_{ij}^{C(\tau_i^{-1}\circ\tau_j)}. \tag{3.133}$$

The key difference between this replica trick and the one for Rényi entropies is that η is not an allowed permutation in the sum. This effects all of the $C(\eta^{-1}\circ\tau_i)$ terms because they are maximized not by $\alpha+m$, but $\alpha+m-1$ which occurs with $\tau_i = \eta_\alpha \times \eta_m \in S_\alpha \times S_m$. This changes rule (1) of the Ford-Fulkerson algorithm to "All τ_i's are in $NC_\alpha \times NC_m$" and rule (4) to "All τ_i's in the connected component of the sink in the residual network are set to $\eta_\alpha \times \eta_m$." The moments are then

$$\begin{aligned}\overline{\mathrm{Tr}\left[\rho_A^\alpha \sigma_A^m\right]} = N^{-(\alpha+m-1)\gamma_A} N^{-(E_A-\gamma_A)} \sum_{\{\tau_i\}} \prod_{i=1}^n \tilde{d}_{Ai}^{C(\eta^{-1}\circ\tau_i)-\alpha-m} \tilde{d}_{Bi}^{C(\tau_i)-\alpha-m} \\ \times \prod_{j=1}^n \tilde{d}_{ij}^{C(\tau_i^{-1}\circ\tau_j)-\alpha-m},\end{aligned} \tag{3.134}$$

where E_A is the weight of the external A edges before applying the Ford-Fulkerson algorithm, which, in the single-tensor case simply equaled the maximum flow. In the special case where

all $\tilde{d}_i$'s are one,

$$\overline{\mathrm{Tr}\left[\rho_A^\alpha \sigma_A^m\right]} = F_{\alpha,m} N^{-(\alpha+m-1)\gamma_A} N^{\gamma_A - E_A}. \tag{3.135}$$

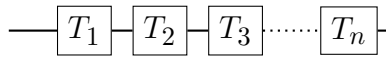
First, consider the two-tensor network when all dimensions equal N. There is a single augmenting path and $E_A = \gamma_A$. However, due to the restriction to $S_\alpha \times S_m$, τ_1 and τ_2 are restricted to non-crossing within the subgroup, such that

$$F_{\alpha,m} = FC_\alpha^{(2)} FC_m^{(2)}, \tag{3.136}$$

which is much smaller than $FC_{\alpha+m}$. The von Neumann relative entropy is then given by

$$D(\rho_A || \sigma_A) = \frac{17}{6}, \tag{3.137}$$

which should be compared with 3/2 which was found for the single-tensor network. Apparently, adding a random tensor makes the state more distinguishable. Generalizing this conclusion, if the tensor network is a string of n tensors



, (3.138)

the combinatorial factor is given by a product of the n^{th} Fuss-Catalan number

$$F_{\alpha,m} = FC_\alpha^{(n)} FC_m^{(n)} = \frac{1}{(n\alpha+1)(nm+1)} \binom{(n+1)\alpha}{\alpha} \binom{(n+1)m}{m}. \tag{3.139}$$

The relative entropy is then

$$D(\rho_A || \sigma_A) = H_{n+1} + n - 1, \tag{3.140}$$

where $H_n := \sum_{k=1}^{n} \frac{1}{k}$ is the harmonic number. This function is monotonically increasing in n.

If we are not concerned with the $O(1)$ contribution, the PRRE will generally be given by

$$D_\alpha(\rho_A||\sigma_A) = \frac{\gamma_A - E_A}{\alpha - 1} \log[N] + O(1). \tag{3.141}$$

This implies that the quantum relative entropy is always divergent if the A edges do not coincide with the minimal cut.[6] We will come back to this point shortly. Similarly, note that Holevo's just-as-good fidelity is exponentially small in this case

$$F_H(\rho_A||\sigma_A) = N^{-2(E_A - \gamma_A)}. \tag{3.142}$$

This means that for many tensor networks, two independent states will be easily distinguishable, no matter the relative size of A and B. When $E_A = \gamma_A$, the $O(1)$ and subleading terms are very interesting.

Recall that holographic random tensor networks are tensor networks composed of random tensors that are arranged geometrically as discretized hyperbolic space (see Fig. 3.6) [81]. Due to the negative curvature of this space, the minimal surfaces for boundary regions always lie in the bulk. This means that we always have $E_A > \gamma_A$, so independent states will always be completely distinguishable. We will see that this is naively in in tension with the holographic results of Chapter 4.1. There, we find that single-tensor networks, which have no built in locality, exactly match certain holographic states while the tensor networks that naively look like Anti-de Sitter space do not share any information theoretic properties with holography except for the entropy.

At face value, the above conclusions are a bit unsettling. Fortunately, this can be reme-

6. There appears to be some similarities of this result with those on holographic Holevo information in Ref. [168].

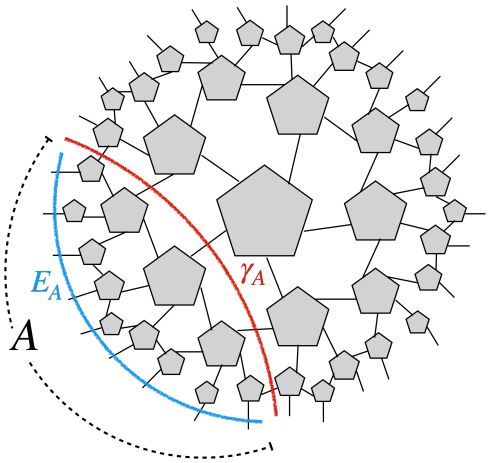


Figure 3.6: A discretization of hyperbolic space is shown as a tensor network. For boundary subregion A, the minimal cut through the network, γ_A, always dips into the bulk and is smaller than the boundary cut E_A

.

died by more carefully stating how a tensor network should model holographic states. Tensor network models represent the holographic map as a quantum error correcting code where the bulk degrees of freedom play the role of "logical qubits" that are protected by being embedded in the larger boundary Hilbert space. The logical qubits live in a code subspace. In the random tensor networks we have been considering, the code subspace (the ensemble of states we are sampling from) is identical to the Hilbert space on the boundary. This equality between bulk effective field theory and boundary Hilbert space dimensions only occurs in AdS/CFT when one has a large black hole whose horizon approaches the asymptotic boundary of the space. This is the reason for the requirement that $E_A = \gamma_A$; all minimal surfaces in the large black hole geometry hug the asymptotic boundary. In order to model other holographic states using tensor networks, we must make the code subspace significantly smaller than the total Hilbert space. Additionally, the bulk density matrices should not be orthogonal. For example, when considering perturbations about vacuum AdS, the total Hilbert space dimension is $O(e^{1/G})$ while the code subspace is $O(e^{G^0})$. In practice, this means that for the two states, ρ and σ, we must take the random tensors to be correlated with each other i.e. the measure for each random tensor only has support on a proper subset

of the Hilbert space. This will be addressed in Section 4.3.

Another important class of random tensor networks is random unitary circuits. In these tensor networks, all tensors are random unitary operators drawn from the Haar measure. Such networks have been the focus of intense study because they present an exactly solvable minimal model of chaotic many-body dynamics, only preserving locality and unitarity. Using the replica trick and Weingarten calculus, many measures of entanglement and operator growth have been computed using geometric quantities in these circuits. [143, 186, 144, 96, 94, 110, 187, 125, 107, 111]. Analogously, the distinguishibility measures discussed in this paper will be computable. For the dynamics of evolving a state from a product state with a random unitary circuit, twice the time (depth of the circuit) plays the role of γ_A when it is smaller than the length of the region A, l_A, which plays the role of E_A. We therefore will find that states are easily distinguishable for times $t < l_A/2$ and very hard to distinguish afterwards. This describes process of thermalization where different initial states become indistinguishable at late times. We note that this is the exact time at which the entanglement entropy saturates after a quench [30, 75]. The details of this calculation, including the precise approach to equilibrium, are left to future work.

3.9.1 Equivalence with Haar unitary tensor networks

Frequently, random tensor networks are constructed by projected Haar unitary states. This is in fact equivalent to the Gaussian random networks we use. The reason is the following. In the Haar random construction, every vertex of degree k is a state of k qudits projected to a random state $U|0\rangle$ where U is a Haar random unitary and $|0\rangle$ is any state. This gives the state $\langle 0|^{\otimes k} U^\dagger |i_1\rangle \cdots |i_k\rangle$. Denoting the set of k indices by one index i, this exactly corresponds to the Gaussian tensor network with the identification $U_{i,0} \leftrightarrow X_i^*$. Every edge corresponds to a maximally entangled pair in the projected Haar random network, which is just the index contraction in the Gaussian network. The projected unitaries indeed have a

Gaussian distribution, since (see e.g. Ref. [34])

$$\begin{aligned}\langle X^*_{i_1}\cdots X^*_{i_n}X_{j_1}\cdots X_{j_n}\rangle &= \langle U_{i_1,0}\cdots U_{i_n,0}U^\dagger_{0,j_1}\cdots U^\dagger_{0,j_n}\rangle = \\ &= \sum_{\sigma,\tau\in S_n}\delta_{i_1,j_{\sigma(1)}}\cdots\delta_{i_n,j_{\sigma(n)}}\mathrm{Wg}(n,\tau\circ\sigma^{-1}) \propto \sum_{\sigma}\delta_{i_1,j_{\sigma(1)}}\cdots\delta_{i_n,j_{\sigma(n)}}\end{aligned} \tag{3.143}$$

and zero for a different number of X and X^*'s, just as for Gaussian variables.

CHAPTER 4

BLACK HOLES

"Where black is the color, where none is the number."

— A Hard Rain's A-Gonna Fall

While studying random states is interesting in its own right, the physical implications of our results becomes significantly richer when we apply them to gravitational systems. We will explain how the connection between random states and gravitational systems is more than an analogy and in some cases, quantitatively identical.

4.1 Fixed-area states in holography

In quantum field theory, we compute the moments of reduced density matrices by evaluating the partition function on certain replica manifolds [90, 29]. These are glued according to the relevant trace structure. If the quantum field theory is holographic, we may map the calculation to an evaluation of the gravitational path integral with boundary conditions prescribed by this trace structure. In the gravitational path integral, we are instructed to sum over all geometries with the given boundary conditions. In the derivation of the Ryu-Takayanagi formula [119], only replica symmetric geometries were considered. In contrast, we find that replica symmetry breaking saddles are important for the evaluation of relative entropies.

In general, it is very difficult to evaluate the gravitational path integral for multiple replicas. This is because the nontrivial coupling between the replicas leads to backreaction, changing the bulk geometry [46]. A great simplification in the gravitational path integral can be made if we focus on "fixed area states" [6, 47]. These are states where the area of one or more surface is fixed and not integrated over. In general, the different replicas will not backreact among themselves, so we are left with copies of the original bulk geometry except

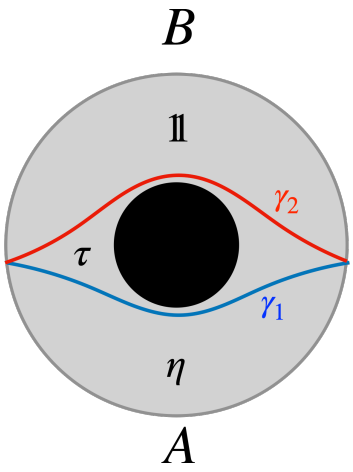


Figure 4.1: An AdS black hole is shown with the asymptotic boundary partitioned into two regions A and B. The two candidate RT surfaces are shown in blue and red respectively. In the gravitational replica trick, the region bounded by γ_1 and A is glued cyclically (η) among the replicas. The region bounded by γ_2 and B is glued according to the identity ($\mathbb{1}$). The region between γ_1 and γ_2 is not fixed by the asymptotic boundary conditions and may be glued among the replicas according to arbitrary S_α permutations (τ).

for potential conical singularities appearing at the locations of the fixed surfaces.

As an example, consider the Rényi entropies of a region on the boundary of a pure state black hole background.[1] There exist two extremal surfaces that are candidate Ryu-Takayanagi surfaces, γ_1 and γ_2, each wrapping the black hole horizon in topologically distinct manners. Denote the areas of these two surfaces A_1 and A_2 respectively. The moments of the reduced density matrix are

$$\mathrm{Tr}\left[\rho_A^\alpha\right] = \frac{\mathcal{Z}(\rho_A^\alpha)}{\mathcal{Z}(\rho_A)^\alpha}, \tag{4.1}$$

where the numerator is the gravitational path integral on the replicated geometry and the denominator is the path integral on a single copy, necessary for normalization. Because the geometry is identical in both geometries away from the conical singularities, the numerator and denominator will almost completely cancel. The nontrivial terms come from the actions

1. On the CFT side of the duality, these states should be thought of as high-energy pure states, though not necessarily energy eigenstates.

of the conical singularities which are determined by their opening angles

$$\mathrm{Tr}\left[\rho_A^\alpha\right] = \sum_{\tau\in S_\alpha} e^{(C(\eta^{-1}\circ\tau)-\alpha)\frac{A_1}{4G}+(C(\tau)-\alpha)\frac{A_2}{4G}} \mathrm{Tr}\left[\rho_b^{n_1}\right]\mathrm{Tr}\left[\rho_b^{n_2}\right]\ldots\mathrm{Tr}\left[\rho_b^{n_{C(\tau)}}\right], \tag{4.2}$$

where G is Newton's constant and the n_k's are the lengths of the cycles in τ and ρ_b is the bulk state labeling the black hole microstate. These account for the bulk entropy term in the FLM formula [64]. The sum over the permutation group arises from all of the ways the replicas may be glued together in the codimension-one region bounded by the two fixed surfaces (see Fig. 4.1). We have chosen the bulk state to be pure such that all of the bulk traces are one

$$\mathrm{Tr}\left[\rho_A^\alpha\right] = \sum_{\tau\in S_\alpha} e^{(C(\eta^{-1}\circ\tau)-\alpha)\frac{A_1}{4G}+(C(\tau)-\alpha)\frac{A_2}{4G}}. \tag{4.3}$$

This sum should now look familiar as it is identical to the sum needed for the Rényi entropies of Haar random states, (3.17), once identifying $d_A \leftrightarrow e^{A_1/4G}$ and $d_B \leftrightarrow e^{A_2/4G}$. In this way, entropies in fixed-area states in holography are identical to entropies in Haar random states.[2]

This connection becomes even richer when we consider more than one gravitational state to compute the relative entropies. Consider the following moments needed for the von Neumann relative entropy

$$\mathrm{Tr}\left[\rho_A\sigma_A^{\alpha-1}\right] = \frac{\mathcal{Z}(\rho_A\sigma_A^{\alpha-1})}{\mathcal{Z}(\rho_A)\mathcal{Z}(\sigma_A)^{\alpha-1}}. \tag{4.4}$$

Both states have fixed areas and the same semiclassical geometry, but come from different black hole microstates, ρ_b and σ_b. In the language of Refs. [8, 74], they are orthogonal states in the same code subspace. Just as before, the gravitational path integral instructs us to

2. This was pointed out in Ref. [157].

sum over all topologies, meaning that the region between the two fixed-area surfaces can be glued according to *any* S_α permutation. This seems different than the calculation in Haar random states which only contained a sum over a subgroup $\mathbb{1} \times S_{\alpha-1}$, (3.23)

$$\mathrm{Tr}\left[\rho_A \sigma_A^{\alpha-1}\right] = \sum_{\tau \in S_\alpha} e^{(C(\eta^{-1}\circ\tau)-\alpha)\frac{A_1}{4G}+(C(\tau)-\alpha)\frac{A_2}{4G}} \mathrm{Tr}\left[\rho_b \sigma_b^{n_1-1}\right] \mathrm{Tr}\left[\sigma_b^{n_2}\right] \ldots \mathrm{Tr}\left[\sigma_b^{n_{C(\tau)}}\right]. \tag{4.5}$$

However, because ρ_b and σ_b are orthogonal, $\mathrm{Tr}\left[\rho_b \sigma_b^{n_1-1}\right]$ is only non-zero if $n_1 = 1$. This reduces the sum to

$$\mathrm{Tr}\left[\rho_A \sigma_A^{\alpha-1}\right] = \sum_{\tau \in \mathbb{1}\times S_{\alpha-1}} e^{(C(\eta^{-1}\circ\tau)-\alpha)\frac{A_1}{4G}+(C(\tau)-\alpha)\frac{A_2}{4G}}, \tag{4.6}$$

which is identical to (3.23) under the same identification. A nearly identical argument holds for the PRRE, SRRE, and trace distance. Therefore we conclude that not only do fixed-area states have the same entropies as Haar random states, but they also have identical Hilbert space geometries.

These results have interesting implications for the distinguishability of black hole microstates. Namely, the asymptotic observer with arbitrarily small information about the state (small A), is able to distinguish between any black hole microstates. This is surprising because we usually consider all black holes to look the same from outside the horizon to any observer, especially local observers. The catch is that the microstates are only distinguishable nonperturbatively in Newton's constant, $O(e^{-1/G})$. This is because all distinguishability measures are linear in d_A/d_B for small region A which translates to proportional to $e^{(A_1-A_2)/4G}$. This means that while distinguishability is in principle possible, the error rates in state discrimination will be very high unless the observer has an exponentially large number of copies of the system. The distinguishability is nonperturbatively small up until region

A is roughly one qubit less than half the boundary system, at which point it becomes $O(1)$. When the observer has access to more than half of the boundary, the black hole microstates become completely distinguishable up to nonperturbatively small corrections.

We also note that these results represent nonperturbative corrections to the JLMS formula which asserts that the boundary relative entropy equals the bulk relative entropy within the entanglement wedge [95]. We have considered bulk states that are pure, orthogonal, and localized between the two extremal surfaces; the bulk states are identical outside of the black hole.[3] When A is sufficiently small, the black hole is not within its entanglement wedge so the bulk states are identical i.e. the bulk relative entropy is zero. Therefore, the JLMS formula asserts that boundary relative entropy is zero. We have shown that there are nonperturbative corrections to this statement. This is examined in greater detail in Section 4.3.

4.2 The PSSY model and replica wormholes

In a landmark achievement, the Page curve [155] for an evaporating black hole was computed for the first time in two independent papers [158, 10]. The key mechanism that "fixed" Hawking's calculation was the inclusion of certain wormhole saddles in the gravitational path integral, referred to as "replica wormholes" [157, 11]. Using the toy model of black hole evaporation presented in Ref. [157] (PSSY), we now show the role of replica wormholes in calculations of relative entropy. This elucidates how the assumptions of Hawking fail. We call this a violation of the no-hair theorem, which is a non-perturbative effect and therefore not present in Hawking's calculation.

The PSSY model consists of two-dimension Jackiw-Teitelboim gravity decorated with

3. Recently, the SRRE between a state with a single fixed-area surface and a state with two fixed-area surfaces was evaluated [82].

end of the world (EOW) branes with k flavors. The Euclidean action is given by

$$I = -\frac{S_0}{2\pi}\left[\frac{1}{2}\int_{\mathcal{M}}\sqrt{g}R + \int_{\partial\mathcal{M}}\sqrt{h}K\right] - \left[\frac{1}{2}\int_{\mathcal{M}}\sqrt{g}\phi(R+2) + \int_{\partial\mathcal{M}}\sqrt{h}\phi K\right] + \mu\int_{brane} ds, \tag{4.7}$$

where S_0 is the large ground state entropy, g (h) is the bulk (asymptotic boundary) metric with curvature R (K), ϕ is the dilaton, and μ is the tension of the EOW brane.

The EOW brane has $k \gg 1$ internal microstates. The global states on the black hole and radiation that we consider are of a maximally entangled form

$$|\Psi\rangle = \frac{1}{\sqrt{k}}\sum_{i=1}^{k} |\psi_i\rangle_B |i\rangle_R , \tag{4.8}$$

where $|i\rangle_R$ represents an orthonormal basis of the states of the radiation. Consider a second microstate

$$|\Psi'\rangle = \frac{1}{\sqrt{k}}\sum_{i=1}^{k} |\psi_i\rangle_B |i+1\rangle_R , \tag{4.9}$$

where $i + k \sim i$ is implied. The definitions of these states are not microscopic in the sense that the $|\psi_i\rangle_B$'s are defined by a gravitational path integral and are not exactly orthogonal. As for the fixed-area state calculation, they may also be thought of as being orthogonal in the code subspace. Because of the non-orthogonality, the reduced states of (4.8) and (4.9) on the radiation are not a priori identical even though the states appear to be related only by a local unitary transformation on B.

The overlap between these two states is

$$\langle\Psi'|\,\Psi\rangle = \frac{1}{k}\sum_{i,l=1}^{k} \langle\psi_l|\,\psi_i\rangle_B \langle l|\,i+1\rangle_R = \frac{1}{k}\sum_{i=1}^{k} \langle\psi_{i+1}|\,\psi_i\rangle_B . \tag{4.10}$$

The overlap on the right hand side is given by a gravity amplitude

$$\langle \psi_i | \, \psi_j \rangle_B = \tag{4.11}$$

Because $i \neq i+1$, connecting the brane in the gravity diagram is incompatible and the amplitude is zero. This means that $|\Psi\rangle$ and $|\Psi'\rangle$ are roughly orthogonal but there are important caveats to this statement because the overlap should be thought of as an ensemble averaged statement. In particular, $|\langle \psi_i | \, \psi_{i+1} \rangle_B|^2$ is non-zero. This is completely analogous to the Haar random story where, on average, two independently chosen vectors will have zero overlap, but the variance is non-zero. The analog of (4.11) for random matrices is

$$\langle \psi_i | \, \psi_{j\neq i} \rangle_B = \tag{4.12}$$

Because, when ensemble averaging, we cannot contract black and red indices, the average, $\overline{\langle \psi_i | \, \psi_{j\neq i} \rangle_B}$, equals zero. In complete analogy, the ensemble average of

$$|\langle \psi_i | \, \psi_{j\neq i} \rangle_B|^2 = \tag{4.13}$$

is non-zero (though very small) because we may now contract red with red and black with black.

We are interested in the relative entropy of the radiation for two different microstates. Hawking and even the island formula papers [158, 10, 13] assumed that the radiation is seen as purely thermal[4] before the Page time in accordance with the no-hair theorem i.e. all black holes of the same mass, charge, and angular momentum look the same from the outside. After the Page time, while the island formula papers did not assume the radiation to be purely thermal, there was no difference between the calculations for different microstates of the

4. Known greybody factors do not qualitatively change this statement in any meaningful way.

black hole. From one perspective, this is great because unitarity can be realized without knowing the microscopic theory. On the other hand, it is disappointing because it bypasses the question of why all initial states appear to lead to the same final state. We resolve this part of the information problem within the PSSY model and believe analogous results should hold in more realistic models of black hole evaporation.

The reduced density matrix on the radiation for the first state is given by

$$\rho_R := \mathrm{Tr}_B \left|\Psi\right\rangle \left\langle\Psi\right| = \frac{1}{k} \sum_{i,j=1}^{k} \left\langle\psi_j\right| \left.\psi_i\right\rangle_B \left|i\right\rangle \left\langle j\right|_R \tag{4.14}$$

and similarly for the second state

$$\rho'_R := \mathrm{Tr}_B \left|\Psi'\right\rangle \left\langle\Psi'\right| = \frac{1}{k} \sum_{i,j=1}^{k} \left\langle\psi_j\right| \left.\psi_i\right\rangle_B \left|i+1\right\rangle \left\langle j+1\right|_R . \tag{4.15}$$

From here on out, we will drop the subscripts labeling the Hilbert spaces as it should be clear.

We now compute the PRRE between two states of the radiation using the replica trick.

$$\begin{aligned}
\mathrm{Tr}\left[\rho_R^{\alpha} \rho_R'^{m}\right] &= \frac{1}{k^{\alpha+m}} \sum_{i_1,\ldots,i_{\alpha+m},j_1,\ldots,j_{\alpha+m}=1}^{k} \left\langle\psi_{j_1}\right| \left.\psi_{i_1}\right\rangle \ldots \left\langle\psi_{j_{\alpha+m}}\right| \left.\psi_{i_{\alpha+m}}\right\rangle \\
&\quad \times \delta_{j_1 i_2} \delta_{j_2 i_3} \ldots \delta_{j_{\alpha-1} i_\alpha} \delta_{j_\alpha i_{\alpha+1}+1} \delta_{j_{\alpha+1}, i_{\alpha+2}} \ldots \delta_{j_{\alpha+m-1}, i_{\alpha+m}} \delta_{j_{\alpha+m}+1, i_1} \\
&= \frac{1}{k^{\alpha+m}} \sum_{i_1,\ldots,i_{\alpha+m}=1}^{k} \left\langle\psi_{i_2}\right| \left.\psi_{i_1}\right\rangle \left\langle\psi_{i_3}\right| \left.\psi_{i_2}\right\rangle \ldots \left\langle\psi_{i_\alpha}\right| \left.\psi_{i_{\alpha-1}}\right\rangle \left\langle\psi_{i_{\alpha+1}+1}\right| \left.\psi_{i_\alpha}\right\rangle \\
&\quad \times \left\langle\psi_{i_{\alpha+2}}\right| \left.\psi_{i_{\alpha+1}}\right\rangle \left\langle\psi_{i_{\alpha+3}}\right| \left.\psi_{i_{\alpha+2}}\right\rangle \ldots \left\langle\psi_{i_{\alpha+m}}\right| \left.\psi_{i_{\alpha+m-1}}\right\rangle \left\langle\psi_{i_1-1}\right| \left.\psi_{i_{\alpha+m}}\right\rangle .
\end{aligned} \tag{4.16}$$

This is a more complicated but still tractable gravitational path integral. As shown in Fig. 4.2, the sum is over only the non-crossing permutations in the $S_\alpha \times S_m$ subgroup of $S_{\alpha+m}$ due to the boundary conditions on the EOW brane i_1 and $i_1 - 1$ being incompatible as well

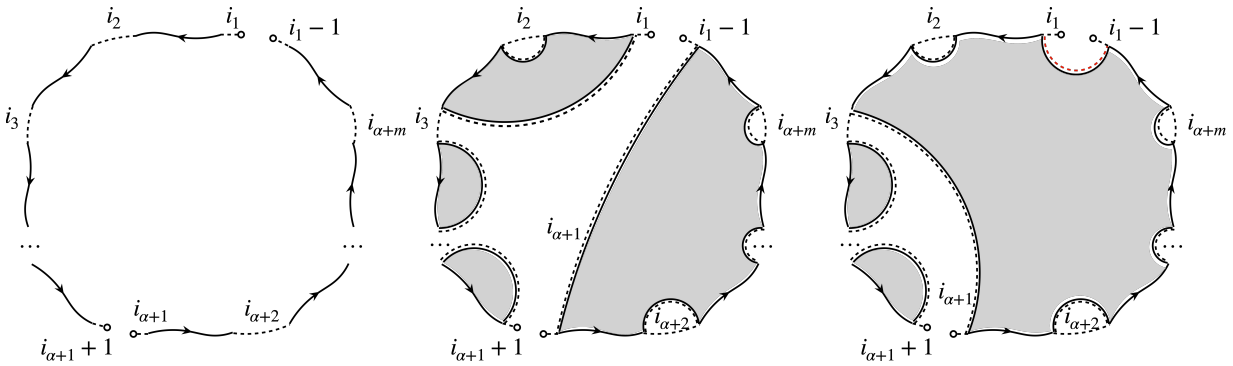


Figure 4.2: Left: The boundary conditions for the path integral in (4.16). Center: An example of a legal way of filling in the geometry. This geometry is also planar, so it contributes at leading order. Right: An example of an illegal way of filling in the geometry because the EOW brane with label i_1 cannot be connected to the EOW brane with label $i_1 - 1$. It is not hard to convince oneself that every diagram that connects the left n and right m boundaries will always lead to an inconsistency as in the right diagram.

as $i_{\alpha+1}$ and $i_{\alpha+1}+1$ being incompatible. There are crossing permutations in $S_{\alpha+m}/S_\alpha \times S_m$ that are compatible with the boundary conditions, but these are subleading. The only other difference in diagrams from the random matrix theory calculation is that the geometries are allowed to have additional handles. This, however, is unimportant because each handle will contribute a factor of e^{-2S_0} and we have assumed the ground state entropy to be large.

To compute the PREE, we must evaluate the gravitational path integral on these replica geometries. For simplicity, we consider the case where the black hole is in the microcanonical ensemble i.e. instead of fixing the lengths of the boundary, we fix the energy, E. The path integral of an n-boundary wormhole is [157]

$$\mathbf{Z}_n = e^{\mathbf{S}} y \left(\sqrt{2E}\right)^n, \tag{4.17}$$

where $\mathbf{S}$ is the microcanonical entropy at energy E. Because of the simply power of n, after normalizing the density matrix, the function y will drop out of the final answer. All calculations are then identical to random matrix theory with the identification of $k \leftrightarrow d_A$

and $e^{\mathbf{S}} \leftrightarrow d_B$.[5] We choose to only write the PRRE

$$D_\alpha(\rho_R||\rho'_R) = \frac{1}{\alpha-1}\begin{cases}\log\left({}_2F_1\left(1-\alpha,-\alpha;2;\frac{k}{e^{\mathbf{S}}}\right)\,{}_2F_1\left(\alpha-1,\alpha;2;\frac{k}{e^{\mathbf{S}}}\right)\right), & k<e^{\mathbf{S}}\\ \log\left(\frac{e^{\mathbf{S}}}{k}\,{}_2F_1\left(1-\alpha,-\alpha;2;\frac{e^{\mathbf{S}}}{k}\right)\,{}_2F_1\left(\alpha-1,\alpha;2;\frac{e^{\mathbf{S}}}{k}\right)\right), & k>e^{\mathbf{S}}\end{cases}. \tag{4.18}$$

The quantum Chernoff bound asserts

$$\begin{aligned}&\lim_{n\to\infty} -\frac{\log\left[\min_A\left[\alpha_n(A)+\beta_n(A)\right]\right]}{n}\\ &\quad=\max_{\alpha\in(0,1)}\begin{cases}-\log\left[{}_2F_1\left(1-\alpha,-\alpha;2;\frac{k}{e^{\mathbf{S}}}\right)\,{}_2F_1\left(\alpha-1,\alpha;2;\frac{k}{e^{\mathbf{S}}}\right)\right], & k<e^{\mathbf{S}}\\ -\log\left[\frac{e^{\mathbf{S}}\,{}_2F_1\left(1-\alpha,-\alpha;2;\frac{e^{\mathbf{S}}}{k}\right)\,{}_2F_1\left(\alpha-1,\alpha;2;\frac{e^{\mathbf{S}}}{k}\right)}{k}\right], & k>e^{\mathbf{S}}\end{cases}\\ &\quad=\begin{cases}-2\log\left[{}_2F_1\left(\frac{1}{2},-\frac{1}{2};2;\frac{k}{e^{\mathbf{S}}}\right)\right], & k<e^{\mathbf{S}}\\ -\log\left[\frac{e^{\mathbf{S}}}{k}\,{}_2F_1\left(\frac{1}{2},-\frac{1}{2};2;\frac{e^{\mathbf{S}}}{k}\right)^2\right], & k>e^{\mathbf{S}}\end{cases},\end{aligned} \tag{4.19}$$

where we found the maximum to be at $\alpha = 1/2$ i.e. Holevo's just-as-good fidelity

$$\min_A\left[\alpha_n(A)+\beta_n(A)\right] \leq F_H(\rho_R||\rho'_R)^{n/2}. \tag{4.20}$$

While (4.20) saturates at large n, the RHS holds as an upper bound for all integer n. Holevo's just-as-good fidelity, plotted in Fig. 4.3, is

$$F_H(\rho_R||\rho'_R) = \begin{cases}{}_2F_1\left(-\frac{1}{2},\frac{1}{2};2;\frac{k}{e^{\mathbf{S}}}\right)^4, & k<e^{\mathbf{S}}\\ \frac{e^{2\mathbf{S}}\,{}_2F_1\left(-\frac{1}{2},\frac{1}{2};2;\frac{e^{\mathbf{S}}}{k}\right)^4}{k^2}, & k>e^{\mathbf{S}}\end{cases}. \tag{4.21}$$

When observing a black hole from the outside, our task is not as simple as distinguishing

5. For analogous reasons, the exact same analysis as in Chapter 3 can be made for the SRRE and trace distance but we do not write these out explicitly to avoid repetition.

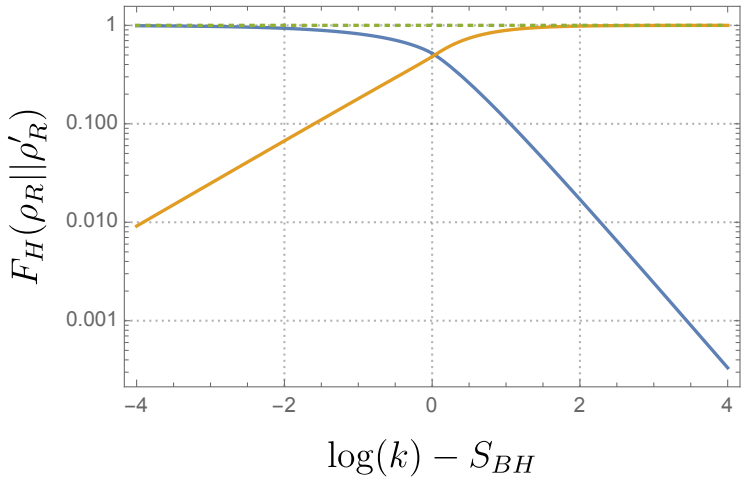


Figure 4.3: Holevo's just-as-good fidelity (blue) and one minus the fidelity (orange) are shown for the PSSY model following (4.21). Before the Page time ($\log[k] = S_{BH}$), the fidelity is exponentially close to one. After the Page, time, the fidelity exponentially decays to zero.

two states. Rather, we need to distinguish between all $e^{\mathbf{S}} \gg 2$ states of the black hole. On the face of it, this seems like an insurmountable task. However, using the multiple quantum Chernoff bound (2.46) and the normal distribution for relative entropies leading to (3.41), we determine that our asymptotic error in the multistate discrimination is identical, at leading order, to that of the two state discrimination.

This has important implications on the nature of black hole evaporation that have not been addressed in the calculations of the entropy. The island formula (or quantum Ryu-Takayanagi formula), stated below, was the main tool in recent calculations of entropy of Hawking radiation [58, 12]

$$S_{vN}(\rho_R) = \min_{\chi} \left[\frac{A_\chi}{4G} + S_{semi\text{-}cl}(\Sigma_\chi) \right], \tag{4.22}$$

where A_χ is the area of the codimension-two quantum extremal surface χ and $S_{semi\text{-}cl}(\Sigma_\chi)$ is the von Neumann entropy of the bulk quantum fields in the codimension-one region, Σ_χ, bounded by χ in the bulk i.e. the entanglement wedge of the radiation, R. While this formula accurately computes the von Neumann entropy of the radiation, restoring *consistency* with

unitarity, it leaves more to be desired. In particular, the bulk entropy term is completely semi-classical and given by a quantum field theory calculation in curved space. The calculation is agnostic to the details of the black hole microstate. One of the central pieces of the apparent paradox was that Hawking radiation always looked the same on the outside regardless of the dynamics in the black hole interior, the phenomenon of *no hair*. (4.21) instead tells us that there is *detectable* hair in the radiation even before the Page time. In fact, information about the particular microstate is present even in the first Hawking quantum. Because the fidelity of the radiation for any two different microstates is strictly less than one, we can always tell the difference between Hawking radiation coming from black holes that are in different microstates, even if they have the same macroscopic parameters mass, charge, and angular momentum.

The caveat is that the difference between states of the radiation coming from distinct black hole microstates is exponentially small in the black hole entropy i.e. the deviation of the fidelity from one before the Page time is $O(e^{-\mathbf{S}})$. This means that while in principle possible, any reasonable observer will be hard-pressed to observe this difference. If we want the probability of error in distinguishing the black hole microstates to be less than ϵ, we need an $O(e^{\mathbf{S}} \log \epsilon)$ number of copies of the state of the radiation, more precisely

$$n = \frac{\log(\epsilon)}{2 \log \left({}_2F_1 \left(-\frac{1}{2}, \frac{1}{2}; 2; \frac{k}{e^{\mathbf{S}}} \right) \right)}. \tag{4.23}$$

After the Page time, there is a different caveat. The fidelity is exponentially close to zero, so the states are essentially fully distinguishable. Precisely, with just one copy of the state, the error probability is bounded above as

$$\min_A \left[\alpha_n(A) + \beta_n(A) \right] \leq \frac{e^{\mathbf{S}} \, {}_2F_1 \left(-\frac{1}{2}, \frac{1}{2}; 2; \frac{e^{\mathbf{S}}}{k} \right)^2}{k} = O(k^{-1}). \tag{4.24}$$

The issue is that the amount of radiation needed to perform this discrimination is of order

the size of the black hole. This means that the observer will have to perform a very complex computation, which again is not so feasible in practice.

Now, consider what we would have concluded if we did not include replica wormholes in the gravitational path integral. This is the analog of Hawking's calculation of the state of the radiation that led to information loss. Removing replica wormholes corresponds to only including the identity permutation in the sum. This means

$$\mathrm{Tr}\left[\rho_R^{\alpha}\rho_R'^{m}\right] = k^{1-\alpha-m}, \tag{4.25}$$

leading to all PRRE's being identically zero, regardless of how much radiation is collected. This is consistent with the initial paradox where the radiation was thought to be in the same state regardless of the black hole microstate. In fact, this was clear from (4.14) and (4.15) because, if the states of the black hole are orthogonal, the reduced density matrices on R would be identical.

Finally, we note that the computation of the relative entropy between two states in the PSSY model was recently studied as a way to detect the violation of global symmetries in theories of quantum gravity [32]. The simpler quantity $\mathrm{Tr}\left[\rho_A\sigma_A\right]$ was evaluated as a proxy with the full relative entropy computation left as an open question. (4.18) is the (generalized) solution to this question. While an $O(1)$ answer was anticipated for the relative entropy after the Page time in Ref. [32], we conclude that the relative entropy is indeed infinite. It is only $O(1)$ slightly prior to the Page time and exponentially small but finite at earlier times.

4.3 Violations of JLMS

We progress to considerations of holography with code subspaces. This brings us to a sharp study of the holographic formula for relative entropy and its corrections and violations. The starting point is the Ryu-Takayanagi (RT) formula with FLM quantum corrections

[173, 172, 93, 64]:

$$S(\rho_R) = \frac{\mathcal{A}(\gamma_R)}{4G} + S_{\text{bulk}}(\rho_r), \tag{4.26}$$

where R is a boundary subregion and γ_R is the minimal-area surface anchored to ∂R. ρ_r is the reduced density matrix on a bulk subregion r such that $\partial r = \gamma_R \cup R$, known as the *entanglement wedge* of R. Equation (4.26), together with the so-called first law of entanglement entropy [25], consequently led to a formula relating the bulk and boundary relative entropy, commonly known as the JLMS formula [95]

$$D(\rho_R||\sigma_R) = D_{\text{bulk}}(\rho_r||\sigma_r). \tag{4.27}$$

While the arguments leading to (4.26) and (4.27) used the Euclidean path integral [64, 119], it was subsequently given a Lorentzian, Hilbert space interpretation by understanding the holographic dictionary as a quantum error-correcting code (QEC) [8, 48, 74]. An exact QEC is an isometric map from the bulk effective field theory (EFT) Hilbert space to the boundary Hilbert space such that it spans a "code" subspace whose information is encoded redundantly on the boundary. It was demonstrated that both (4.26) and (4.27) are straightforward consequences of the existence of such an exact QEC [74]. In fact, the JLMS formula implies the existence of a bulk reconstruction map, the *Petz map*, that can be used to write down bulk operators in terms of their boundary representation [153, 41, 61].

While the above results are quite illuminating, they are by no means the complete story. An emerging theme over the past few years has been that the holographic map is only an approximate QEC. Non-perturbative corrections from the gravitational path integral result in errors in the encoding map, often leading to qualitatively novel features that do not arise for exact QECs [83].

In particular, arguments using the path integral have updated the RT formula with quantum corrections to the quantum extremal surface (QES) formula [58]. Interestingly, the

QES formula can also be given a Lorentzian interpretation by understanding the holographic map as an approximate QEC [5].[6] Once one considers the QES formula, the JLMS formula is updated to a version we refer to as the quantum extremal JLMS (qJLMS) formula [49]

$$D(\rho_R||\sigma_R) = \left\langle \hat{\mathcal{A}}^{\gamma_\sigma} - \log \sigma_r \right\rangle_\rho - \left\langle \hat{\mathcal{A}}^{\gamma_\rho} - \log \rho_r \right\rangle_\rho , \tag{4.28}$$

where γ_σ (γ_ρ) are the codimension-2 surfaces extremizing the first (second) terms in the above equation. For the second term, this is the QES while for the first term, it is the modular extremal surface defined in Ref. [49]. $\hat{\mathcal{A}}$ is the *area operator*, which is a linear operator defined on the gravitational phase space. Further, it is related to the entropy, which is a non-linear observable, as long as one chooses a sufficiently small code subspace [9].

The qJLMS formula arises from a Euclidean path integral analysis under the assumption of replica symmetry. However, in recent times, much has been learned by performing a more careful analysis of the gravitational path integral including replica symmetry breaking contributions [157, 52, 129, 4, 176, 51, 106, 188, 108, 109, 50, 183, 182, 2, 3]. We are therefore motivated to revisit the calculation of the relative entropy using the path integral and analyze corrections to the qJLMS formula. As expected, we find perfect agreement with qJLMS in all cases when evaluating the contributions from only the replica symmetric saddles. However, including all the remaining contributions to the path integral, we find small non-perturbative corrections as well as large violations in certain situations.

In order to do so, we will analyze this problem in the PSSY model. In the PSSY model, we first define a code subspace of states by considering flavor indices on the ETW brane as depicted in Fig. 4.4. We then calculate the boundary relative entropy between various choices of bulk states by employing a replica trick calculation. Doing so, we find the following

6. We will work with stationary spacetimes and thus, we will only be interested in the minimality in the spatial direction. Extremality in the timelike direction has not so far been explained satisfactorily from QEC.

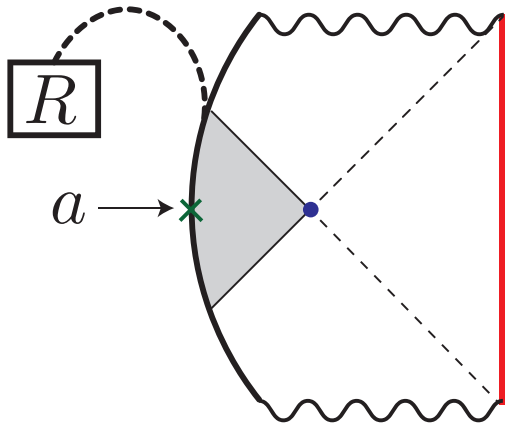


Figure 4.4: The Lorentzian description of the states we consider in the PSSY Model, a JT gravity black hole with an ETW brane. The ETW brane carries k flavour indices that are entangled (dashed lines) with radiation system R. There are additional indices (labelled a) that span the code subspace of bulk states we consider. The extremal surface is denoted in purple and the island that dominates after the Page time is coloured gray.

classes of corrections to the qJLMS formula:

- $O(e^{-1/G})$: Non-perturbative corrections arise from the existence of subdominant saddles. Such configurations exist in the presence of non-minimal QESs.

- $O(G^{-1/2})$: Such corrections, which have appeared previously in the context of entanglement entropy phase transitions [185, 141, 157, 52, 129], also arise for the relative entropy in a similar manner. The key mechanism leading to these corrections is $O(\sqrt{G})$ energy fluctuations in the canonical ensemble.

- $O(G^{-1})$: Leading order corrections to the QES formula are now well understood to arise when considering bulk states that are *incompressible* [4, 188]. The failure of the QES formula, which is the central ingredient entering (4.28), naturally leads to large corrections of $O(G^{-1})$ to the qJLMS formula as well.

- Unbounded: The most novel, surprising violations of the qJLMS formula arise when considering pure states in the bulk. Despite the bulk relative entropy being universally *infinite* for distinct pure states, the boundary quantum system that encodes the

Hawking radiation has full rank, leading to a *finite* boundary relative entropy. The existence of energy fluctuations plays an important role for such violations of the qJLMS formula. In particular, exponentially small tails in the wavefunction that arise from macroscopically large fluctuations result in an avoidance of *rank deficiency* for the boundary quantum system.

The above classes of corrections are rather generic and we expect similar features to arise in more general models of holography from analogous mechanisms.[7]

We would like to emphasize that although the first three mechanisms of violation have been seen previously in other related settings, the infinite violations we find are qualitatively new. They essentially arise from the fact that the relative entropy changes discontinuously under small changes to the state, unlike stable quantities like the von Neumann entropy and reflected entropy [60, 17, 7]. This discontinuous behaviour also extends to the modular Hamiltonian, which is the logarithm of the density matrix. Despite the fact that the bulk and boundary density matrices are related by an approximate isometry in our examples, we find that the operator version of the JLMS formula that relates the bulk and boundary modular Hamiltonians is also violated by a large amount.

Having found these corrections, we discuss the implications of violations of the JLMS formula. An important role that the JLMS formula plays in holography is that of justifying bulk reconstruction and subregion duality. In particular, the JLMS formula being satisfied to a high degree implies the existence of a reconstruction map [98, 41, 31], i.e.,

$$D(\rho_r||\sigma_r) - D(\rho_R||\sigma_R) \geq -2\log\left[F(\rho_r, \mathcal{R}_{\sigma_r,\mathcal{N}} \circ \mathcal{N}[\rho_r])\right], \tag{4.29}$$

where $\mathcal{N}(\rho_r)$ is the completely-positive, trace-preserving map that maps the bulk density

7. In principle, perturbative corrections at all orders in G could lead to other corrections as well. They do not show up in this simple model which is one-loop exact [180], and summing up non-perturbatively suppressed saddles is well justified.

matrix ρ_r to a boundary density matrix ρ_R, and $\mathcal{R}$ is a recovery map, independent of ρ_r, that attempts to recover the information in bulk subregion r on boundary subregion R. If the JLMS formula is satisfied up to small errors, there exists a recovery map that works with good fidelity [41, 31].[8]

However, the degree to which the JLMS formula holds only provides a lower bound on the fidelity of recovery. Thus, not all corrections to the JLMS formula lead to a failure of the bulk reconstruction map. For example, while the $O(G^{-1})$ corrections are expected to lead to a failure of bulk reconstruction since the entanglement wedge is not well defined [4], the infinite violations we find arise in the vanilla situation where the entanglement wedge is sharply defined and one should expects bulk reconstruction to work with good fidelity. We show that this is indeed the case by identifying a high-fidelity reconstruction map. The violations we find essentially arise from the sensitivity of the relative entropy to small changes in the state, and is not always the optimal diagnostic of the fidelity of reconstruction.

We now perform computations of relative entropy in the PSSY model, working through small corrections to large violations of JLMS. First, we write the general form of the two terms in the replica computation of relative entropy when the two states, ρ_R and σ_R, descend from general bulk states, ρ_b and σ_b. Summing over all genus-0 replica wormholes, we have

$$\mathrm{Tr}\left[\rho_R^n\right] = \frac{1}{(kZ_1)^n} \sum_{\tau \in S_n} k^{C(\eta^{-1}\circ\tau)} \prod_{i=1}^{C(\tau)} Z_{n_i} \mathrm{Tr}\left[\rho_b^{n_i}\right], \tag{4.30}$$

where n_i is the length of the i^{th} cycle in permutation τ and Z_p is the JT gravity path integral with p boundaries. We will refer to this term as the *entropy term*. The other term, which

8. These results have also been generalized to the case of Type III_1 von Neumann algebras where the Hilbert space does not factorize into subregions [62, 70, 63].

we refer to as the *relative term*, is given by

$$\mathrm{Tr}\left[\rho_R \sigma_R^{n-1}\right] = \frac{1}{(kZ_1)^n} \sum_{\tau \in S_n} k^{C(\eta^{-1}\circ\tau)} Z_{n_1} \mathrm{Tr}\left[\rho_b \sigma_b^{n_1-1}\right] \prod_{i=2}^{C(\tau)} Z_{n_i} \mathrm{Tr}\left[\sigma_b^{n_i}\right]. \tag{4.31}$$

Consider only including the dominant replica symmetric saddle i.e. τ is either the identity or cyclic permutation. When the identity element is dominant (pre-Page time), the entropy term is identical to the relative term

$$\mathrm{Tr}\left[\rho_R^n\right] = \mathrm{Tr}\left[\rho_R \sigma_R^{n-1}\right] = k^{1-n}, \tag{4.32}$$

so we manifestly find zero relative entropy. In contrast, when the cyclic permutation is dominant,

$$\mathrm{Tr}\left[\rho_R^n\right] = \frac{Z_n}{Z_1^n} \mathrm{Tr}\left[\rho_b^n\right], \quad \mathrm{Tr}\left[\rho_R \sigma_R^{n-1}\right] = \frac{Z_n}{Z_1^n} \mathrm{Tr}\left[\rho_b \sigma_b^{n-1}\right]. \tag{4.33}$$

so the relative entropy is equal to bulk relative entropy

$$D(\rho_R || \sigma_R) = D(\rho_b || \sigma_b). \tag{4.34}$$

Therefore, replica symmetry leads to the JLMS formula. The qJLMS formula is found when noting that the transition in dominance from the identity to cyclic permutations occurs at different locations for the entropy and relative terms.

One natural choice is to take σ_b to be maximally mixed state in the code subspace

$$\sigma_b = \frac{\mathbb{1}_b}{d_{code}}. \tag{4.35}$$

The relative term simplifies to

$$\mathrm{Tr}\left[\rho_R \sigma_R^{n-1}\right] = \frac{1}{(k d_{code} Z_1)^n} \sum_{\tau \in S_n} k^{C(\eta^{-1}\circ\tau)} \prod_{i=1}^{C(\tau)} d_{code} Z_{n_i}. \tag{4.36}$$

This is identical to the entropy term, (4.30), for a pure bulk state ρ_b and renormalized partition function, $Z_p \to d_{code} Z_p$. It is independent of the state ρ_b. Therefore, the relative entropy may be expressed as the difference between two von Neumann entropies

$$D(\rho_R || \sigma_R) = S_{vN}\left(\tilde{\rho}_{pure}\right) - S_{vN}\left(\rho_R\right), \tag{4.37}$$

where $\tilde{\rho}_{pure}$ is the a pure state density matrix with the renormalized partition function. Due to (4.37), observations on corrections to the quantum extremal surface formula for von Neumann entropy naturally carry over to the relative entropy. The other natural choice for σ_b is a pure state that has fidelity t with ρ_b. The relative term becomes

$$\begin{aligned}\mathrm{Tr}\left[\rho_R \sigma_R^{n-1}\right] = &\frac{1}{\left(k e^{\mathbf{S}}\right)^n} \\ &\times \left(\sum_{\tau \in S_n, n_1 = 1} k^{C(\eta^{-1}\circ\tau)} \left(e^{\mathbf{S}}\right)^{C(\tau)} + t \sum_{\tau \in S_n, n_1 \neq 1} k^{C(\eta^{-1}\circ\tau)} \left(e^{\mathbf{S}}\right)^{C(\tau)} \right).\end{aligned} \tag{4.38}$$

This choice leads to qualitatively new violations of the qJLMS.

4.3.1 $O\left(e^{-1/G}\right)$: Wormholes

The smallest corrections to JLMS are non-perturbative, arising from subleading saddles in the gravitational path integral. These are $O\left(e^{-1/G}\right)$ because the gravitational action is inversely proportional to Newton's constant. For this purpose, we will only need to consider

the black hole in the microcanonical ensemble, where both terms simplify

$$\mathrm{Tr}\left[\rho_R^n\right] = \frac{1}{\left(ke^{\mathbf{S}}\right)^n} \sum_{\tau \in S_n} k^{C(\eta^{-1}\circ\tau)} \left(e^{\mathbf{S}}\right)^{C(\tau)} \prod_{i=1}^{C(\tau)} \mathrm{Tr}\left[\rho_b^{n_i}\right], \tag{4.39}$$

$$\mathrm{Tr}\left[\rho_R \sigma_R^{n-1}\right] = \frac{1}{\left(ke^{\mathbf{S}}\right)^n} \sum_{\tau \in S_n} k^{C(\eta^{-1}\circ\tau)} \left(e^{\mathbf{S}}\right)^{C(\tau)} \mathrm{Tr}\left[\rho_b \sigma_b^{n_1-1}\right] \prod_{i=2}^{C(\tau)} \mathrm{Tr}\left[\sigma_b^{n_i}\right]. \tag{4.40}$$

For additional simplicity, we may consider ρ_b to be a pure state in the code subspace such that all of its moments are equal to one, so

$$\mathrm{Tr}\left[\rho_R^n\right] = \frac{1}{\left(ke^{\mathbf{S}}\right)^n} \sum_{\tau \in S_n} k^{C(\eta^{-1}\circ\tau)} \left(e^{\mathbf{S}}\right)^{C(\tau)}. \tag{4.41}$$

For $k, e^{\mathbf{S}} \propto N \to \infty$, only permutations that maximize the exponents remain, so we find that the familiar sum can be re-expressed in terms of hypergeometric functions

$$\mathrm{Tr}\left[\rho_A^n\right] = \begin{cases} k^{1-n}\, {}_2F_1\left(1-n, -n; 2; \frac{k}{e^{\mathbf{S}}}\right), & k < e^{\mathbf{S}} \\ (e^{\mathbf{S}})^{1-n}\, {}_2F_1\left(1-n, -n; 2; \frac{e^{\mathbf{S}}}{k}\right), & k > e^{\mathbf{S}} \end{cases}. \tag{4.42}$$

Taking the analytic continuation, we find

$$S_{vN}(\rho_R) = \lim_{n\to 1} \frac{1}{1-n} \log\left[\mathrm{Tr}\left[\rho_A^n\right]\right] = \begin{cases} \log[k] - \frac{k}{2e^{\mathbf{S}}}, & k < e^{\mathbf{S}} \\ \log\left[e^{\mathbf{S}}\right] - \frac{e^{\mathbf{S}}}{2k}, & e^{\mathbf{S}} < k \end{cases}, \tag{4.43}$$

which is Page's formula [154]. If we choose σ_b to be maximally mixed, we may use (4.37) to find

$$D(\rho_R||\sigma_R) = \begin{cases} \frac{k(d_{code}-1)}{2d_{code}e^{\mathbf{S}}}, & k < e^{\mathbf{S}} \\ \log\left[\frac{k}{e^{\mathbf{S}}}\right] + \frac{e^{2\mathbf{S}}d_{code}-k^2}{2d_{code}e^{\mathbf{S}}k}, & e^{\mathbf{S}} < k < d_{code}e^{\mathbf{S}} \\ \log\left[d_{code}\right] - \frac{e^{\mathbf{S}}(d_{code}-1)}{2k}, & d_{code}e^{\mathbf{S}} < k \end{cases} . \tag{4.44}$$

To compare to the qJLMS formula, we need to identify the QES. Prior to the Page time, defined by $\log[k] = \mathbf{S}$, the QES and consequently the entanglement wedge of the radiation for any state in the code subspace is the empty set. The area terms of (4.28) are trivially zero while the bulk entropy terms cancel due to ρ_r and σ_r being identical, maximally mixed states. After the Page time but before $\log[k] = \log\left[d_{code}\right] + \mathbf{S}$, the QES for σ_R remains the empty set due to the mixedness of σ_b that contributes a bulk entropy term equal to $\log\left[d_{code}\right]$, while the QES for ρ_R becomes nontrivial, located at the black hole horizon. The entanglement wedge for ρ_R thus contains the black hole interior and ρ_b. The area term for ρ_R is then the black hole entropy while the bulk entropy term disappears because ρ_b is pure. Finally, when $\log[k] > \log\left[d_{code}\right] + \mathbf{S}$, the QESs for both ρ_R and σ_R are nontrivial, with areas canceling in (4.28). The bulk term for σ_b is $\log\left[d_{code}\right]$. In total, the qJLMS formula gives

$$D_{qJLMS}(\rho_R||\sigma_R) = \begin{cases} 0, & k < e^{\mathbf{S}} \\ \log\left[\frac{k}{e^{\mathbf{S}}}\right], & e^{\mathbf{S}} < k < d_{code}e^{\mathbf{S}} \\ \log\left[d_{code}\right], & d_{code}e^{\mathbf{S}} < k \end{cases} , \tag{4.45}$$

which is only different from (4.44) by terms that are non-perturbatively small. These corrections become $O(1)$ around the transitions between the three regimes.

We may also take σ_b to be a pure state which, in general, has a fidelity of t with the pure

state ρ_b. We have previously considered the $t = 0$ case where it was noted that the dominant permutations for the first term of (4.38) are those that are the identity on the first element and non-crossing on the rest. There are now $N_{n-1,r}$ such non-crossing permutations with $C(\eta^{-1} \circ \tau) = r$, so the first sum may be written once more as a hypergeometric function. The second sum consists of all non-crossing permutations that are left over, so there are $N_{n,r} - N_{n-1,r}$ of these. We have

$$\mathrm{Tr}\left[\rho_R \sigma_R^{n-1}\right] = \begin{cases} k^{1-n}\left((1-t)_2F_1\left(1-n, 2-n; 2; \frac{k}{e^{\mathbf{S}}}\right) + t_2F_1\left(1-n, -n; 2; \frac{k}{e^{\mathbf{S}}}\right)\right), & k < e^{\mathbf{S}} \\ \left(e^{\mathbf{S}}\right)^{1-n}\left((1-t)_2F_1\left(1-n, 2-n; 2; \frac{e^{\mathbf{S}}}{k}\right) + t_2F_1\left(1-n, -n; 2; \frac{e^{\mathbf{S}}}{k}\right)\right), & e^{\mathbf{S}} < k \end{cases} \tag{4.46}$$

Taking the replica limit, we find

$$D(\rho_R||\sigma_R) = \begin{cases} (1-t)\left(1 + \frac{k}{2e^{\mathbf{S}}} + \left(\frac{e^{\mathbf{S}}}{k} - 1\right)\log\left[1 - \frac{k}{e^{\mathbf{S}}}\right]\right), & k < e^{\mathbf{S}} \\ \infty, & e^{\mathbf{S}} < k \end{cases}, \tag{4.47}$$

where in the second line, we have assumed that $t < 1$. If $t = 1$, the states are identical and the relative entropy is always zero. Because ρ_b and σ_b are both pure, the QESs for ρ_R and σ_R are identical in all regimes. Before the Page time, these are the empty set, leading to trivial relative entropy. After the Page time, they are nontrivial and the infinite relative entropy between ρ_b and σ_b lead to infinite relative entropy in the radiation

$$D_{qJLMS}(\rho_R||\sigma_R) = \begin{cases} 0, & k < e^{\mathbf{S}} \\ \infty, & e^{\mathbf{S}} < k \end{cases}. \tag{4.48}$$

Again, the corrections from the exact formula are non-perturbatively small up until the Page time where they become $O(1)$. We note that all results from this section also apply to

AdS/CFT in higher dimensions in fixed-area states.

4.3.2 $O\left(G^{-1/2}\right)$: Energy Fluctuations

New phenomena arise when we allow the energy of the black hole to fluctuate in the canonical ensemble. Indeed, energy fluctuations in condensed matter systems were first understood to give rise to $O(\sqrt{\text{Volume}})$ corrections to volume-law entanglement near entanglement phase transitions [185, 141]. In order to observe analogous corrections to qJLMS, we consider σ_b in the maximally mixed state and a sufficiently large code subspace to find a separation of time scales. We can then use (4.37) such that all we need is to evaluate von Neumann entropies. For the same reasons as (4.45), the qJLMS predicts

$$D_{qJLMS}(\rho_R||\sigma_R) = \begin{cases} 0, & k < e^{S_{BH}} \\ \log\left[\frac{k}{e^{S_{BH}}}\right], & e^{S_{BH}} < k < d_{code}e^{S_{BH}} \\ \log\left[d_{code}\right], & d_{code}e^{S_{BH}} < k \end{cases}. \tag{4.49}$$

In the canonical ensemble, it is not easy to explicitly evaluate the sum over permutations and then analytically continue n to one to find the von Neumann entropy. Instead, one may solve for the entanglement spectrum, $D(\lambda)$, explicitly, then evaluate the von Neumann entropy as

$$S_{vN}(\rho_R) = -\int d\lambda D(\lambda)\lambda \log\left[\lambda\right]. \tag{4.50}$$

The entanglement spectrum can be extracted from the resolvent

$$D(\lambda) = -\frac{1}{\pi}\lim_{\epsilon\to 0}\text{Im}\left[R(\lambda+i\epsilon)\right], \quad R(\lambda) = \text{Tr}\left[\frac{1}{\lambda-\rho_R}\right]. \tag{4.51}$$

An implicit formula for the resolvent for the canonical ensemble was determined in Ref. [157]

using a Schwinger-Dyson (self-consistency) equation

$$\lambda R(\lambda) = k + \int_0^\infty ds \frac{\rho(s)w(s)R(\lambda)}{k - w(s)R(\lambda)}. \tag{4.52}$$

There is no closed-form solution to this equation but an approximation to the spectrum that works to high accuracy was found to be a *shifted cutoff thermal spectrum*

$$D(\lambda) = \int_0^{s_k} ds\rho(s)\delta(\lambda - \lambda_0 - w(s)), \quad k = \int_0^{s_k} ds\rho(s), \quad \lambda_0 = \frac{1}{k}\int_{s_k}^\infty \rho(s)w(s). \tag{4.53}$$

As the name suggests, this is the thermal spectrum of JT gravity shifted by λ_0 with a cutoff after the first k eigenvalues.

Before the Page time, both density matrices in (4.37) are nearly maximally mixed and the relative entropy is consequently $O\left(\frac{k}{e^{S_{BH}}}\right)$, the familiar non-perturbative corrections to qJLMS found in the previous section. More interestingly, in an $O(G^{-1/2})$ window in $\log[k]$ around the Page time, there is a *negative* $O(G^{-1/2})$ correction to the naive Page curve of $\min\left[\log[k], S_{BH}\right]$ [157]. In (4.37), this only occurs for the second term because the first term has yet to reach the Page time such that $D(\rho_R^{(a)}||\sigma_R)$ is $O\left(G^{-1/2}\right)$ and large in the semi-classical limit. This represents a significant breakdown of qJLMS. After the Page time, the entropy of ρ_R becomes exponentially close to S_{BH}, while the entropy of $\tilde{\rho}_{pure}$ remains exponentially close to $\log\left[k\right]$, such that the relative entropy is large.

The qJLMS formula fails once again when $\tilde{\rho}_{pure}$ undergoes its Page transition at $\log[k] = S_{BH} + \log\left[d_{code}\right]$. At this point, the entropy of ρ_R is still exponentially close to S_{BH}, but the entropy of $\tilde{\rho}_{pure}$ receives an $O\left(G^{-1/2}\right)$ correction. After this transition, the entropies are exponentially close to S_{BH} and $S_{BH} + \log\left[d_{code}\right]$ respectively and the corrections to

qJLMS are, once more, non-perturbatively small. All together, we have

$$D(\rho_R||\sigma_R) = \begin{cases} O\left(\frac{k}{e^{S_{BH}}}\right), & k \ll e^{S_{BH}} \\ O\left((\beta G)^{-1/2}\right), & k \simeq e^{S_{BH}} \\ \log\left[\frac{k}{e^{S_{BH}}}\right] + O\left(\frac{e^{S_{BH}}}{k}, \frac{k}{d_{code}e^{S_{BH}}}\right), & e^{S_{BH}} \ll k \ll d_{code}e^{S_{BH}} \\ \log\left[d_{code}\right] - O\left((\beta G)^{-1/2}\right), & k \simeq d_{code}e^{S_{BH}} \\ \log\left[d_{code}\right] + O\left(\frac{d_{code}e^{S_{BH}}}{k}\right), & k \simeq d_{code}e^{S_{BH}} \end{cases} . \quad (4.54)$$

In higher dimensional AdS/CFT, analogous $O(G^{-1/2})$ corrections to the von Neumann entropy have been evaluated near entanglement phase transitions [129, 52]. These have been attributed to fluctuations in the *areas* of the QESs. The fluctuation of area is the analog of energy fluctuation in the canonical ensemble and we expect the $O(G^{-1/2})$ corrections to qJLMS to persist.

4.3.3 $O\left(G^{-1}\right)$: Incompressibility

So far, we have mainly been concerned with pure states for ρ_b. Pure states fall under the umbrella of *perfectly compressible* quantum states, those that may be well-approximated by keeping only $e^{S(\rho_b)}$ of the states in their support. Interestingly, it was argued in Ref. [4] that incompressible bulk states lead to leading order ($O(G^{-1})$) corrections to the QES formula. By (4.37), these states lead to $O(G^{-1})$ violations of JLMS. As an example of an incompressible state, we now take ρ_b to be a probabilistic mixture of a pure state and the maximally mixed state

$$\rho_b = p\left|\psi\right\rangle\left\langle\psi\right| + (1-p)\frac{\mathbb{1}_b}{d_b}. \quad (4.55)$$

The eigenvalues and hence entropy of ρ_b is easily computable and for large code subspaces $S(\rho_b) = (1-p)\log[d_b] + O(d_b^0)$.

The entropy in such a mixture was computed for fixed-area states (equivalently the microcanonical ensemble) [4]

$$S(\rho_R) = \begin{cases} \log[k], & k \ll \frac{e^{\mathbf{S}}}{p} \\ p\mathbf{S} + (1-p)\log[k], & \frac{e^{\mathbf{S}}}{p} \ll k \ll e^{\mathbf{S}+S(\rho_b)}, \\ \mathbf{S} + S(\rho_b), & e^{\mathbf{S}+S(\rho_b)} \ll k \end{cases} \tag{4.56}$$

where we have dropped $O(1)$ and more subleading corrections. At this order we thus obtain

$$D(\rho_R||\sigma_R) = \begin{cases} 0, & k \ll \frac{e^{\mathbf{S}}}{p} \\ p(\log[k] - \mathbf{S}), & \frac{e^{\mathbf{S}}}{p} \ll k \ll e^{\mathbf{S}+S(\rho_b)} \\ \log[k] - \mathbf{S} - S(\rho_b), & e^{\mathbf{S}+S(\rho_b)} \ll k \ll d_{code}e^{\mathbf{S}} \\ \log[d_{code}] - S(\rho_b), & d_{code}e^{\mathbf{S}} \ll k \end{cases}, \tag{4.57}$$

which may bee seen to involve $O(G^{-1})$ corrections to the qJLMS answer

$$D(\rho_R||\sigma_R) = \begin{cases} 0, & k \leq e^{\mathbf{S}+S(\rho_b)} \\ \log[k] - \mathbf{S} - S(\rho_b), & e^{\mathbf{S}+S(\rho_b)} \leq k \leq d_{code}e^{\mathbf{S}}. \\ \log[d_{code}] - S(\rho_b), & d_{code}e^{\mathbf{S}} \leq k \end{cases} \tag{4.58}$$

4.3.4 $O(\infty)$: Rank Deficiency

In the canonical ensemble, we can also take ρ_b and σ_b to be pure states with fidelity t. The description of the QESs is identical to the microcanonical case (4.48) with the microcanonical

black hole entropy replaced by the canonical S_{BH}

$$D_{qJLMS}(\rho_R||\sigma_R) = \begin{cases} 0, & k < e^{S_{BH}} \\ \infty, & e^{S_{BH}} < k \end{cases}. \tag{4.59}$$

Summing all terms, we have

$$\begin{aligned} \mathrm{Tr}\left[\rho_R \sigma_R^{n-1}\right] &= \frac{1}{(kZ_1)^n} \sum_{\tau \in S_n, n_1=1} k^{C(\eta^{-1}\circ\tau)} \prod_{i=1}^{C(\tau)} Z_{n_i} + \frac{t}{(kZ_1)^n} \sum_{\tau \in S_n, n_1 \neq 1} k^{C(\eta^{-1}\circ\tau)} \prod_{i=1}^{C(\tau)} Z_{n_i} \\ &= (1-t)\frac{\mathrm{Tr}\left[\rho_R^{n-1}\right]}{k} + t\mathrm{Tr}\left[\rho_R^n\right]. \end{aligned} \tag{4.60}$$

We need to take care of the replica limit

$$\lim_{n\to 1} \frac{1}{1-n} \log\left[\mathrm{Tr}\left[\rho_R \sigma_R^{n-1}\right]\right] = -\frac{1-t}{k}\partial_n \mathrm{Tr}\left[\rho_R^{n-1}\right]\big|_{n=1} - t\partial_n \mathrm{Tr}\left[\rho_R^n\right]\big|_{n=1}. \tag{4.61}$$

Re-expressing the traces as integrals over the entanglement spectrum of ρ_R, we have

$$\begin{aligned} \lim_{n\to 1} \frac{1}{1-n} \log\left[\mathrm{Tr}\left[\rho_R \sigma_R^{n-1}\right]\right] &= -\frac{1-t}{k}\partial_n \int d\lambda D(\lambda)\lambda^{n-1}\big|_{n=1} - t\partial_n \int d\lambda D(\lambda)\lambda^n\big|_{n=1} \\ &= -\frac{1-t}{k}\int d\lambda D(\lambda)\log[\lambda] - t\int d\lambda D(\lambda)\lambda\log[\lambda]. \end{aligned} \tag{4.62}$$

In total, the relative entropy is thus

$$D\left(\rho_R||\sigma_R\right) = -(1-t)\int d\lambda D(\lambda)\left(\frac{1}{k} - \lambda\right)\log[\lambda]. \tag{4.63}$$

Prior to the Page time, the entanglement spectrum is a sharply peaked semi-circle of eigenvalues centered around $\frac{1}{k}$ due to the dominance of the identity permutation in the sum. Thus, we see that the relative entropy is close to zero, with the non-perturbative corrections to qJLMS arising strictly from the finite width of the spectrum.

While, as before, there will be large correction near the transition,[9] we focus on a more striking feature of (4.63), namely that after the Page transition, the violations to qJLMS are *infinite*. The qJLMS formula after the Page time predicts an infinite relative entropy because the states in the code subspace are pure. From (4.63), we see that that relative entropy is finite as long as there are no eigenvalues at zero. Indeed, this is the case, as carefully examined in Ref. [157]. This is reasonable because JT gravity has an infinite number of eigenstates. Seemingly innocuous small tails in the spectrum at high energies drastically change the relative entropy. Note that in the microcanonical ensemble (or equivalently random tensors), there is only a finite number of states occupiable in JT gravity, leading to a rank deficiency (zero eigenvalues) following the Page time and an infinite relative entropy.

While the upshot is that there is an infinite violation of JLMS, to get a sense of the true magnitude of the relative entropy, we can approximate the entanglement spectrum using the shifted cutoff thermal spectrum. Assuming the validity of this approximation in the computation of relative entropy,

$$D\left(\rho_R||\sigma_R\right) = -\int d\lambda \rho(s) \left(\frac{1}{k} - \lambda_0 - w(s)\right) \log[\lambda_0 + w(s)], \tag{4.64}$$

which we plot in Fig. 4.5 alongside the Page curve. While the relative entropy is finite, unlike (4.54), it is unbounded at large k.

It is instructive to analyze the relative entropy of the black holes rather than the radiation. Indeed, it was this complementary relative entropy that was important in the exact reconstruction analysis of Ref. [48]. The entropy term is identical to that for the radiation because the global state is pure. The relative term does not have this symmetry and instead

9. In fact, we expect these to be $O\left(G^{-1}\right)$ as in the logarithmic negativity [50] because the integral is most sensitive to the smallest eigenvalues.

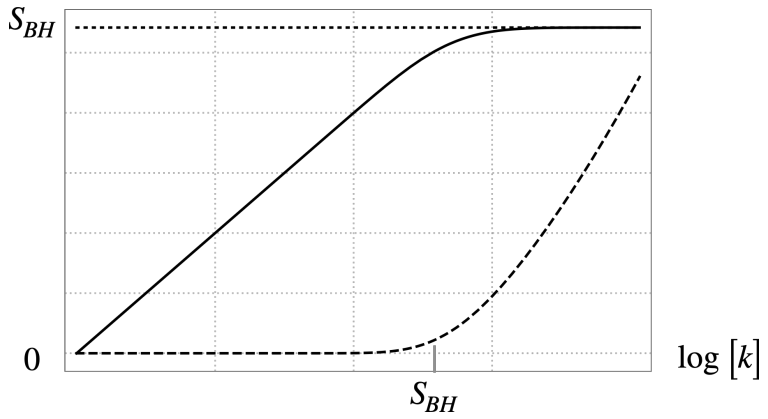


Figure 4.5: The relative entropy between pure states in the code subspace (dashed line) is shown alongside the von Neumann entropy (solid line) and black hole entropy (dotted line) using the approximate entanglement spectrum (4.53). Unlike the microcanonical ensemble, the relative entropy is smooth and finite across the Page transition.

reads

$$\mathrm{Tr}\left[\rho_B \sigma_B^{n-1}\right] = \frac{1}{(kZ_1)^n} \sum_{\tau \in S_n, \tilde{n}_1 = 1} k^{C(\tau)} \prod_{i=1}^{C(\eta^{-1}\circ\tau)} Z_{\tilde{n}_i} + \frac{t}{(kZ_1)^n} \sum_{\tau \in S_n, \tilde{n}_1 \neq 1} k^{C(\tau)} \prod_{i=1}^{C(\eta^{-1}\circ\tau)} Z_{\tilde{n}_i}, \tag{4.65}$$

where $\tilde{n}_i$ is the length of the i^{th} cycle of $\eta^{-1}\circ\tau$. After the Page time, the entanglement wedge of the black hole does not include the interior because the interior is part of the entanglement wedge of the radiation. Therefore the qJLMS formula predicts zero relative entropy for the black hole. In this regime, the identity permutation dominates the sum, leading to

$$\mathrm{Tr}\left[\rho_B \sigma_B^{n-1}\right] \simeq \frac{Z_n}{Z_1^n} \simeq \mathrm{Tr}\left[\rho_B^n\right]. \tag{4.66}$$

Therefore, the true relative entropy is close to zero. The leading corrections to zero relative entropy arise from the $\binom{n}{2}$ replica wormholes that involve a single two-boundary wormhole and $(n-2)$ single-boundary spacetimes. This will be $O\left(\frac{e^{S_{BH}}}{k}\right)$, so we conclude that even when there is an infinite violation of qJLMS for the radiation system, there is only a non-

perturbative correction to qJLMS for the black hole.

4.3.5 General Tensor Networks

We return to general random tensor networks, this time, with additional bulk degrees of freedom representing the code subspace, generalization the results of Section 3.9.

Consider a random tensor network with tensors labeled $i = 1, \ldots, m$. The dimensions of the bonds connecting tensors i and j are denoted d_{ij}. There are also external bonds at the boundary of the network with dimensions $d_{i\partial}$. There may be bulk states in the tensor network, represented by additional bonds on the internal tensors. For the state ρ, we will always consider bulk states that can be formed from random unitaries. These can be pure states if the random unitary entangles the bonds from tensors in the bulk. They can also be mixed states if the unitary entangles the bonds with an auxiliary Hilbert space. For the state σ, the bulk state can be any state created by random unitaries on some bulk bonds and the maximally mixed state on the rest. When the bulk state is created by random unitaries, a random tensor is effectively added to the network.

For this class of tensor network states, we find

$$\mathrm{Tr}\left[\rho_A \sigma_A^{n-1}\right] = \frac{1}{\prod_{\{i,j\}} d_{ij}^n} \sum_{\{\tau_i\}} \prod_{\langle ij \rangle} d_{ij}^{C(\tau_i^{-1} \circ \tau_j)} \prod_{\langle i\partial_A \rangle} d_{i\partial_A}^{C(\eta^{-1} \circ \tau_i)} \prod_{\langle i\partial_B \rangle} d_{i\partial_B}^{C(\tau_i)} \prod_{\langle iMM \rangle} d_{iMM}^{C(\tau_i)}. \tag{4.67}$$

Here, d_{iMM} represent the bond dimensions for the bulk degrees of freedom prepared in the mixed state for σ. There is a permutation τ_i for each tensor in the network. For the bulk tensors, this permutation runs over the entire S_n permutation group, while for the tensors representing independent random unitaries preparing the bulk states of ρ and σ, the sum is only over the $\mathbb{1} \times S_{n-1}$ subgroup.

We can use an adaptation of the Ford-Fulkerson algorithm to solve for the dominant

permutations, and in many cases, the relative entropy at large-N.[10] We apply the following rules for the dominant permutations in the residual network

1. All τ_i's are non-crossing.

2. τ_i's are non-decreasing along each augmenting path in the network i.e. each permutation is contained within all permutations further along the path.

3. τ_i's in the connected component of the source are set to the identity.

4. τ_i's in the connected component of the sink are set to η.

5. All other τ_i's in the same connected component are identical.

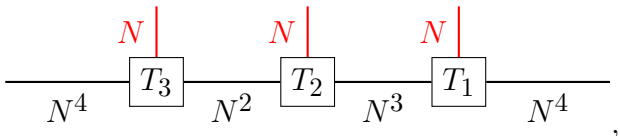
The same strategy can be applied to the entropy term, with minor differences. Firstly, all permutations run over the entire permutation group. The geometry of the network is also slightly modified because we are agnostic to the state σ. The bulk states are generated by random unitaries, which effectively connect an additional tensor to the bonds that this unitary acts on. To have a mixed state in the bulk, the unitary acts on the bulk bond and is connected to the source, effectively adding a boundary edge corresponding to region B with dimension $d_{i\partial B}$. In this new network, we have

$$\mathrm{Tr}\left[\rho_A^n\right] = \frac{1}{\prod_{\{i,j\}} d_{ij}^n} \sum_{\{\tau_i\}} \prod_{\langle ij\rangle} d_{ij}^{C(\tau_i^{-1}\circ\tau_j)} \prod_{\langle i\partial_A\rangle} d_{i\partial_A}^{C(\eta^{-1}\circ\tau_j)} \prod_{\langle i\partial_B\rangle} d_{i\partial_B}^{C(\tau_j)}. \tag{4.68}$$

With the above rules in hand, it is straightforward to compute the relative entropy in nearly any random tensor network. To demonstrate the new tools, we work with a three-tensor network. As discussed in Ref. [4], this can describe a state with two fixed-area surfaces with bulk fields present in each region. For demonstration, we consider the following bond

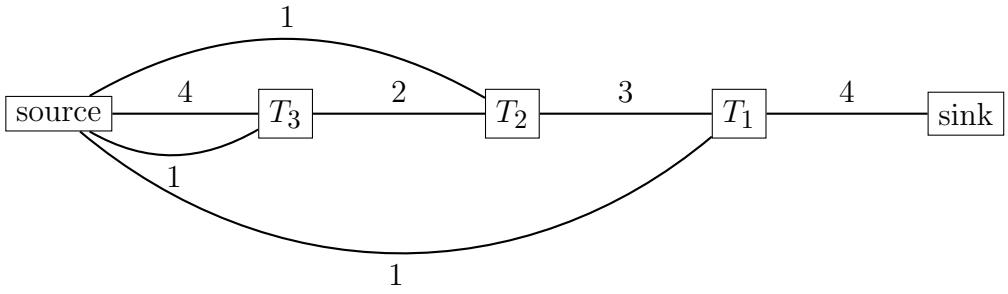
10. We take each dimension to scale with N as $d_{ij} = \tilde{d}_{ij} N^{w_{ij}}$ where $\tilde{d}_{ij}, w_{ij} = O(1)$.

dimensions



, (4.69)

where the red edges are bulk degrees of freedom, the right external edge is region A and the left is region B. Considering the relative entropy between a random pure state in the code subspace, entangling the three bulk regions, and the maximally mixed state in the code subspace, we find the following flow network for the $\mathrm{Tr}\left[\rho_A \sigma_A^{n-1}\right]$ term



. (4.70)

The Ford-Fulkerson algorithm involves taking three augmenting paths: source $\to T_3 \to T_2 \to T_1 \to$ sink, source $\to T_2 \to T_1 \to$ sink, and source $\to T_1 \to$ sink. The total flow equals 4. We are left with the following residual network

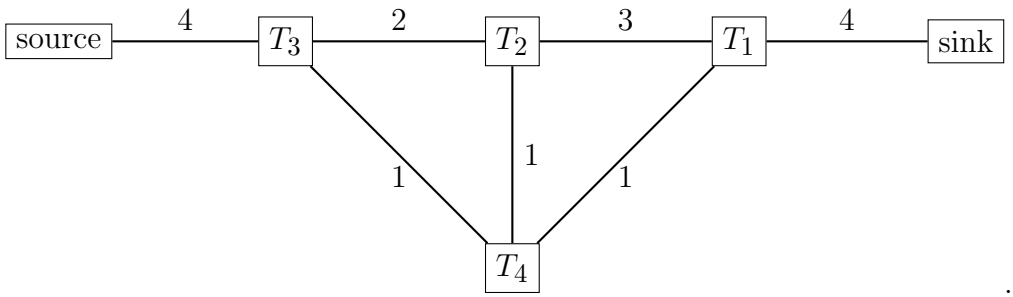
. (4.71)

The rules tell us that all permutations are non-crossing and $\mathbb{1} = \tau_3 \leq \tau_2 \leq \tau_1 \leq \eta$. The number of $\{\tau_i\}$'s satisfying these conditions are known in the combinatorics literature to be given by the second Fuss-Catalan number [15]

$$FC_n^{(2)} := \frac{1}{2n+1}\binom{3n}{n}. \tag{4.72}$$

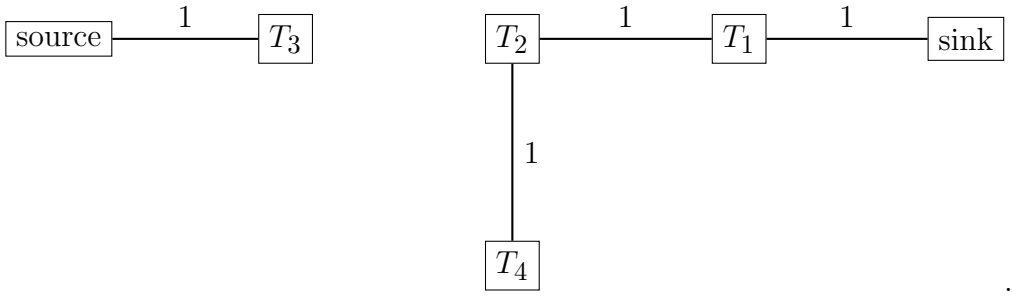
We thus find $\mathrm{Tr}\left[\rho_A \sigma_A^{n-1}\right] = F C_n^{(2)} N^{4(1-n)}$.

Moving on to the entropy term, we have a new flow network



. (4.73)

This time, we need just two augmenting paths[11]: source $\to T_3 \to T_2 \to T_1 \to$ sink and source $\to T_3 \to T_4 \to T_1 \to$ sink. The total flow is 3 and we are left with the following residual network



. (4.74)

The rules now assert that $\tau_3 = \mathbb{1}$ and $\tau_1 = \tau_2 = \tau_4 = \eta$, so $\mathrm{Tr}\left[\rho_A^n\right] = N^{3(1-n)}$. All together, we have

$$D(\rho_A||\sigma_A) = \lim_{n\to 1} \frac{1}{1-n}\left(\log\left[F C^{(2)} N^{4(1-n)}\right] - \log\left[N^{3(1-n)}\right]\right) = \log[N] - \frac{5}{6}. \quad (4.75)$$

11. While the choice of paths are not unique, the final answer following from the rules is.

4.4 Information transfer from black hole to radiation

In this section, we investigate how information is transferred from the black hole to the radiation during the evaporation process using the Hayden-Preskill thought experiment [84]. Suppose we throw a secret diary D into a black hole B at an early stage in the evaporation process. As the black hole evaporates, we collect all of the radiation and refer to it as R, and refer to the remaining black hole as B'. Suppose we know the initial state of B, and have access to a universal quantum computer with which we can act on the radiation R. We then ask how much of the radiation we need in order to learn the initial state of the diary, ρ_D.

We can make this question more explicit in the following way. The relevant decompositions of the Hilbert space at the initial and final times respectively are $\mathcal{H} = \mathcal{H}_D \otimes \mathcal{H}_B$ and $\mathcal{H} = \mathcal{H}_{B'} \otimes \mathcal{H}_R$. We can consider the following quantum channel from $\mathcal{H}_D$ to $\mathcal{H}_R$,

$$\mathcal{N}(\rho_D) = \mathrm{Tr}_{B'} \left[U \left(\rho_D \otimes \rho_B \right) U^\dagger \right], \tag{4.76}$$

where U is the time-evolution operator for the black hole and radiation, together with the diary, and $\rho_B = |\psi_0\rangle \langle \psi_0|_B$ is some fixed initial pure state of the black hole. Asking whether we can learn the state of the diary from the radiation is equivalent to asking whether there exists a universal recovery channel $\mathcal{R}$ such that $\mathcal{R} \circ \mathcal{N}(\rho_D) = \rho_D$ for all ρ_D.

Note that while we will mostly discuss the above question in terms of evaporating black holes in this section, it can also be seen as a more general question about thermalized states in quantum many-body systems. Suppose we put some information in a small subsystem D, and then let D evolve together with the rest of the system, B. Then given some subsystem R in the thermalized state, can we learn the initial state of D?

We will take three different approaches to this question, each of which reveals different aspects of the transfer of information from the black hole to the radiation. We will use the equilibrium approximation, a technique reviewed in Appendix A, in each of these approaches,

which allows us to understand the transfer of information both at infinite and finite temperature. Note that all statements at finite temperature below are based on the canonical ensemble universality class, where we take $\mathcal{I}_\alpha = e^{\beta_1 H_{B'}} \otimes e^{\beta_2 H_R}$. We will assume that the Hilbert space of the diary has a finite dimension d_D. We will in general consider the case where the diary can be large, so that d_D can be $O(e^{1/G})$.

1. We first introduce a reference system Q with the same Hilbert space dimension as D, and consider an initial state in which D is maximally entangled with Q. The time-evolution on $\mathcal{H}_Q$ is trivial, while $\mathcal{H}_B \otimes \mathcal{H}_D$ is evolved with U. We then look at the mutual information $I(Q, R)$ under time-evolution, which can be seen as a way of quantifying the extent to which the radiation contains information about the diary. The behavior of this quantity motivates us to define two natural time-scales:

   - t_{p_1}, the time at which $S_R^{(\mathrm{eq})} = S_{B'}^{(\mathrm{eq})} - \log d_D$. $I(Q, R)$ starts increasing from its initial value of zero at this time.
   - t_{p_2}, at which $S_R^{(\mathrm{eq})} = S_{B'}^{(\mathrm{eq})} + \log d_D$. $I(Q, R)$ reaches its maximal value of $2 \log d_D$ at this time.

   The standard Page time t_p, at which we have $S_R^{(\mathrm{eq})} = S_{B'}^{(\mathrm{eq})}$, lies between these two time scales.

   When $I(Q, R)$ reaches its maximal value at t_{p_2}, this can be interpreted by saying that all the information that was initially in the diary is now present in the radiation. This statement can be understood operationally in terms of the quantum channel $\mathcal{N}$ in (4.76). From the results of [147], $I(Q, R) = 2 \log d_D$ implies the existence of a universal recovery channel $\mathcal{R}$ for $\mathcal{N}$, i.e. $\mathcal{R} \circ \mathcal{N}(\rho_D) = \rho_D$ for all ρ_D. This is consistent with the fact that from a gravitational perspective, t_{p_2} is the latest time at which an island can form [158] (see Fig. 4.6).

2. For some choice of reference state σ_D, and any state ρ_D, we compute the difference in

the relative entropies before and after applying the channel $\mathcal{N}$,

$$D(\mathcal{N}(\rho_D)||\mathcal{N}(\sigma_D)) - D(\rho_D||\sigma_D), \tag{4.77}$$

in order to find a lower bound for the fidelity of some recovery channel $\mathcal{R}$ using (2.52). The relevant time scales for this quantity turn out in general to depend on the choice of σ_D and ρ_D, and are given by

- $t_p(\rho_D)$, the time at which $S_R^{(\text{eq})} = S_{B'}^{(\text{eq})} + S(\rho_D)$. The lower bound on the fidelity of some recovery channel $\mathcal{R}$ first starts to increase from an exponentially small value at this time.
- $t_{p_2}(\sigma_D, \rho_D)$, at which $S_R^{(\text{eq})} = S_{B'}^{(\text{eq})} + D(\rho_D||\sigma_D) + S(\rho_D) = S_{B'}^{(\text{eq})} - \text{Tr}[\rho_D \log \sigma_D]$. The lower bound on the fidelity reaches its maximal value of 1 at this time-scale.

Note that if we take ρ_D to be pure, then $t_p(\rho_D) = t_p$, the standard Page time. If we take $\sigma_D = \mathbf{1}/d_D$, then $t_{p_2}(\sigma_D, \rho_D)$ becomes equal to t_{p_2} defined in the previous point independently of ρ_D, implying that universal recovery is possible after t_{p_2}. Recall that in the previous point, universal recovery at t_{p_2} was deduced from the complementary perspective that the mutual information $I(Q, R)$ becomes maximal at this time.

The lower bound on the fidelity provides an operational way of seeing the gradual transfer of information from the black hole to the radiation between times t_p and t_{p_2}, as the fidelity increases from its minimal value to one in this range of times. However, this quantity does not seem to have a regime which reflects the growth of $I(Q, R)$ from time t_{p_1} to t_p which we observed in the previous point.

3. We explicitly calculate the fidelity of the Petz recovery map $\mathcal{P}$ for $\mathcal{N}$, taking the initial state ρ_D to be pure to simplify calculations. This reveals the following new time scale, which the lower bound in terms of the relative entropy from the previous point is not

sensitive to:

- t_b, the time at which $S^{(\mathrm{eq})}_{\frac{1}{2},R} = S^{(\mathrm{eq})}_{2,B'}$. For a large diary, the fidelity of the Petz map starts growing from its initial value of $F(\rho_D||\sigma_D)$ at this time. This is also the time scale at which the logarithmic negativity between two parts of the radiation starts to become extensive in the canonical ensemble [182, 183].

The time at which the fidelity saturates to unity is t_{p_2}, consistent with the conclusions of the two other approaches in the previous points.

Note that at infinite temperature and for small diaries such that $\log d_D \ll S^{(\mathrm{eq})}_{B'}, S^{(\mathrm{eq})}_{R}$, all the different time scales above coincide, i.e. $t_{p_1} = t_b = t_p = t_{p_2}$. More generally, allowing for finite temperature and large diaries, we have $t_{p_1} \leq t_p \leq t_{p_2}$ and $t_b \leq t_p \leq t_{p_2}$.

4.4.1 From the perspective of mutual information

We first discuss the transfer of information from the perspective of the mutual information of the radiation with a reference system Q. This approach was considered at infinite temperature in the original discussion of [84] by using random states, and has been recently generalized to generic chaotic systems at infinite temperature using properties of operator growth [127]. We now provide a generalization to finite temperature using the equilibrium approximation.

We take the initial state of the full system to be

$$|\Psi_0\rangle = \frac{1}{\sqrt{d_D}} \sum_{n=0}^{d_D-1} |n\rangle_Q \otimes |n\rangle_D \otimes |\psi_0\rangle_B \,. \tag{4.78}$$

We will take the time evolution in Q to be trivial, so that the n-th Renyi entropy of the time-evolved state in subsystem A is given by

$$\mathcal{Z}_{n,A} = \langle \eta_A \otimes e_{\bar{A}} \,|\, (\mathbf{1}\otimes\mathbf{1})^n_Q \otimes (U\otimes U^\dagger)^n_{DB} \,|\, \rho_0 \,, e\rangle \,, \quad \rho_0 = |\Psi_0\rangle\langle\Psi_0| \,. \tag{4.79}$$

From the fact that the time-evolution in Q is trivial, we immediately have at all times

$$S_{n,Q} = \log d_D \,. \tag{4.80}$$

With the initial state (4.78), (4.79) for $A = R$ becomes

$$\mathcal{Z}_{n,R} = \frac{1}{d_D^n} \langle e_{B'} \otimes \eta_R | (U \otimes U^\dagger)^n | (\tilde{\rho}_0)_B \otimes \frac{\mathbf{1}_D}{d_D}, e \rangle \tag{4.81}$$

where $\tilde{\rho}_0 = |\psi_0\rangle \langle\psi_0|$. Note that (4.81) can also be seen as the time-evolution of the entanglement entropy of a mixed state with initial entropy $\log d_D$. We can then use (A.32) to find

$$S_{n,R} = \begin{cases} S_{n,R}^{(\mathrm{eq})} & S_{n,R}^{(\mathrm{eq})} < S_{n,B'}^{(\mathrm{eq})} + \log d_D \\ S_{n,B'}^{(\mathrm{eq})} + \log d_D & S_{n,R}^{(\mathrm{eq})} > S_{n,B'}^{(\mathrm{eq})} + \log d_D \end{cases} \tag{4.82}$$

Since the state on the full system is pure,

$$\mathcal{Z}_{n,QR} = \mathcal{Z}_{n,B'} = \langle \eta_{B'} \otimes e_R | (U \otimes U^\dagger)^n | (\tilde{\rho}_0)_B \otimes \frac{\mathbf{1}_D}{d_D}, e \rangle \,. \tag{4.83}$$

Again using (A.32), we have

$$S_{n,QR} = S_{n,B'} = \begin{cases} S_{n,R}^{(\mathrm{eq})} + \log d_D & S_{n,R}^{(\mathrm{eq})} + \log d_D < S_{n,B'}^{(\mathrm{eq})} \\ S_{n,B'}^{(\mathrm{eq})} & S_{n,R}^{(\mathrm{eq})} + \log d_D > S_{n,B'}^{(\mathrm{eq})} \end{cases} . \tag{4.84}$$

The n-th Renyi mutual information between Q and R is then given by

$$I_n(Q,R) = \begin{cases} 0 & S_{n,R}^{(\mathrm{eq})} < S_{n,B'}^{(\mathrm{eq})} - \log d_D \\ \log d_D + S_{n,R}^{(\mathrm{eq})} - S_{n,B'}^{(\mathrm{eq})} & S_{n,B'}^{(\mathrm{eq})} - \log d_D < S_{n,R}^{(\mathrm{eq})} < S_{n,B'}^{(\mathrm{eq})} + \log d_D \\ 2 \log d_D & S_{n,R}^{(\mathrm{eq})} > S_{n,B'}^{(\mathrm{eq})} + \log d_D \end{cases} \tag{4.85}$$

and by analytic continuation,

$$I(Q,R) = \begin{cases} 0 & S_R^{(\mathrm{eq})} < S_{B'}^{(\mathrm{eq})} - \log d_D \\ \log d_D + S_R^{(\mathrm{eq})} - S_{B'}^{(\mathrm{eq})} & S_{B'}^{(\mathrm{eq})} - \log d_D < S_R^{(\mathrm{eq})} < S_{B'}^{(\mathrm{eq})} + \log d_D \\ 2\log d_D & S_R^{(\mathrm{eq})} > S_{B'}^{(\mathrm{eq})} + \log d_D \end{cases} . \tag{4.86}$$

Note that at all times, $I(Q,B') = 2\log d_D - I(Q,R)$. From [147], when $I(Q,R) = 2\log d_D$ and $I(Q,B') = 0$, there exists a recovery channel $\mathcal{R}$ satisfying $\mathcal{R} \circ \mathcal{N}(\rho_D) = \rho_D$ for all ρ_D under the channel $\mathcal{N}$ defined in (4.76). Hence, the universal recovery channel exists after a time scale t_{p_2}, at which $S_R^{(\mathrm{eq})} = S_{B'}^{(\mathrm{eq})} + \log d_D$. This result from the equilibrium approximation is consistent with the expectations from [83] and [158].

Note from (4.86) that while the mutual information between the radiation and the auxiliary system becomes maximal at t_{p_2}, it begins to grow from zero at an earlier time t_{p_1}, at which $S_R^{(\mathrm{eq})} = S_B^{(\mathrm{eq})} - \log d_D$.

4.4.2 Recovery channel: bound from relative entropy

For the quantum channel that we are interested in, consider a system $D \cup B = R \cup B'$ in the initial state $\rho_D \otimes \rho_B$, evolve it for a while, and then trace out a portion B' of the full system. If we take ρ_B to be some fixed state, this gives a quantum channel from D to R,

$$\mathcal{N}(\rho_D) = \mathrm{Tr}_{B'}\left[U\left(\rho_D \otimes \rho_B\right)U^\dagger\right] . \tag{4.87}$$

The corresponding $\mathcal{N}^\dagger$ from R to D is then given by

$$\mathcal{N}^\dagger(\phi_R) = \mathrm{Tr}_B\left(\rho_B U^\dagger\left(\phi_R \otimes \mathbb{1}_{B'}\right)U\right) . \tag{4.88}$$

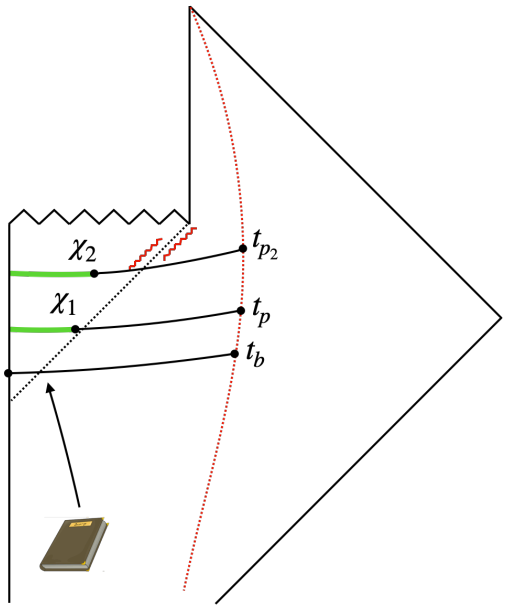


Figure 4.6: A Penrose diagram is shown for an evaporating black hole with a large diary thrown into it. We ask when the diary can be reconstructed from the radiation (the fields to the right of the dotted red line). At early times, the entanglement wedge of the radiation is trivial, so the diary cannot be recovered. If the diary was initialized in a pure state, an island (green) forms at t_p where the nontrivial extremal surface, χ_1, becomes smaller than the entropy of the radiation. If the diary was initialized in the maximally mixed state, the island will not form until t_{p_2} because the generalized entropy includes the entropy of the diary. While the diary cannot be reconstructed with $O(1)$ fidelity until the island forms, the fidelity exponentially increases from its minimal value starting at the earlier time t_b.

Applying (2.51) to this case and imposing a replica trick, we have

$$\mathcal{P}_{\sigma_D,\mathcal{N}} \circ \mathcal{N}(\rho_D) = \lim_{m \to -\frac{1}{2}} \sigma_D^{\frac{1}{2}} \mathrm{Tr}_B \left[\rho_B U^\dagger \left(\mathcal{N}(\sigma_D)^m \mathcal{N}(\rho_D) \mathcal{N}(\sigma_D)^m \otimes \mathbb{1}_{B'} \right) U \right] \sigma_D^{\frac{1}{2}} \ . \quad (4.89)$$

In some of our discussion below, an alternative analytic continuation will also be useful:

$$\mathcal{P}_{\sigma_D,\mathcal{N}} \circ \mathcal{N}(\rho_D) = \lim_{\substack{n_1 \to -\frac{1}{2}, \\ n_2 \to -\frac{1}{2}}} \sigma_D^{\frac{1}{2}} \mathrm{Tr}_B \left[\rho_B U^\dagger \left(\mathcal{N}(\sigma_D)^{n_1} \mathcal{N}(\rho_D) \mathcal{N}(\sigma_D)^{n_2} \otimes \mathbb{1}_{B'} \right) U \right] \sigma_D^{\frac{1}{2}} \ . \quad (4.90)$$

When we take B' to be empty, then $\mathcal{N}$ is unitary, and $\mathcal{N}^\dagger$ is the inverse evolution, for which the above equation gives ρ_D. In the opposite limit, with B' being the full system, we have $\mathcal{P}_{\sigma_D,\mathcal{N}} \circ \mathcal{N}(\rho_D) = \sigma_D$ for any ρ_D.

In the case where ρ_D is a pure state, the fidelity of the Petz map can be written as an overlap

$$\begin{aligned} F(\rho_D, \mathcal{P}_{\sigma_D,\mathcal{N}} \circ \mathcal{N}(\rho_D)) &= \mathrm{Tr}_D \left(\rho_D \mathcal{P}_{\sigma_D,\mathcal{N}} \circ \mathcal{N}(\rho_D) \right) \\ &= \lim_{m \to -\frac{1}{2}} \mathrm{Tr} \left[U \left(\sigma_D^{\frac{1}{2}} \rho_D \sigma_D^{\frac{1}{2}} \otimes \rho_B \right) U^\dagger \mathcal{N}(\sigma_D)^m \mathcal{N}(\rho_D) \mathcal{N}(\sigma_D)^m \right] \ . \end{aligned} \quad (4.91)$$

or a similar overlap using (4.90). (4.89), (4.90), and (4.91) can all be evaluated by using the equilibrium approximation. If the Petz map works perfectly, without error, the fidelity will be one. At worst, the Petz map should output a random answer in which case the fidelity would be exponentially small in the entropy.

We now consider the evolution of the fidelity of recovery channels during the evaporation process. In this subsection, we use (2.52) to put a lower bound on the fidelity of some recovery channel $\mathcal{R}$, by calculating the change of the relative entropy under $\mathcal{N}$ using the equilibrium approximation. In next subsection, we calculate the fidelity of the Petz map

explicitly. Note that in both this subsection and the next, we do not include a reference system Q, and the input state for $\mathcal{N}$ is simply ρ_D.

Infinite temperature

For simplicity, let us first consider the infinite temperature case, and take σ_D to be $\mathbf{1}_D/d_D$ and ρ_D to be a pure state. Then for the first term in the expression (5.50) for $D(\mathcal{N}(\rho_D)||\mathcal{N}(\sigma_D))$, in the infinite temperature case we can sum over all permutations to find (see for instance [126])

$$\mathrm{Tr}\left[\mathcal{N}(\rho_D)^n\right] = \frac{1}{(d_R d_{B'})^n} \sum_{\tau \in \mathcal{S}_n} d_R^{C(\eta^{-1}\circ\tau)} d_{B'}^{C(\tau)} = d_R^{1-n} {}_2F_1\left(1-n, -n, 2, \frac{d_R}{d_{B'}}\right), \tag{4.92}$$

where we consider the regime where d_R and $d_{B'}$ are both large, but with no restrictions on the relative sizes. The resulting von Neumann entropy coincides with Page's formula [154]

$$S(\mathcal{N}(\rho_D)) = \lim_{n\to 1} \frac{1}{1-n} \log \mathrm{Tr}\left[\mathcal{N}(\rho_D)^n\right] = \begin{cases} \log d_R - \frac{d_R}{2d_{B'}}, & d_R < d_{B'} \\ \log d_{B'} - \frac{d_{B'}}{2d_R}, & d_R > d_{B'} \end{cases}. \tag{4.93}$$

Using (5.55), the second term in (5.50) is given by a similar sum

$$\mathrm{Tr}\left[\mathcal{N}(\rho_D)\mathcal{N}(\sigma_D)^{n-1}\right] = \frac{1}{(d_D d_R d_{B'})^n} \sum_{\tau \in S_n} d_R^{C(\eta^{-1}\circ\tau)} \left(d_D d_{B'}\right)^{C(\tau)}. \tag{4.94}$$

This can be interpreted as the entropy of the radiation if it were coupled to an additional bath with a Hilbert space dimension identical to that of the diary. The $n \to 1$ limit is given by

$$\lim_{n\to 1} \frac{1}{1-n} \log \mathrm{Tr}\left[\mathcal{N}(\rho_D)\mathcal{N}(\sigma_D)^{n-1}\right] = \begin{cases} \log d_R - \frac{d_R}{2d_{B'}d_D}, & d_R < d_{B'}d_D \\ \log d_{B'}d_D - \frac{d_{B'}d_D}{2d_R}, & d_R > d_{B'}d_D \end{cases}. \tag{4.95}$$

The relative entropy is therefore

$$D(\mathcal{N}(\rho_D)||\mathcal{N}(\sigma_D)) = \begin{cases} \frac{d_R(d_D-1)}{2d_{B'}d_D}, & d_R < d_{B'} \\ \log\frac{d_R}{d_{B'}} + \frac{d_{B'}}{2d_R} - \frac{d_R}{2d_{B'}d_D}, & d_{B'} < d_R < d_{B'}d_D \\ \log d_D + \frac{d_{B'}(1-d_D)}{2d_R}, & d_R > d_{B'}d_D \end{cases}. \tag{4.96}$$

Plugging back into (2.52), we obtain a lower bound on the fidelity:

$$F(\rho, [\mathcal{R}_{\sigma,\mathcal{N}} \circ \mathcal{N}](\rho)) \geq \begin{cases} \frac{1}{d_D}\exp\left(\frac{d_R(d_D-1)}{2d_{B'}d_D}\right), & d_R < d_{B'} \\ \frac{d_R}{d_{B'}d_D}\exp\left(\frac{d_{B'}}{2d_R} - \frac{d_R}{2d_{B'}d_D}\right), & d_{B'} < d_R < d_{B'}d_D \\ \exp\left(\frac{d_{B'}(1-d_D)}{2d_R}\right), & d_R > d_{B'}d_D \end{cases}. \tag{4.97}$$

The transitions between the different lines of (4.97) occur at the infinite-temperature versions of the times t_p and t_{p_2} respectively. The fidelity is exponentially small prior to t_p. After t_p, the fidelity is still small but exponentially increases until it reaches a value close to one at t_{p_2}. While the formula for $\exp(D(\mathcal{N}(\rho_D)||\mathcal{N}(\sigma_D)) - D(\rho_D||\sigma_D))$ on the RHS of (4.97) is only exact in the limit of large Hilbert space dimensions, we find that it is remarkably accurate even for small system sizes in Fig. 4.7.

Finite temperature

Let us now consider finite temperature, and allow for general choices of the initial state of the diary ρ_D and the reference state σ_D. Now, in general, we can no longer perform the sum over all permutations appearing in the equilibrium approximation. We can instead use the

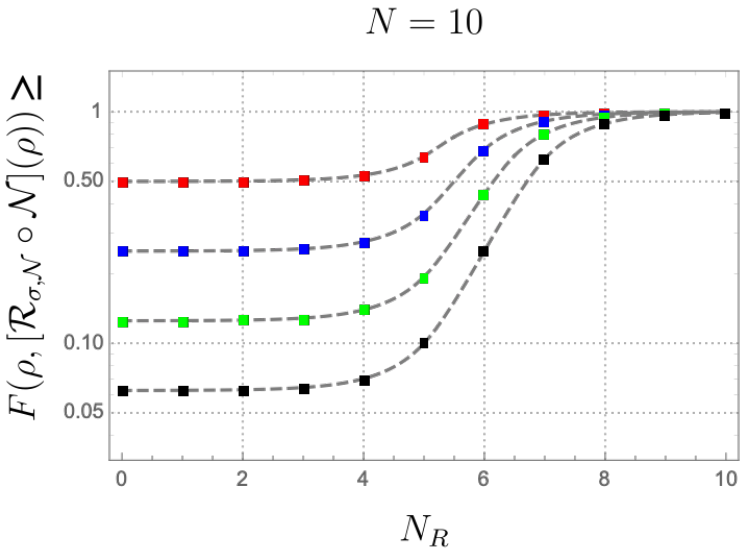


Figure 4.7: The lower bound on the fidelity of the Petz map to recover the diary thrown into the black hole from the radiation as a function of the number of qubits N_R in the radiation, taking the total number of qubits in the system to be $N = 10$. The diary is composed of one (red), two (blue), three (green), or four (black) qubits. The dashed lines are the analytic predictions calculated at large N, (4.97), although the agreement is already very precise. In the numerical data, we take a fiducial pure state tensored with the diary state and apply a random unitary matrix, computing the relative entropy before and after the random channel. We take 10^3 disorder realizations. Error bars are plotted, though barely visible due to small fluctuations.

formula (5.62) based on the dominant permutations in different regimes, to find[12]

$$D(\mathcal{N}(\rho_D)||\mathcal{N}(\sigma_D)) \simeq \begin{cases} 0, & S_R^{(\mathrm{eq})} < S_{B'}^{(\mathrm{eq})} + S(\rho_D) \\ S_R^{(\mathrm{eq})} - S_{B'}^{(\mathrm{eq})} - S(\rho_D), & S_{B'}^{(\mathrm{eq})} + S(\rho_D) < S_R^{(\mathrm{eq})} \\ & < S_{B'}^{(\mathrm{eq})} + D(\rho_D||\sigma_D) + S(\rho_D) \\ D(\rho_D||\sigma_D), & S_R^{(\mathrm{eq})} > S_{B'}^{(\mathrm{eq})} + D(\rho_D||\sigma_D) \end{cases} . \quad (4.98)$$

As an aside, note that we can interpret this result in terms of quantum hypothesis testing, from the standard operational interpretation of the relative entropy. Say Alice were to ask Bob a yes or no question and Bob responded to Alice by writing his answer in his diary (i.e. encoding one of two states ρ_D and σ_D in the diary), then throwing it into the black hole, knowing that Alice would collect the radiation from the black hole in order to decode his message using a quantum measurement. If this is repeated over and over, the error rate of Alice misidentifying Bob's answer is given by $e^{-D(\mathcal{N}(\rho_D)||\mathcal{N}(\sigma_D))}$.

(4.98) implies the following bound on the fidelity of recovery on using (2.52):

$$F(\rho, [\mathcal{R}_{\sigma,\mathcal{N}} \circ \mathcal{N}](\rho) \geq \begin{cases} e^{-D(\rho_D||\sigma_D)}, & S_R^{(\mathrm{eq})} < S_{B'}^{(\mathrm{eq})} + S(\rho_D) \\ e^{S_R^{(\mathrm{eq})} - S_{B'}^{(\mathrm{eq})} - S(\rho_D) - D(\rho_D||\sigma_D)}, & S_{B'}^{(\mathrm{eq})} + S(\rho_D) < S_R^{(\mathrm{eq})} \\ & < S_{B'}^{(\mathrm{eq})} + D(\rho_D||\sigma_D) + S(\rho_D) \\ 1, & S_R^{(\mathrm{eq})} > S_{B'}^{(\mathrm{eq})} + D(\rho_D||\sigma_D) + S(\rho_D) \end{cases} . \quad (4.99)$$

We now see that the lower bound on fidelity first starts increasing at a time $t_p(\rho_D)$, at which $S_R^{(\mathrm{eq})} = S_{B'}^{(\mathrm{eq})} + S(\rho_D)$, and reaches its maximal value at $t_{p_2}(\sigma_D, \rho_D)$, at which $S_R^{(\mathrm{eq})} = S_{B'}^{(\mathrm{eq})} + D(\rho_D||\sigma_D) + S(\rho_D) = S_{B'}^{(\mathrm{eq})} - \mathrm{Tr}[\rho_D \log \sigma_D]$.

12. Note that this formula breaks down and becomes infinite in the large R regime if the initial diary states ρ_D and σ_D are close to orthogonal. This is guaranteed not to be the case if we choose σ_D to be full rank.

4.4.3 Recovery channel: the Petz Map and its fidelity

We progress to explicitly evaluating the fidelity of the Petz recovery map. The quantum channel that we seek to reverse is

$$\begin{aligned} \mathcal{N} : \mathcal{L}(\mathcal{H}_D) &\to \mathcal{L}(\mathcal{H}_R) \\ \rho_D &\mapsto \mathrm{Tr}_{B'} \left[U^\dagger \left(\rho_D \otimes \rho_B \right) U \right], \end{aligned} \tag{4.100}$$

As a reminder, the Petz map is given by

$$\begin{aligned} \mathcal{P}_{\sigma,\mathcal{N}} : \mathcal{L}(\mathcal{H}_R) &\to \mathcal{L}(\mathcal{H}_D) \\ X &\mapsto \sigma^{\frac{1}{2}} \mathcal{N}^\dagger \left(\mathcal{N}(\sigma)^{-\frac{1}{2}} X \mathcal{N}(\sigma)^{-\frac{1}{2}} \right) \sigma^{\frac{1}{2}}. \end{aligned} \tag{4.101}$$

Infinite temperature

We first study Petz recovery at infinite temperature, again taking ρ_D to be pure and σ_D to be maximally mixed. Using the equilibrium approximation for this quantity with $\mathcal{I}_\alpha = \mathbf{1}$, we have $F(\rho, [\mathcal{P}_{\sigma,\mathcal{N}} \circ \mathcal{N}](\rho)) = \lim_{m \to \frac{1}{2}} F_m$, where

$$F_m = \frac{1}{d_D^{2m+3} (d_R d_{B'})^{2m+2}} \sum_{\tau \in \mathcal{S}_{2m+2}} d_R^{C(\eta^{-1} \circ \tau)} d_{B'}^{C(\tau)} d_D^{C(\tau)+\zeta(\tau)} \tag{4.102}$$

where $\zeta(\tau)$ is zero if the first and $(m+2)^{th}$ elements are in different cycles in the permutation τ, and 1 if they are in the same cycle.

At early times, $d_{B'} \gg d_R$ and the identity element will dominate. The identity has $\zeta(e) = 0$, so that

$$F_m = \frac{1}{d_D d_R^{2m+1}} + O\left(d_{B'}^{-1} \right), \quad F(\rho, [\mathcal{P}_{\sigma,\mathcal{N}} \circ \mathcal{N}](\rho)) = \frac{1}{d_D} = F(\rho_D, \sigma_D). \tag{4.103}$$

At late times,when $d_R \gg d_{B'} d_D$ the cyclic permutation will dominate, which has $\zeta(\eta) = 1$, so that

$$F_m = \frac{1}{(d_{B'} d_D)^{2m+1}} + O\left(d_R^{-1}\right), \quad F(\rho, [\mathcal{P}_{\sigma,\mathcal{N}} \circ \mathcal{N}](\rho)) = 1\,. \tag{4.104}$$

To understand how the recovery process improves from (4.103) to (4.104) as the size of the radiation grows, we can perform the full sum in (4.102), as we will soon explain. Before turning to this detailed calculation, we can try to understand relevant time scales by looking at the leading corrections at both early and late times:

1. To find the leading corrections to (4.103), note that the permutations with the next largest value of $C(\tau)$ after the identity are those with a single transposition, e.g. (12). If it were not for the $d_D^{\zeta(\tau)}$ term, we would need to sum over all such permutations. However, only the permutation that transposes the first and $(m+2)^{th}$ elements has $\zeta(\tau) = 1$. Therefore, this permutation gives the leading correction, and on including it we find

$$F_m = \frac{1}{d_D d_R^{2m+1}} + \frac{1}{d_D d_{B'} d_R^{2m}} + O\left(d_D^{-2} d_{B'}^{-1}\right). \tag{4.105}$$

   so that we find the leading and next-to-leading order contributions to the fidelity at early times to be

$$F(\rho, [\mathcal{P}_{\sigma,\mathcal{N}} \circ \mathcal{N}](\rho)) = \frac{1}{d_D} + \frac{d_R}{d_D d_{B'}}, \tag{4.106}$$

   which grows as more radiation is collected, reflecting the improved recovery. This correction starts to give a contribution comparable to the leading term at time t_p, which indicates that the fidelity starts to grow from its initial value at t_p. Note also that this correction has the same scaling as the correction in the lower bound (4.97),

but is twice as large.

When the diary is small, the terms with $\zeta(\tau) = 0$ can also be important. There are $\binom{2m+2}{2} - 1$ leading terms, so that

$$F_m = \frac{1}{d_D d_R^{2m+1}} + \frac{1}{d_D d_{B'} d_R^{2m}} + \frac{m(3+2m)}{d_D^2 d_{B'} d_R^{2m}} + O(d_{B'}^{-2} d_D^{-2}). \tag{4.107}$$

The replica limit $m \to -\frac{1}{2}$ gives

$$F(\rho, [\mathcal{P}_{\sigma,\mathcal{N}} \circ \mathcal{N}](\rho)) = \frac{1}{d_D}\left(1 + \frac{d_R}{d_{B'}} - \frac{d_R}{d_D d_{B'}}\right), \tag{4.108}$$

2. To understand the time scale at which the late-time value (4.104) is reached, let us now evaluate the leading correction to (4.104). There are now $m(m+1)$ permutations satisfying both $C(\eta^{-1} \circ \tau) = 2m+1$ and $\zeta(\tau) = 1$, leading to

$$F_m = \frac{1}{(d_{B'} d_D)^{2m+1}} + \frac{m(m+1)}{d_R (d_{B'} d_D)^{2m}} + O\left(d_R^{-1} d_D^{-2m-1}\right). \tag{4.109}$$

The leading and next-to-leading contributions to the fidelity at late times are therefore

$$F(\rho, [\mathcal{P}_{\sigma,\mathcal{N}} \circ \mathcal{N}](\rho)) = 1 - \frac{d_{B'} d_D}{4 d_R} + O\left(d_R^{-1} d_D^{0}\right). \tag{4.110}$$

Note that the correction had to be negative because the fidelity is bounded above by unity. The correction becomes comparable to the leading term at times earlier than t_{p_2}. Note also that this correction has the same scaling as, but is half the size of, the lower bound correction in (4.97).

For small diaries, there are $(m+1)^2$ other terms which are also important in the

correction to the late-time value, and we find

$$F(\rho, [\mathcal{P}_{\sigma,\mathcal{N}} \circ \mathcal{N}](\rho)) = 1 - \frac{d_{B'} d_D}{4 d_R} + \frac{d_{B'}}{4 d_R} + O\left(d_R^{-2} d_D^{-1}\right). \tag{4.111}$$

Hence, for the infinite temperature case, (4.106) and (4.110) together give the same predictions for the time scales at which the fidelity of $\mathcal{P}$ first increases and the time scale at which it saturates, as the predictions based on the lower bound for some recovery channel in (4.97).

Finally, we can also examine the crossover regime of (4.102), where the fidelity becomes $O(1)$. At the time t_{p_2}, $d_R = d_{B'} d_D := d$, so the sum simplifies to

$$F_m \simeq \frac{1}{d^{4m+4}} \sum_{\tau \in \mathcal{S}_{2m+2}} d^{C(\eta^{-1} \circ \tau) + C(\tau)} d_D^{\zeta(\tau) - 1}. \tag{4.112}$$

The exponent is maximized at $2m+4$ when τ is a non-crossing permutation that has the first and $(m+2)^{th}$ factors in the same cycle. Out of the C_{2m+2} total non-crossing permutations (where C_n is the $n-$th Catalan number), only C_{m+1}^2 have $\zeta(\tau) = 1$, a statement we will soon prove. Therefore,

$$F_m \simeq \frac{C_{m+1}^2}{d^{2m+1}}. \tag{4.113}$$

Taking the $m \to -\frac{1}{2}$ limit, we find the fidelity at t_{p_2} to have an $O(1)$ value independent of the dimensions,

$$F(\rho, [\mathcal{P}_{\sigma,\mathcal{N}} \circ \mathcal{N}](\rho)) = \frac{64}{9\pi^2} \simeq 0.72. \tag{4.114}$$

This fidelity is markedly larger than the lower bound from (4.97), which is $e^{-\frac{1}{2}} \simeq 0.61$. Again, for small diaries, we include the subleading non-crossing permutation with $\zeta(\tau) = 0$

to find

$$F(\rho, [\mathcal{P}_{\sigma,\mathcal{N}} \circ \mathcal{N}](\rho)) = \frac{64}{9\pi^2} + \frac{1 - \frac{64}{9\pi^2}}{d_D}. \tag{4.115}$$

We now turn to the full calculation of the fidelity in the planar limit, i.e., we carry out the full sum in (4.102) analytically. Recall that the planar limit corresponds to non-crossing permutations τ where $C(\eta^{-1} \circ \tau) + C(\tau) = 2m + 3$. We claim that the subset of these permutations with $\zeta(\tau) = 1$ and $C(\eta^{-1} \circ \tau) = p$ is given by a product of Narayana numbers

$$\begin{aligned} NC_{2m+2,\zeta=1,p} &= \sum_{p_1+p_2=p} N_{m+1,p_1} N_{m+1,p_2} = \sum_{p_1=1}^{p} N_{m+1,p_1} N_{m+1,p-p_1} \\ &= \frac{\binom{m+1}{p-2}\binom{m+1}{p-1}\, {}_4F_3(-m-1, -m, 1-p, 2-p; 2, m-p+3, m-p+4; 1)}{m+1}. \end{aligned} \tag{4.116}$$

We note that the previous statement regarding the square of Catalan numbers above (4.113) is the special case where we sum the above equation from $p = 2$ to $p = 2m + 2$.

(4.116) can be proven by considering the elements of the permutation group as a circular lattice of $2m+2$ points then moving to the dual lattice (see Fig. 4.8). In the dual lattice, we count the number of non-crossing permutations that factorize into non-crossing permutations of $m + 1$ and $m + 1$ elements. There is a unique non-crossing permutation of the original lattice corresponding to each one of these permutations, commonly known as the "Kreweras complement." This is the maximally extended permutation in the dual graph that does not cross the permutation. It is clear that for each factorized permutation in the dual lattice, the Kreweras complement in the original lattice has $\zeta(\tau) = 1$. Indeed, the number of non-crossing permutations in the dual graph with $C(\tau) = p$ is the same as the number of non-crossing permutations in the original graph with $C(\eta^{-1} \circ \tau) = p$. Using this insight and the prior knowledge of counting the number of non-crossing permutations with a given

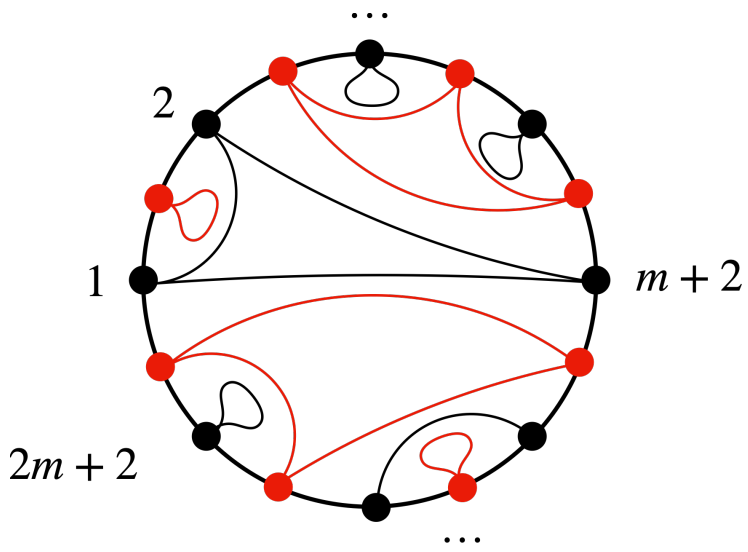


Figure 4.8: A circular lattice of $2m+2$ elements is shown as black dots. The dual lattice is the red dots. We show an example of a non-crossing permutation on the red lattice that factorizes across the first and $m+2$nd elements of the original lattice. Clearly, the Kreweras complement (shown in black lines) connects the first and $m+2$nd elements such that $\zeta(\tau) = 1$.

$C(\tau)$, we arrive at (4.116).

We can now find the sum over all permutations with $\zeta(\tau) = 1$ in (4.102), which we call $\tilde{F}_m$:

$$\begin{aligned}\tilde{F}_m &= \frac{1}{(d_R d_{B'} d_D)^{2m+2}} \sum_{p=2}^{2m+2} NC_{2m+2,\zeta=1,p} d_R^p (d_D d_{B'})^{2m+3-p} \\ &= \frac{1}{d_{B'} d_D} \begin{cases} \left(d_R^{-m} {}_2F_1\left(-m, -m-1, 2, \frac{d_R}{d_D d_{B'}}\right)\right)^2, & d_R < d_D d_{B'} \\ \left((d_D d_{B'})^{-m} {}_2F_1\left(-m, -m-1, 2, \frac{d_D d_{B'}}{d_R}\right)\right)^2, & d_R > d_D d_{B'} \end{cases} . \end{aligned} \tag{4.117}$$

To enumerate the terms with $\zeta(\tau) = 0$ and $C(\eta^{-1} \circ \tau) = p$, we note that this must simply be the remaining non-crossing permutations of which there are $N_{2m+2,p} - NC_{2m+2,\zeta=1,p}$.

We refer to this second contribution to F_m as $\bar{F}_m$, and immediately find

$$\bar{F}_m = \begin{cases} \frac{1}{d_D d_R^{2m+1}} \, {}_2F_1\left(-2m-1, -2m-2, 2, \frac{d_R}{d_D d_{B'}}\right) - \frac{\tilde{F}_m}{d_D}, & d_R < d_D d_{B'} \\ \frac{1}{d_D (d_D d_{B'})^{2m+1}} \, {}_2F_1\left(-2m-1, -2m-2, 2, \frac{d_D d_{B'}}{d_R}\right) - \frac{\tilde{F}_m}{d_D}, & d_R > d_D d_{B'} \end{cases} \tag{4.118}$$

This expression simplifies in the $m \to -\frac{1}{2}$ limit to

$$\bar{F}_{-\frac{1}{2}} = \frac{1}{d_D} - \frac{\tilde{F}_{-\frac{1}{2}}}{d_D} \tag{4.119}$$

Therefore, the fidelity is

$$F(\rho, [\mathcal{P}_{\sigma,\mathcal{N}} \circ \mathcal{N}](\rho)) = \begin{cases} (1 - \frac{1}{d_D}) \frac{d_R}{d_{B'} d_D} \, {}_2F_1\left(\frac{1}{2}, -\frac{1}{2}, 2, \frac{d_R}{d_D d_{B'}}\right)^2 + \frac{1}{d_D} & d_R < d_D d_{B'} \\ (1 - \frac{1}{d_D}) \, {}_2F_1\left(\frac{1}{2}, -\frac{1}{2}, 2, \frac{d_D d_{B'}}{d_R}\right)^2 + \frac{1}{d_D} & d_R > d_D d_{B'} \end{cases} \tag{4.120}$$

This expression is plotted in Fig. 4.9, where it matches perfectly with numerical tests for finite system sizes.

Finite temperature

At finite temperature, we have far less control over the contributions from subleading permutations, and hence we are able to find the value of the fidelity corresponding to the sum over all permutations in the equilibrium approximation only in some special cases. First, note that based on leading permutations, we expect again that at early times when the identity permutation dominates, the fidelity is given by

$$F(\rho, [\mathcal{P}_{\sigma,\mathcal{N}} \circ \mathcal{N}](\rho)) = F(\rho_D, \sigma_D). \tag{4.121}$$

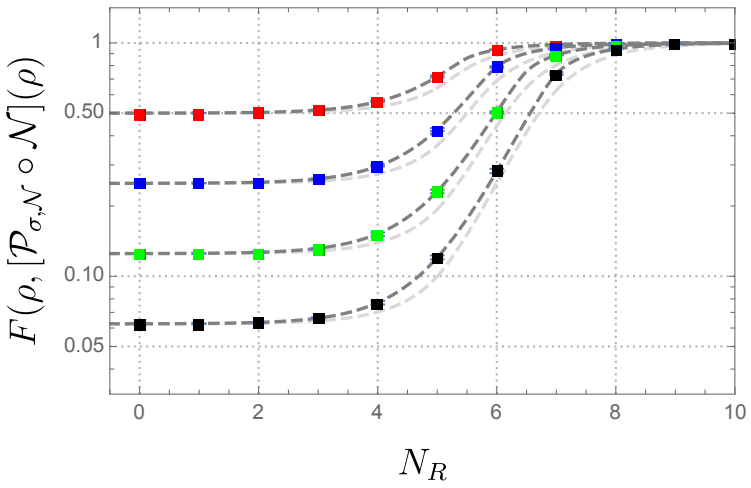


Figure 4.9: The fidelity of the Petz map for the Hayden-Preskill protocol at infinite temperature, as a function of the number of qubits N_R in the radiation. The total number of qubits in the black hole and the radiation is 10 in all cases, and the different curves correspond to different sizes of the diary, ranging from 1 to 4 qubits from top to bottom. The gray dashed curve is the analytic result given by (4.120). The light gray curves are the lower bounds set by the change in relative entropy. In the numerical data, we take a fiducial pure state tensored with the diary state and apply a random unitary matrix and partial trace, followed by the Petz recovery map. We numerically evaluate the between this recovered state and the initial state. Again, we show the average over 10^3 disorder realizations and the corresponding error bars.

where in the last expression we take σ_D to be an equilibrium density matrix on D, and ρ_D to be some pure state. Similarly, at sufficiently late times, the cyclic permutation should dominate, so that we have

$$F(\rho, [\mathcal{P}_{\sigma,\mathcal{N}} \circ \mathcal{N}](\rho)) = 1\,. \tag{4.122}$$

At intermediate times, the fidelity will interpolate between these values. To probe this regime and find the relevant time scales, let us first look at the leading corrections to the dominant permutations. The first correction to early-time contribution from the identity permutation comes from the permutation that swaps the first and $(m+2)^{th}$ copies,

$$F_m = \frac{Z_{2m+2,R}}{Z_{1,R}^{2m+2}} F(\rho_D, \sigma_D) + \frac{Z_{2,B'} Z_{m+1,R}^2}{Z_{1,B'}^2 Z_{1,R}^{2m+2}} \mathrm{Tr}\left[\sigma_D^{\frac{1}{2}} \rho_D \sigma_D^{\frac{1}{2}} \rho_D\right]. \tag{4.123}$$

In the replica limit, this can be rewritten in terms of Renyi entropies as

$$F(\rho, [\mathcal{P}_{\sigma,\mathcal{N}} \circ \mathcal{N}](\rho)) = F(\rho_D, \sigma_D) + e^{S^{(\mathrm{eq})}_{\frac{1}{2},R} - S^{(\mathrm{eq})}_{2,B'}} \mathrm{Tr}\left[\sigma_D^{\frac{1}{2}} \rho_D \sigma_D^{\frac{1}{2}} \rho_D\right]. \tag{4.124}$$

We can assume $\mathrm{Tr}\left[\sigma_D^{\frac{1}{2}} \rho_D \sigma_D^{\frac{1}{2}} \rho_D\right]$ is of roughly the same magnitude as $F(\rho_D, \sigma_D)$ (note that if we take σ_D to be that maximally mixed state and ρ_D to be pure, then both are equal to $\frac{1}{d_D}$). This suggests that the fidelity begins to grow significantly from its initial value at the time scale t_b, at which $S^{(\mathrm{eq})}_{\frac{1}{2},R} = S^{(\mathrm{eq})}_{2,B'}$.

However, in order to reliably calculate the fidelity past t_b, we must sum over the contributions from all permutations corresponding to planar diagrams, a calculation we explain in Section 4.4.4 using a generating functional method. For this calculation, we assume that σ_D is maximally mixed, ρ_D is pure, the radiation is at infinite temperature, and the black hole is at inverse temperature β, with the density of states as in AdS_3, $\rho(E) = e^{cV\sqrt{E/V}}$. Now if we define

$$x \equiv \frac{\log d_R}{V}, \quad y \equiv \frac{\log d_D}{V}, \tag{4.125}$$

where V is the volume of B' and d_R is the dimension of the radiation, then the various relevant time scales are given by:

$$t_b: \quad x = \frac{3c^2}{8\beta}, \quad t_p: \quad x = \frac{c^2}{2\beta}, \quad t_b: \quad x = \frac{c^2}{2\beta} + y. \tag{4.126}$$

We find that for small diaries (i.e. y is $O(V^{-1})$), the fidelity grows rapidly at t_p and the time scale t_b does not turn out to be relevant,

$$F(\rho, [\mathcal{P}_{\sigma,\mathcal{N}} \circ \mathcal{N}](\rho)) \approx \begin{cases} \frac{1}{d_D} & x < \frac{c^2}{2\beta} \\ 1 & x > \frac{c^2}{2\beta} \end{cases} \tag{4.127}$$

with the transition occurring in a region of size $O((V\beta)^{-\frac{1}{2}})$ in x. On the other hand, for a sufficiently large diary such that there is a regime where $\frac{3c^2}{8\beta} < x < \frac{5c^2}{16\beta} + y$, we have

$$F(\rho, [\mathcal{P}_{\sigma,\mathcal{N}} \circ \mathcal{N}](\rho)) = \begin{cases} \frac{1}{d_D} & x < \frac{3c^2}{8\beta} \\ e^{V(-\frac{3c^2}{8\beta}+x-y)} = e^{\log d_R - S_{2,B'} - \log d_D} & \frac{3c^2}{8\beta} < x < \frac{5c^2}{16\beta} + y \\ e^{V(c\sqrt{-\frac{c^2}{4\beta^2}+\frac{x-y}{\beta}}-x+y)} & \frac{5c^2}{16\beta} + y < x < \frac{c^2}{2\beta} + y \\ 1 & x > \frac{c^2}{2\beta} + y \end{cases} \tag{4.128}$$

So for a sufficiently large diary, the fidelity starts increasing exponentially from its initial value of $\frac{1}{d_D}$ at time t_b, and the initial increase is precisely as predicted by (4.124). Note that the lower bound on the fidelity from the change in relative entropy from (4.99) is not sensitive to the time scale t_b, and only starts growing at the Page time t_p.

When y is $O(1)$ but the diary is not sufficiently large such that such that there is a regime where $\frac{3c^2}{8\beta} < x < \frac{5c^2}{16\beta} + y$, the fidelity starts increasing from $\frac{1}{d_D}$ at a time t_r between t_b and

t_p defined by

$$t_r : \quad x = \frac{c^2}{2\beta} + 2y - c\sqrt{\frac{y}{\beta}} \tag{4.129}$$

so that we have

$$F(\rho, [\mathcal{P}_{\sigma,\mathcal{N}} \circ \mathcal{N}](\rho)) = \begin{cases} \frac{1}{d_D} & x < \frac{c^2}{2\beta} + 2y - c\sqrt{\frac{y}{\beta}} \\ e^{V(c\sqrt{-\frac{c^2}{4\beta^2} + \frac{x-y}{\beta}} - x + y)} & \frac{c^2}{2\beta} + 2y - c\sqrt{\frac{y}{\beta}} < x < \frac{c^2}{2\beta} + y \\ 1 & x > \frac{c^2}{2\beta} + y \end{cases} \tag{4.130}$$

t_r is increasingly earlier for larger $\log d_D$, and eventually becomes equal to t_b.

(4.127), (4.128) and (4.130) are derived with several approximations for the thermodynamic limit. We plot a more exact expression resulting from the calculations of Section 4.4.4 for finite but large volumes in Fig. 4.10. Note in particular that for the case where d_D is large, we see the initial exponential growth of fidelity $e^{S^{(\mathrm{eq})}_{\frac{1}{2},R} - S^{(\mathrm{eq})}_{2,B'} - \log d_D}$ predicted by (4.124). For a small diary, the curves obtained from increasing system sizes with fixed $\log d_D$ gradually approach an increasingly sharp transition, as expected from (4.127).

The discussion in this section so far has entirely been for the canonical ensemble case. In Section 4.4.4, we also consider $\mathcal{I}_\alpha$ in the microcanonical example, corresponding to a case where the radiation is divided into two parts R_1 and R_2 such that there is energy conservation between R_1 and B' while R_2 is at infinite temperature. For a small diary, we again find that the fidelity grows from $\frac{1}{d_D}$ to 1 rapidly at t_p. For a large diary, we find that the fidelity reaches 1 at t_{p_2}, and that there is a time scale $t_r < t_p$ when the fidelity starts to grow above $\frac{1}{d_D}$. But for this example, t_r does not seem to be related to the time t_b at which $\mathcal{E}(R_1, R_2)$ starts to grow in any regime.

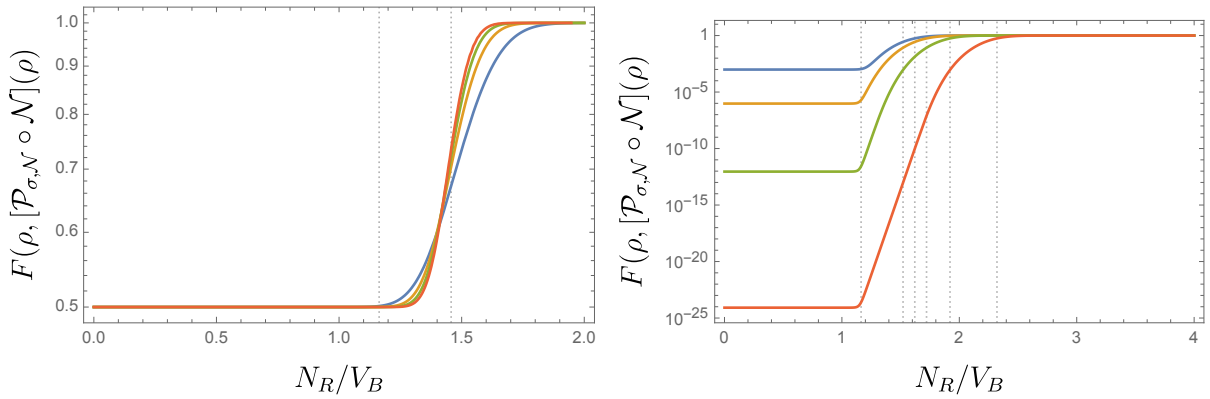


Figure 4.10: Left: The fidelity of the Petz recovery map is shown for a diary consisting of a single qubit. The different plots are for different $V = 100$, 200, 300, and 400 (from blue to red). The transition approaches the second vertical line, which corresponds to t_p in the thermodynamic limit. The first vertical line is t_b. Right: The fidelity of the Petz recovery map for large diaries of sizes 10, 20, 40, and 80 qubits (top to bottom). The first two vertical lines denote t_b and t_p which are separated even in the thermodynamic limit. We also display the t_{p_2}'s for the respective diary sizes, always with $V = 100$. We take R to be at infinite temperature and B' to be at finite temperature with Cardy-like density of states $\rho(E) = e^{V\sqrt{\frac{E}{V}}}$ in both plots, taking $\beta = 1/2$. Notably, the fidelity increases significantly between t_b and t_p and does not saturate to one until after t_{p_2}.

4.4.4 Evolution of Petz map fidelity at finite temperature

Canonical ensemble

To evaluate the fidelity of the Petz map reconstruction $F(\rho, [\mathcal{P}_{\sigma,\mathcal{N}} \circ \mathcal{N}](\rho))$ systematically at finite temperature for the canonical ensemble, we use a method similar to Section 3 of [157]. We take the reference state σ_D to be maximally mixed and ρ_D to be pure. For simplicity, we take R to be at infinite temperature, so that the effective identity operator is

$$\mathcal{I}_\alpha = e^{-\beta H_{B'}} \otimes \mathbf{1}_R \,. \tag{4.131}$$

It is useful to define

$$F_{n_1,n_2} \equiv F^c_{n_1,n_2} + F^d_{n_1,n_2} \tag{4.132}$$

for two non-negative integers n_1, n_2, where

$$F^c_{n_1,n_2} \equiv \frac{1}{Z_1^{n_1+n_2+2}} \frac{1}{d_D^{n_1+n_2+2}} \sum_{\tau \in P_c} \langle \eta_R \otimes e_{B'} | \mathcal{I}_\alpha, \tau \rangle \, d_D^{C(\tau)} \tag{4.133}$$

$$F^d_{n_1,n_2} \equiv \frac{1}{Z_1^{n_1+n_2+2}} \frac{1}{d_D^{n_1+n_2+3}} \sum_{\tau \in P_d} \langle \eta_R \otimes e_{B'} | \mathcal{I}_\alpha, \tau \rangle \, d_D^{C(\tau)} \tag{4.134}$$

where P_c refers to the set of permutations in $S_{n_1+n_2+2}$ such that the first and n_1+2-th element are in the same cycle, and P_d refers to the rest. The equilibrium approximation for the fidelity is then given by[13]

$$F(\rho, [\mathcal{P}_{\sigma,\mathcal{N}} \circ \mathcal{N}](\rho)) = F^c + F^d, \quad F^{c,d} = \lim_{n_1 \to -\frac{1}{2}, n_2 \to -\frac{1}{2}} F^{c,d}_{n_1,n_2} \tag{4.135}$$

13. Note that here we use the alternative analytic continuation from (4.90).

Note that

$$F^c_{n_1,n_2} + d_D F^d_{n_1,n_2} = \frac{1}{(Z_1 d_D)^{n_1+n_2+2}} \sum_{\tau \in S_{n_1+n_2+2}} \langle \eta_R \otimes e_{B'} \otimes e_{D'} | (\mathcal{I}_\alpha)_{B'R} \otimes \mathbf{1}_{D'}, \tau \rangle$$

$$= \mathcal{Z}_{n_1+n_2+2,R} \tag{4.136}$$

where we get the last expression by identifying the second-to-last expression to be the equilibrium approximation for $\mathcal{Z}_{n_1+n_2+2,R}$ for an equilibrated pure state $|\Psi\rangle$ in a Hilbert space $\mathcal{H}_{B'R} \otimes \mathcal{H}_{D'}$, with the effective identity operator $(\mathcal{I}_\alpha)_{B'R} \otimes \mathbf{1}_{D'}$, where D' is an auxiliary system with Hilbert space dimension d_D. This implies that

$$F^c + d_D F^d = \mathcal{Z}_{1,R} = 1\,, \tag{4.137}$$

so that it is sufficient to calculate F^c and express the fidelity as

$$F(\rho, [\mathcal{P}_{\sigma,\mathcal{N}} \circ \mathcal{N}](\rho)) = \left(1 - \frac{1}{d_D}\right) F^c + \frac{1}{d_D}\,. \tag{4.138}$$

In order to systematically obtain F^c using $F^c_{n_1,n_2}$, we can define the generating functional

$$F^c(\lambda_1, \lambda_2) = \sum_{n_1,n_2=0}^{\infty} \frac{F^c_{n_1,n_2}}{\lambda_1^{n_1} \lambda_2^{n_2}}\,. \tag{4.139}$$

Then for non-negative integers n_1, n_2,

$$F^c_{n_1,n_2} = \oint_\infty \frac{d\lambda_1}{2\pi i} \frac{d\lambda_2}{2\pi i}\, \lambda_1^{n_1-1} \lambda_2^{n_2-1}\, F^c(\lambda_1, \lambda_2) \tag{4.140}$$

where the contour is taken to be around the point at infinity. Now in all permutations τ contributing to (4.139), the first and $(n_1+2)^{th}$ element are in a common cycle. Suppose this cycle also includes m_1 elements out of the n_1 elements between the first and $(n_1+2)^{th}$

element, and m_2 elements after the $(n_1+2)^{th}$ element. We can then consider the total contribution from a fixed m_1, m_2 to all n_1, n_2, as shown in Figure 4.11. Note that in the process we have introduced factors of $R(\lambda)$, the resolvent for $\rho_R \equiv \mathrm{Tr}_{B'D'}[|\Psi\rangle\langle\Psi|]$, with $|\Psi\rangle$ as defined above (4.137).

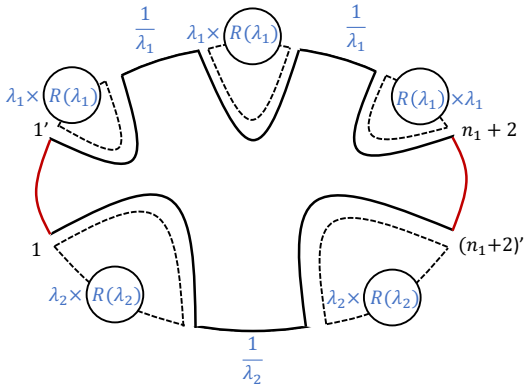


Figure 4.11: Contribution from a fixed $m_1 = 2$ and $m_2 = 1$ to all n_1 and n_2 in the expression for $F^c(\lambda_1, \lambda_2)$ in (4.139). The solid loop gives a factor of $\frac{d_D Z_{m_1+m_2+2,B'}}{(d_D d_R Z_{1,B'})^{m_1+m_2+2}}$, and we get m_1, m_1+1, m_2 and m_2+1 factors respectively of $\frac{1}{\lambda_1}$, $\lambda_1 R(\lambda_1)$, $\frac{1}{\lambda_2}$ and $\lambda_2 R(\lambda_2)$, leading to (4.141).

We then find

$$F^c(\lambda_1,\lambda_2) = \sum_{m_1,m_2=0}^{\infty} \frac{d_D Z_{m_1+m_2+2,B'}}{(Z_{1,B'}\, d_R\, d_D)^{m_1+m_2+2}}\, R(\lambda_1)^{m_1+1} R(\lambda_2)^{m_2+1}\, \lambda_1\lambda_2\,. \tag{4.141}$$

and using (4.140),

$$\begin{aligned} F^c_{n_1,n_2} = d_D \int_0^\infty dE\, \rho(E) \left(\oint_\infty \frac{d\lambda}{2\pi i} \lambda^{n_1} R(\lambda) \frac{e^{-\beta E}}{Z_{1,B'} d_D} \frac{1}{d_R - \frac{e^{-\beta E} R(\lambda)}{Z_{1,B'} d_D}} \right) \\ \times \left(\oint_\infty \frac{d\lambda}{2\pi i} \lambda^{n_2} R(\lambda) \frac{e^{-\beta E}}{Z_{1,B'} d_D} \frac{1}{d_R - \frac{e^{-\beta E} R(\lambda)}{Z_{1,B'} d_D}} \right) \end{aligned} \tag{4.142}$$

Note that $\lambda^{-\frac{1}{2}}$ has a branch point at infinity, so the above expression cannot be analytically

continued as written. However, we can first deform the contour for integer n_1, n_2 to surround the branch cut on the real axis coming from $R(\lambda)$, and then analytically continue to $n_1 = n_2 = -\frac{1}{2}$, to get

$$F^c = d_D \int_0^\infty dE\, \rho(E) \left(\oint_{\mathcal{C}} \frac{d\lambda}{2\pi i} \lambda^{-\frac{1}{2}} R(\lambda) \frac{e^{-\beta E}}{Z_{1,B'} d_D} \frac{1}{d_R - \frac{e^{-\beta E} R(\lambda)}{Z_{1,B'} d_D}} \right)^2 \tag{4.143}$$

where $\mathcal{C}$ is the contour that surrounds the branch cut on the real axis. In order to further simplify this expression, we use the following approximation for the resolvent:

$$R(\lambda) \approx \frac{d_R}{\lambda - \lambda_0}\,, \tag{4.144}$$

where λ_0 is the smallest eigenvalue of ρ_R, which turns out to be

$$\lambda_0 = \frac{1}{d_R} g, \quad g = \frac{1}{Z_{1,B'}} \int_{E_0}^\infty dE\, \rho(E)\, e^{-\beta E} \tag{4.145}$$

with E_0 defined through

$$\int_0^{E_0} dE\, \rho(E) = \frac{d_R}{d_D}\,. \tag{4.146}$$

In deriving (4.144)-(4.146), we use similar methods to [157], which we review for this context at the end of this subsection. Putting (4.144) into (4.143), we find

$$F^c \approx \frac{1}{Z_{1,B'}} \int_0^\infty dE\, \rho(E)\, \frac{e^{-2\beta E}}{d_D Z_{1,B'} \lambda_0 + e^{-\beta E}}\,. \tag{4.147}$$

Now to further analyse (4.147), let us take the density of states to be as in AdS_3, $\rho(E) = e^{cV\sqrt{E/V}}$, where V is the volume of B'. Then $Z_{1,B'}$ can be evaluated using the saddle point

approximation and has the behavior

$$Z_{1,B'} = c' V^{\frac{1}{2}} e^{-\beta E_c + cV\sqrt{\epsilon_c}} = c' V^{\frac{1}{2}} e^{\frac{c^2 V}{4\beta}}, \quad E_c = V\epsilon_c, \quad \sqrt{\epsilon_c} = \frac{c}{2\beta}, \quad S = V\frac{c^2}{2\beta} \tag{4.148}$$

where S is the thermal entropy and c' is some $O(1)$ constant. The second equilibrium Renyi entropy can also be evaluated from the saddle point

$$e^{-S_{2,B'}} = \frac{Z_{1,B'}(2\beta)}{Z_{1,B'}^2(\beta)} \sim V^{-\frac{1}{2}} e^{-\frac{3c^2 V}{8\beta}}, \quad S_{2,B'} = \frac{3c^2 V}{8\beta} \ . \tag{4.149}$$

Now if we define

$$x \equiv \frac{\log d_R}{V}, \quad y \equiv \frac{\log d_D}{V}, \tag{4.150}$$

then the various relevant time scales are given by:

$$t_b : \quad x = \frac{3c^2}{8\beta}, \quad t_p : \quad x = \frac{c^2}{2\beta}, \quad t_{p_2} : \quad x = \frac{c^2}{2\beta} + y \ . \tag{4.151}$$

We will always take $x \sim O(1)$ while take $V \to \infty$, but y can be $O(V^{-1})$ or $O(1)$ depending on whether the diary is "small" or "large."

Now let us examine the behavior of g defined in (4.145). Note that for $\epsilon_0 = E_0/V$,

$$V\int_0^{\epsilon_0} d\epsilon \, e^{cV\sqrt{\epsilon}} = \frac{2}{cV}\sqrt{\epsilon_0} e^{cV\sqrt{\epsilon_0}} = \frac{d_R}{d_D} \tag{4.152}$$

$$\Rightarrow \quad \sqrt{\epsilon_0} \approx \frac{x-y}{c} + O(V^{-1}\log V), \quad . \tag{4.153}$$

so that g has the form

$$g = \frac{1}{2}\left[1 + \mathrm{erf}\left(\sqrt{V\beta}(\sqrt{\epsilon_c} - \sqrt{\epsilon_0})\right)\right] + \sqrt{\frac{\beta}{\pi c^2 V}} e^{-V\beta(\sqrt{\epsilon_c}-\sqrt{\epsilon_0})^2} \tag{4.154}$$

$$= \begin{cases} 1 + O(e^{-V}) & x - y < \frac{c^2}{2\beta} - O(V^{-\frac{1}{2}}) \\ \frac{\sqrt{\frac{\epsilon_0}{\epsilon_c}} e^{-V\beta(\sqrt{\epsilon_c}-\sqrt{\epsilon_0})^2}}{2\sqrt{\pi V\beta}(\sqrt{\epsilon_0}-\sqrt{\epsilon_c})} & x - y > \frac{c^2}{2\beta} + O(V^{-\frac{1}{2}}) \end{cases} . \tag{4.155}$$

During the evaporation process, $\log d_R$ increases while V decreases. We would like to find the evolution of F^c as a function of x to see how the information recovery improves during the evaporation process.

Note first that from (4.155), for $x - y > \frac{c^2}{2\beta} + O(V^{-\frac{1}{2}})$, the first term in the denominator of the integrand of (4.147) can always be ignored relative to the second, implying that the $F^c \approx 1$ and hence $F \approx 1$ at t_{p_2} and later times. For $t < t_{p_2}$, $\lambda_0 \approx \frac{1}{d_R}$, and we can simplify (4.147) in this regime to

$$F^c \approx \sqrt{V} e^{-V\frac{c^2}{4\beta}} \int_0^\infty d\epsilon\, e^{cV\sqrt{\epsilon}} \frac{e^{-2\beta V\epsilon}}{e^{V(\frac{c^2}{4\beta} - x + y)} + e^{-\beta V\epsilon}} . \tag{4.156}$$

The second term in the denominator dominates for

$$\epsilon < \epsilon_u \equiv -\frac{c^2}{4\beta^2} + \frac{x-y}{\beta} \tag{4.157}$$

and the first term dominates otherwise. (Note that the regime (4.157) exists only for $x - y \geq \frac{c^2}{4\beta}$.) So we can write (ignoring sub-exponential prefactors)

$$F^c \sim e^{-Vc^2/(4\beta)} \int_0^{\epsilon_u} e^{cV\sqrt{\epsilon}} e^{-\beta V\epsilon} + e^{-Vc^2/(4\beta)} e^{\beta V\epsilon_u} \int_{\epsilon_u}^\infty e^{cV\sqrt{\epsilon}} e^{-2\beta V\epsilon}. \tag{4.158}$$

Recall that the saddle point for the first integral is at $\epsilon_c = \frac{c^2}{4\beta^2}$, and the saddle point for the second integral is at $\epsilon_c^{(2)} = \frac{c^2}{16\beta^2}$. We consider three regimes:

1. $x - y < \frac{c^2}{4\beta}$: We only have the second term in (4.158), which can be approximated with its saddle point, so

$$F^c \sim e^{V(-\frac{3c^2}{8\beta}+x-y)}, \quad x - y < \frac{c^2}{4\beta} \tag{4.159}$$

2. $\frac{c^2}{4\beta} < x - y < \frac{5c^2}{16\beta}$: Here the first term should be approximated with the boundary value at ϵ_u, and the second term with the saddle point at $\epsilon_c^{(2)}$. So we find

$$F^c \sim e^{V(c\sqrt{-\frac{c^2}{4\beta^2}+\frac{x-y}{\beta}}-x+y)} + e^{V(-\frac{3c^2}{8\beta}+x-y)} \tag{4.160}$$

The second term is always greater than the first in this regime, so we have

$$F^c \approx e^{V(-\frac{3c^2}{8\beta}+x-y)}, \quad \frac{c^2}{4\beta} < x - y < \frac{5c^2}{16\beta}\,. \tag{4.161}$$

3. $\frac{5c^2}{16\beta} < x - y < \frac{c^2}{2\beta}$: In this regime, both terms are approximated with their boundary values, which turn out to be the same, so we have

$$F^c \approx 2e^{V(c\sqrt{-\frac{c^2}{4\beta^2}+\frac{x-y}{\beta}}-x+y)}, \quad \frac{5c^2}{16\beta} < x - y < \frac{c^2}{2\beta} \tag{4.162}$$

For a small diary, we can neglect y in (4.159)-(4.162), and we therefore find that $(1 - \frac{1}{d_D})F^c$ is exponentially suppressed in volume before t_p, while $\frac{1}{d_D}$ is $O(1)$. We therefore have

$$F(\rho, [\mathcal{P}_{\sigma,\mathcal{N}} \circ \mathcal{N}](\rho)) \approx \begin{cases} \frac{1}{d_D} & x < \frac{c^2}{2\beta} \\ 1 & x > \frac{c^2}{2\beta} \end{cases} \tag{4.163}$$

So the fidelity for the recovery of a small diary improves rapidly from $\frac{1}{d_D}$ to 1 at the Page time.

For a diary sufficiently large such that there is a regime where $\frac{3c^2}{8\beta} < x < \frac{5c^2}{16\beta} + y$, we have

$$F(\rho, [\mathcal{P}_{\sigma,\mathcal{N}} \circ \mathcal{N}](\rho)) = \begin{cases} \frac{1}{d_D} & x < \frac{3c^2}{8\beta} \\ e^{V(-\frac{3c^2}{8\beta}+x-y)} = e^{\log d_R - S_{2,B'} - \log d_D} & \frac{3c^2}{8\beta} < x < \frac{5c^2}{16\beta} + y \\ e^{V(c\sqrt{-\frac{c^2}{4\beta^2}+\frac{x-y}{\beta}}-x+y)} & \frac{5c^2}{16\beta} + y < x < \frac{c^2}{2\beta} + y \\ 1 & x > \frac{c^2}{2\beta} + y \end{cases}. \tag{4.164}$$

So in this case, the fidelity starts increasing from its initial value of $\frac{1}{d_D}$ at time t_b (independently of $\log d_D$ as long as there is a non-trivial regime corresponding to the second line), and the initial improvement is exponential in $\log d_R - S_{2,B'}$, precisely as predicted by the leading correction in (4.124).

For a diary that is $O(1)$ but not sufficiently large such that there is a regime where $\frac{3c^2}{8\beta} < x < \frac{5c^2}{16\beta} + y$, we have

$$F(\rho, [\mathcal{P}_{\sigma,\mathcal{N}} \circ \mathcal{N}](\rho)) = \begin{cases} \frac{1}{d_D} & x < \frac{c^2}{2\beta} + 2y - c\sqrt{\frac{y}{\beta}} \\ e^{V(c\sqrt{-\frac{c^2}{4\beta^2}+\frac{x-y}{\beta}}-x+y)} & \frac{c^2}{2\beta} + 2y - c\sqrt{\frac{y}{\beta}} < x < \frac{c^2}{2\beta} + y \\ 1 & x > \frac{c^2}{2\beta} + y \end{cases}. \tag{4.165}$$

So in this case, the fidelity starts to increase from $\frac{1}{d_D}$ at a time

$$t_r : \quad x = \frac{c^2}{2\beta} + 2y - c\sqrt{\frac{y}{\beta}}, \tag{4.166}$$

which is macroscopically earlier than t_p when y is $O(1)$, and becomes increasingly earlier as the diary becomes larger.

Further, note that in (4.158), we can evaluate both integrals exactly using error functions.

The first integral is approximately

$$F^c \approx \frac{\sqrt{\pi}c\left(1-\text{erf}\left(\frac{\sqrt{V}\left(c-\sqrt{4\beta(x-y)-c^2}\right)}{2\sqrt{\beta}}\right)\right)}{2\beta^{3/2}}. \tag{4.167}$$

Near t_{p_2}, this term dominates over the second integral. This shows that the transition regime is $O((V\beta)^{-\frac{1}{2}})$ in $x-y$, as this is the regime where the error function becomes $O(1)$.

We can also use (4.146) and (4.145) to get an exact expression for λ_0 in terms of d_R, d_D, V, which can then be used to evaluate the integral (4.147) for large values of V numerically. This leads to the evolution of the fidelity shown in Fig. 4.10.

Let us now justify the approximation we used for $R(\lambda)$ in (4.144). We use similar steps to Section 2 of [157]. The Schwinger-Dyson equation for $R(\lambda)$ is

$$\lambda R = d_R + \sum_{n=1}^{\infty} d_D Z_{n,B'} \frac{R^n}{d_R^n Z_{1,B'}^n d_D^n} \tag{4.168}$$

$$\Rightarrow \lambda R = d_R + d_D R \int_0^{\infty} dE\, \rho(E)\, \frac{e^{-\beta E}}{d_R d_D Z_{1,B'} - e^{-\beta E} R} \tag{4.169}$$

It is useful to first find the smallest eigenvalue λ_0 of ρ_R. For $\lambda < \lambda_0$, we can see from the definition of R that R is real and negative, and we can write

$$\lambda \approx \frac{d_R}{R} - \frac{d_D}{R}\int_0^{E_0'} dE\, \rho(E) + d_D \int_{E_0'}^{\infty} dE\, \rho(E)\, \frac{e^{-\beta E}}{d_R d_D Z_{1,B'}} \tag{4.170}$$

where E_0' for a given set of $d_R, d_D, V_{B'}$ and λ is defined implicitly by

$$-e^{-\beta E_0'} = \frac{d_R d_D Z_{1,B'}}{R(\lambda)}\,. \tag{4.171}$$

In the two terms in (4.170), we have assumed that different terms dominate in the denominator of the integrand of (4.169). Since $\lambda = \lambda_0$ is a branch point of $R(\lambda)$, we have $\frac{d\lambda}{dR} = 0$

at this point, which implies

$$\int_0^{E_0} dE\,\rho(E) \approx \frac{d_R}{d_D} \tag{4.172}$$

where we use E_0 to refer to the value of E_0' at λ_0. We can treat (4.146) as the definition of E_0. In terms of E_0, λ_0 is given by

$$\lambda_0 = \frac{1}{d_R Z_{1,B'}} \int_{E_0}^{\infty} dE\,\rho(E)\, e^{-\beta E} \tag{4.173}$$

Let us now return to (4.169) to obtain an approximation for the resolvent. We can divide the integral into two parts at E_0:

$$\begin{aligned} \lambda R = d_R + d_D R \int_0^{E_0} dE\,\rho(E)\, \frac{e^{-\beta E}}{d_R d_D Z_{1,B'} - e^{-\beta E} R} \\ + d_D R \int_{E_0}^{\infty} dE\,\rho(E)\, e^{-\beta E} \frac{e^{-\beta E}}{d_R d_D Z_{1,B'} - e^{-\beta E} R} \end{aligned} \tag{4.174}$$

Note that $e^{-\beta E_0} R(\lambda_0) = -d_R d_D Z_{1,B'}$, and $e^{-\beta E}|R(\lambda)| < e^{-\beta E_0}|R(\lambda_0)|$ for $E > E_0$ and $\lambda > \lambda_0$, using the definition of R. Hence, the denominator in the second term of the integrand can be approximated as $d_R d_D Z_{1,B'}$, and we have

$$\begin{aligned} \lambda R &\approx d_R + d_D R \int_0^{E_0} dE\,\rho(E)\, \frac{e^{-\beta E}}{d_R d_D Z_{1,B'} - e^{-\beta E} R} \\ &\quad + \frac{1}{d_R Z_{1,B'}} R \int_{E_0}^{\infty} dE\,\rho(E)\, e^{-\beta E} \\ \Rightarrow \lambda R &\approx d_R + d_D R \int_0^{E_0} dE\,\rho(E)\, \frac{e^{-\beta E}}{d_R d_D Z_{1,B'} - e^{-\beta E} R} + R\lambda_0 \\ \Rightarrow R &= \frac{d_R}{\lambda - \lambda_0} + \frac{d_D}{\lambda - \lambda_0} R \int_0^{E_0} dE\,\rho(E)\, \frac{e^{-\beta E}}{d_R d_D Z_{1,B'} - e^{-\beta E} R} \end{aligned} \tag{4.175}$$

Treating the second term as a perturbation, we find to first order

$$R \approx \frac{d_R}{\lambda - \lambda_0} + \frac{d_D}{\lambda - \lambda_0} \int_0^{E_0} dE \rho(E)\, \frac{e^{-\beta E}}{Z_{1,B'} d_D}\, \frac{1}{\lambda - \lambda_0 - e^{-\beta E}/(Z_{1,B'} d_D)} \tag{4.176}$$

Let us now understand whether our approximation in treating the second term as a perturbation is self-consistent. Suppose we require that for some small ϵ,

$$\lambda - \lambda_0 - \frac{e^{-\beta\epsilon}}{Z_{1,B'} d_D} > 0\,. \tag{4.177}$$

The integral in the second term can be divided into two parts,

$$\begin{aligned} I = &\frac{d_D}{\lambda - \lambda_0} \int_0^{\epsilon} dE \rho(E) \frac{e^{-\beta E}}{Z_{1,B'} d_D} \frac{1}{\lambda - \lambda_0 - e^{-\beta E}/(Z_{1,B'} d_D)} \\ &+ d_D \int_{\epsilon}^{E_0} dE \rho(E) \frac{e^{-\beta E}/(Z_{1,B'} d_D)}{\lambda - \lambda_0} \frac{1}{\lambda - \lambda_0 - e^{-\beta E}/(Z_{1,B'} d_D)} \end{aligned} \tag{4.178}$$

If ϵ is small, the first term is small. The integrand in the second term is positive and $< \rho(E)$ due to (4.177), so from (4.146), the second term is $< d_R$, and hence smaller than the leading term in (4.176). So the approximation (4.176) is valid as long as (4.177) is satisfied for small ϵ, and to a first approximation, we can ignore the second term in (4.176) and write $R(\lambda)$ as in (4.144).

Microcanonical ensemble

Let us now consider the question of Petz map fidelity taking the effective identity operator to be

$$\mathcal{I}_\alpha = \mathbf{1}_{R_2} \otimes \sum_{E_p^{R_1} + E_r^B \in I_{E,\Delta}} (|p\rangle \langle p|)_{R_1} \otimes (|r\rangle \langle r|)_B. \tag{4.179}$$

with R_1 and R_2 corresponding to two parts of the radiation, and B to the black hole. We again take ρ_D to be pure and σ_D to be maximally mixed. R_2 is taken to be at infinite temperature while there is energy conservation between R_1 and B. We take the average energy density ϵ in $R_1 B$ to be

$$\epsilon = \frac{E}{V_{R_1} + V_B} \tag{4.180}$$

and view the evaporation process as a process where $S_R^{(\mathrm{eq})} = V_{R_1} s(\epsilon) + \log d_{R_2}$ increases while $S_B^{(\mathrm{eq})} = V_B s(\epsilon)$ decreases. t_p is the time when

$$t_p : \quad V_{R_1} s(\epsilon) + \log d_{R_2} = V_B s(\epsilon) \tag{4.181}$$

and t_{p_2} is defined as the time when

$$t_{p_2} : \quad V_{R_1} s(\epsilon) + \log d_{R_2} = V_B s(\epsilon) + \log d_D \ . \tag{4.182}$$

Recall that in this setup, the time t_b at which the logarithmic negativity starts to grow is given for sufficiently small $\lambda \equiv \frac{\log d_{R_2}}{V_{R_1} s(\epsilon) + \log d_{R_2}}$ by

$$t_b : \quad \log d_{R_2} + 2S^{(\mathrm{eq})}_{\frac{1}{3}, R_1} - V_{R_1} s(\epsilon) = V_B s(\epsilon) \ . \tag{4.183}$$

The equilibrium approximation for the Petz map fidelity is given by

$$F(\rho, [\mathcal{P}_{\sigma,\mathcal{N}} \circ \mathcal{N}](\rho)) = \lim_{m \to \frac{1}{2}} F_m^E, \tag{4.184}$$

where

$$\begin{aligned} F_m^E &\equiv \sum_{E_1} \frac{1}{d_D^{2m+3} (N_E d_{R_2})^{2m+2}} \sum_{\tau \in \mathcal{S}_{2m+2}} (d_{E_1}^{R_1} d_{R_2})^{C(\eta^{-1} \circ \tau)} (d_{E-E_1}^{B'})^{C(\tau)} d_D^{C(\tau)+\zeta(\tau)} \\ &= \sum_{E_1} \left(\frac{d_{E_1}^{R_1} d_{E-E_1}^{B'}}{N_E} \right)^{2m+2} F_m^{\infty}(d_D,\, d_{E_1}^{R_1} d_{R_2},\, d_{E-E_1}^{B'}) \end{aligned} \tag{4.185}$$

with

$$N_E \equiv \sum_{E_1} d_{E_1}^{R_1} d_{E-E_1}^{B'} \approx c_1 e^{(V_1 + V_B) s(\epsilon)} \tag{4.186}$$

where c_1 and all c_i we will introduce below are $O(1)$ constants, and

$$F_m^{\infty}(d_D, d_R, d_{B'}) \equiv \frac{1}{d_D^{2m+3}(d_R d_B)^{2m+2}} \sum_{\tau \in \mathcal{S}_{2m+2}} d_R^{C(\eta^{-1}\circ\tau)} d_{B'}^{C(\tau)} d_D^{C(\tau)+\zeta(\tau)} \tag{4.187}$$

can be identified to be the infinite-temperature value of F_m in (4.102). Then using (4.184) and (4.120),

$$F(\rho, [\mathcal{P}_{\sigma,\mathcal{N}} \circ \mathcal{N}](\rho)) = \frac{1}{d_D} + \left(1 - \frac{1}{d_D}\right)(F_1 + F_2) \tag{4.188}$$

where

$$F_1 = \frac{1}{d_D}\frac{d_{R_2}}{N_E} \sum_{E_1 < V_{R_1}\epsilon_u} (d_{E_1}^{R_1})^2 \; {}_2F_1\left(\frac{1}{2}, -\frac{1}{2}, 2, \frac{d_{R_2} d_{E_1}^{R_1}}{d_{E-E_1}^{B'} d_D}\right)^2, \tag{4.189}$$

$$F_2 = \sum_{E_1 > V_{R_1}\epsilon_u} \frac{d_{E_1}^{R_1} d_{E-E_1}^{B'}}{N_E} \; {}_2F_1\left(\frac{1}{2}, -\frac{1}{2}, 2, \frac{d_{E-E_1}^{B'} d_D}{d_{R_2} d_{E_1}^{R_1}}\right)^2 \tag{4.190}$$

and ϵ_u is defined as the solution to

$$V_{R_1} s(\epsilon_u) + \log d_{R_2} = \log d_D + V_B s(\bar{\epsilon}_u), \quad \bar{\epsilon}_u = \frac{E - V_{R_1}\epsilon_u}{V_B}. \tag{4.191}$$

We note that F_1 is always dominated by the value at upper limit of the sum in (4.189),

$$F_1 = c_2 e^{\Lambda_1}, \quad \Lambda_1 = 2V_{R_1} s(\epsilon_u) - (V_{R_1} + V_B) s(\epsilon) + \log d_{R_2} - \log d_D = \Lambda_2, \tag{4.192}$$

$$\Lambda_2 = V_{R_1}(s(\epsilon_u) - s(\epsilon)) + V_B(s(\bar{\epsilon}_u) - s(\epsilon)) \tag{4.193}$$

and F_2 is dominated by either its saddle point or its lower limit, i.e.

$$F_2 = \begin{cases} 1 & \epsilon_u < \epsilon \\ c_3 e^{\Lambda_2} & \epsilon_u > \epsilon \end{cases}. \tag{4.194}$$

Note that by definition $\Lambda_2 \leq 0$, and thus F_1 is always exponentially suppressed except at $\epsilon_u = \epsilon$.

From (4.191) when we decrease the values of V_1 and $\log d_{R_2}$ (i.e. going to earlier times), the value of ϵ_u should increase. We can assume that ϵ_u is a smooth and monotonically decreasing function of t. At t_{p2} we have $\epsilon_u = \epsilon$ while for $t < t_{p2}$ ($t > t_{p2}$) we have $\epsilon_u > \epsilon$ ($\epsilon_u < \epsilon$). From (4.192)–(4.194), we then conclude that for any d_D, $F_1 + F_2$ is exponentially small in volume for $t < t_{p_2}$ and becomes 1 at t_{p_2} over a very short range of time (volume suppressed).

Then for a small diary, precisely as in (4.163) for the canonical example, we have

$$F(\rho, [\mathcal{P}_{\sigma,\mathcal{N}} \circ \mathcal{N}](\rho)) = \begin{cases} \frac{1}{d_D} & t < t_p \\ 1 & t > t_p \end{cases}. \tag{4.195}$$

For a large diary, we have

$$F(\rho, [\mathcal{P}_{\sigma,\mathcal{N}} \circ \mathcal{N}](\rho)) = \begin{cases} \frac{1}{d_D} & t < t_r \\ \frac{1}{d_D} e^{V_{R_1}(2s(\epsilon_u) - s(\epsilon)) + \log d_{R_2} - V_B s(\epsilon)} & \\ = e^{V_{R_1}(s(\epsilon_u) - s(\epsilon)) + V_B(s(\bar{\epsilon}_u) - s(\epsilon))} & t_r < t < t_{p_2} \\ 1 & t > t_{p_2} \end{cases}. \tag{4.196}$$

The growth of the fidelity above $1/d_D$ for a large diary starts at the time scale

$$t_r : \quad V_{R_1}(2s(\epsilon_u) - s(\epsilon)) + V_{R_2} s_0 = V_B s(\epsilon) \tag{4.197}$$

Comparing this to t_p defined in (4.181) and noting that $\epsilon_u > \epsilon$ for $t < t_{p_2}$, we can see that $t_r < t_p$. Note that unlike in the canonical ensemble example in the previous subsection, t_r does not converge to t_b defined by (4.183) for sufficiently large diaries. For a given λ,

depending on $\log d_D$, t_r can be either earlier or later than t_b.

CHAPTER 5

THERMALIZATION

"Truth is chaos. Maybe beauty is chaos."

— Interview by Nora Ephron & Susan Edmiston (1965)

5.1 Quantum Chaos and Eigenstate Thermalization

The eigenstate thermalization hypothesis (ETH) was a major development in understanding the emergence of thermal physics from isolated quantum many-body systems in pure states [43, 179, 45]. The statement of eigenstate thermalization is that given two energy eigenstates, $|E_i\rangle$ and $|E_j\rangle$, and a "simple" few-body operator $\mathcal{O}$, the expectation value varies smoothly with the macroscopic, thermodynamic quantities such as energy

$$\langle E_i| \mathcal{O} |E_j\rangle = f_{\mathcal{O}}(E)\delta_{ij} + e^{-S(E)/2}R_{ij}, \quad E := \frac{E_i + E_j}{2}, \tag{5.1}$$

where $f_{\mathcal{O}}(E)$ is a smooth function of the energy, $S(E)$ is the thermodynamic entropy, and R_{ij} is an $O(1)$ pseudorandom matrix. The ETH is expected to hold for generic nonintegrable systems and violated in integrable systems. Morally, it states that expectation values of simple observables appear thermal.

Note that the standard, *local* ETH is a statement only about local or few-body operators. A significant strengthening of the ETH can be made by asserting that the entire reduced density matrix supported on a finite spatial region appears thermal. More precisely, the subsystem eigenstate thermalization hypothesis states that the reduced density matrices of eigenstates, $\rho_A(\psi)$, are exponentially close in trace distance to a universal thermal density matrix, $\rho_{\text{univ}}(E)$, that only depends on the total energy [53, 69]

$$|\rho_A(\psi) - \rho_{\text{univ}}(E)|_1 = O\left(e^{-S(E)/2}\right). \tag{5.2}$$

In addition, "off-diagonal" matrices are exponentially suppressed

$$\left|\mathrm{Tr}_B\left[|E_i\rangle\langle E_j|\right]\right|_1 = \delta_{ij} + O\left(e^{-S(E)/2}\right). \tag{5.3}$$

These conditions imply the local ETH for *all* operators in region A and are significantly stronger [117].

It is important to understand which systems obey the subsystem ETH. Of course, for all systems, when A is the entire system, the subsystem ETH completely fails because the distance between a pure state and a thermal state is $O(1)$. It is then nontrivial to determine at which point the subsystem ETH breaks down and thermal physics no longer applies. In the following, we show that holographic CFTs obey the subsystem ETH whenever A is smaller than half the total system size for a class of high-energy states. More generally, we find generic chaotic Hamiltonian systems, whose eigenstate ansatzes were put forward in Refs. [44, 128, 141], obey the subsystem ETH in its strongest form.

5.1.1 Generic chaotic Hamiltonians

We use the following ansatz for the tensor product decomposition of energy eigenstates of energy E for generic chaotic quantum many-body systems [44]

$$|E\rangle = \mathcal{N}^{-1/2} \sum_{E-\Delta<E_i+E_J<E+\Delta} c_{iJ}\,|E_i\rangle_A\,|E_J\rangle_B\,, \tag{5.4}$$

where $\Delta \ll E$, $\mathcal{N}$ is the normalization, $|E_i\rangle_A$ and $|E_J\rangle_B$ are subsystem energy eigenstates[1], and the coefficients are complex Gaussian random variables

$$\overline{c_{iJ}c^*_{i'J'}} = \delta_{ii'}\delta_{JJ'}, \tag{5.5}$$

where, with the proper normalization, the variance is set to one. The reduced density matrix is

$$\rho_A = \mathcal{N}^{-1} \sum_{E_i-2\Delta<E_j<E_i+2\Delta} \sum_{E-E_i-\Delta<E_J<E-E_i+\Delta} c_{iJ}c^*_{jJ} \left|E_i\right\rangle \left\langle E_j\right|_A . \tag{5.6}$$

Using this ansatz for the reduced density matrix, we perform the replica trick for the PRRE. In analogy with Refs. [44, 128, 141, 52] where the Rényi and von Neumann entropies were evaluated for this ansatz, we find, in analogy with Section 3.2, that in the thermodynamic limit, after ensemble averaging

$$\overline{\mathrm{Tr}\left[\rho_A^\alpha \sigma_A^{1-\alpha}\right]} = \mathcal{N}^{-1} \int d\mathcal{E} e^{S_A(\mathcal{E})+S_B(E-\mathcal{E})} G_\alpha(\mathcal{E}), \tag{5.7}$$

1. In fact, it is not quite correct to consider these subsystem energy eigenstates due to the interaction terms in the Hamiltonian coupling A and B that lead to correlations near the boundary, as emphasized in Ref. [128]. It is more accurate to consider very similar states referred to as "many-body Berry (MBB) states" in Ref. [128]. These are constructed from perturbing an integrable Hamiltonian by an integrability breaking term. In MBB states, the subsystem eigenstates are not *energy* eigenstates, but local product states. The following calculation is unchanged. Moreover, preliminary numerical tests suggest that our results are somewhat universal for non-integrable many-body systems.

where

$$G_\alpha(\mathcal{E}) := \begin{cases} {}_2F_1\left(1-\alpha,-\alpha;2;e^{S_A(\mathcal{E})-S_B(E-\mathcal{E})}\right)\,{}_2F_1\left(\alpha-1,\alpha;2;e^{S_A(\mathcal{E})-S_B(E-\mathcal{E})}\right), \\ \qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad S_A(\mathcal{E}) < S_B(E-\mathcal{E}), \\ \\ e^{S_B(E-\mathcal{E})-S_A(\mathcal{E})}\,{}_2F_1\left(1-\alpha,-\alpha;2;e^{S_B(E-\mathcal{E})-S_A(\mathcal{E})}\right)\,{}_2F_1\left(\alpha-1,\alpha;2;e^{S_B(E-\mathcal{E})-S_A(\mathcal{E})}\right), \\ \qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad S_A(\mathcal{E}) > S_B(E-\mathcal{E}) \end{cases} \tag{5.8}$$

and

$$\mathcal{N} = \int d\mathcal{E} e^{S_A(\mathcal{E})+S_B(E-\mathcal{E})}. \tag{5.9}$$

We use the following ansatz for the thermodynamic entropies

$$S_A(\mathcal{E}) = fVs\left(\frac{\mathcal{E}}{Vf}\right), \quad S_B(E-\mathcal{E}) = (1-f)Vs\left(\frac{E-\mathcal{E}}{V(1-f)}\right), \tag{5.10}$$

where $s(u)$ is the entropy density, V is the volume of the total system and f is the fractional volume of subsystem A. The saddle point equation for the main integral is

$$s'\left(\frac{\mathcal{E}_1}{Vf}\right) = s'\left(\frac{E-\mathcal{E}_1}{V(1-f)}\right) - \frac{\partial_{\mathcal{E}} G_\alpha(\mathcal{E}_1)}{G_\alpha(\mathcal{E}_1)} \tag{5.11}$$

and for the normalization

$$s'\left(\frac{\mathcal{E}_2}{Vf}\right) = s'\left(\frac{E-\mathcal{E}_2}{V(1-f)}\right). \tag{5.12}$$

We can now evaluate the PRRE in various regimes. The saddle point equation for the

normalization is simple to solve because $s'(u)$ is single-valued

$$\mathcal{E}_2 = fE. \tag{5.13}$$

This is not so surprising as it implies a constant energy density. The normalization is then evaluated to

$$\mathcal{N} = \sqrt{\frac{2\pi V f(1-f)}{s''\left(\frac{E}{V}\right)}} e^{Vs\left(\frac{E}{V}\right)} \tag{5.14}$$

We now specify to $f < 1/2$ where we claim that $\mathcal{E}_1 < E/2$. In this regime, we can expand the hypergeometric function

$$G_\alpha(\mathcal{E}) \simeq 1 + \alpha(\alpha-1)e^{S_A(\mathcal{E}) - S_B(E-\mathcal{E})}. \tag{5.15}$$

In this approximation,

$$\frac{\partial_{\mathcal{E}} G_\alpha(\mathcal{E}_1)}{G_\alpha(\mathcal{E}_1)} = \frac{(\alpha-1)\alpha e^{fVs\left(\frac{\mathcal{E}_1}{fV}\right)}\left(s'\left(\frac{E-\mathcal{E}_1}{(1-f)V}\right) + s'\left(\frac{\mathcal{E}_1}{fV}\right)\right)}{e^{(1-f)Vs\left(\frac{E-\mathcal{E}_1}{(1-f)V}\right)} + (\alpha-1)\alpha e^{fVs\left(\frac{\mathcal{E}_1}{fV}\right)}}. \tag{5.16}$$

This term is exponentially small for $f < 1/2$. Therefore, the saddle point equation for $\mathcal{E}_1$ can be treated as an expansion around $\mathcal{E}_2$

$$\mathcal{E}_1 = \mathcal{E}_2(1+\delta), \quad \delta \ll 1. \tag{5.17}$$

To leading order, the saddle point equation is

$$\begin{aligned}\frac{Es''\left(\frac{E}{V}\right)}{V}\delta &= -\frac{Es''\left(\frac{E}{V}\right)f}{V(1-f)}\delta - \frac{2(\alpha-1)\alpha e^{(2f-1)Vs\left(\frac{E}{V}\right)}s'\left(\frac{E}{V}\right)}{(\alpha-1)\alpha e^{(2f-1)Vs\left(\frac{E}{V}\right)}+1} \\ &\simeq -\frac{Es''\left(\frac{E}{V}\right)f}{V(1-f)}\delta - 2(\alpha-1)\alpha e^{(2f-1)Vs\left(\frac{E}{V}\right)}s'\left(\frac{E}{V}\right). \end{aligned} \tag{5.18}$$

Solving for δ, we find self-consistency with the claim that $\mathcal{E}_1$ is very close to $\mathcal{E}_2$

$$\delta \simeq \frac{2(\alpha-1)\alpha(f-1)Ve^{(2f-1)Vs\left(\frac{E}{V}\right)}s'\left(\frac{E}{V}\right)}{Es''\left(\frac{E}{V}\right)}. \tag{5.19}$$

The saddle point solution is then

$$\int d\mathcal{E}e^{S_A(\mathcal{E})+S_B(E-\mathcal{E})}G_\alpha(\mathcal{E}) \simeq \sqrt{\frac{2\pi Vf(1-f)}{s''\left(\frac{E}{V}\right)}}e^{Vs\left(\frac{E}{V}\right)+\alpha(\alpha-1)e^{(2f-1)Vs\left(\frac{E}{V}\right)}+O(\delta^2)}. \tag{5.20}$$

Therefore, the moments are

$$\overline{\mathrm{Tr}\left[\rho_A^\alpha\sigma_A^{1-\alpha}\right]} \simeq e^{\alpha(\alpha-1)e^{(2f-1)Vs\left(\frac{E}{V}\right)}}, \tag{5.21}$$

so the PRRE is exponentially suppressed in the entropy for all values of α

$$\overline{D_\alpha(\rho_A||\sigma_A)} \simeq \alpha e^{(2f-1)Vs\left(\frac{E}{V}\right)}. \tag{5.22}$$

This places strict bounds on the trace distance

$$\frac{e^{(2f-1)Vs\left(\frac{E}{V}\right)}}{4} \leq \overline{T(\rho_A||\sigma_A)} \leq \frac{e^{(f-1/2)Vs\left(\frac{E}{V}\right)}}{\sqrt{2}}. \tag{5.23}$$

This provides further evidence for the refined subsystem eigenstate thermalization hypothesis in Ref. [53] that postulated the scaling of the trace distance on the size of subregion A because the upper bound precisely matches that scaling when replacing N_A, the number of qubits in region A, with the subsystem thermodynamic entropy.

Note also that these results generalize our result for random matrix theory because at infinite temperature, we can identify

$$fVs\left(\frac{E}{V}\right) = \log d_A, \quad (1-f)Vs\left(\frac{E}{V}\right) = \log d_B. \tag{5.24}$$

Next, consider the $f > 1/2$. Unfortunately, the maximum of the integral occurs right near the transition $S_A(\mathcal{E}_1) = S_B(E - \mathcal{E}_1)$. Because of this, we cannot simply make a saddle point approximation. However, it is straightforward to argue that the PRRE will be large for $\alpha < 1$ and infinite for $\alpha \geq 1$. Note that the integrands of the numerator and denominator of $\mathrm{Tr}\left[\rho_A^\alpha \sigma_A^{1-\alpha}\right]$ are exponentially close for $S_A(\mathcal{E}_1) < S_B(E - \mathcal{E}_1)$ while the numerator is exponentially suppressed in relation to the denominator for $S_A(\mathcal{E}_1) > S_B(E - \mathcal{E}_1)$. Because the saddle point for the denominator occurs when $S_A(\mathcal{E}_1) > S_B(E - \mathcal{E}_1)$, $\mathrm{Tr}\left[\rho_A^\alpha \sigma_A^{1-\alpha}\right]$ will be exponentially small i.e. $\log\left[\mathrm{Tr}\left[\rho_A^\alpha \sigma_A^{1-\alpha}\right]\right]$ will be negative and of order the entropy. Due to the factor of $(\alpha - 1)^{-1}$ in the PRRE, this implies that the PRRE for $\alpha < 1$ is of order the entropy and ill-defined (infinite) for $\alpha \geq 1$, in analogy with the random matrix theory result (3.35).

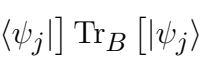
We have found that whenever A is less than half the total system size, the PRRE and therefore the subsystem trace distance between any two eigenstates of the same energy is exponentially suppressed in the entropy. The trace distance is a metric on the space of density matrices, so these eigenstates lie within a ball with radius $O(e^{-S(E)})$. The universal density matrix then must also lie within this ball such that (5.2) is satisfied.

For (5.3), we need to perform an additional computation. The off diagonal ($i \neq j$) matrix

for two random states is represented as

$$\mathrm{Tr}_B\left[|\psi_i\rangle\langle\psi_j|\right] := \text{  }. \tag{5.25}$$

To compute the trace norm, we need the integer powers of

$$\mathrm{Tr}_B\left[|\psi_i\rangle\langle\psi_j|\right]\mathrm{Tr}_B\left[|\psi_j\rangle\langle\psi_i|\right] = \text{  }. \tag{5.26}$$

The moments are given by a new sum over permutations

$$\overline{\mathrm{Tr}\left[\left(\mathrm{Tr}_B\left[|\psi_i\rangle\langle\psi_j|\right]\mathrm{Tr}_B\left[|\psi_j\rangle\langle\psi_i|\right]\right)^{\alpha}\right]} = \frac{1}{(d_A d_B)^{2\alpha}}\sum_{\tau\in S_{odd}} d_A^{C(\eta^{-1}\circ\tau)} d_B^{C(\tau)}. \tag{5.27}$$

Here, S_{odd} represents the set of permutations where, within each cycle, the difference between consecutive numbers is always odd. For example, the cycle $(1,2,5,6)$ is allowed, but $(2,4,5)$ is not. Crucially, the identity permutation is not an allowed permutation. If we want to maximize the number of dashed loops for small d_A/d_B, τ must be composed of α noncrossing two-cycles. The degeneracy is given by the Catalan number so that

$$\overline{\mathrm{Tr}\left[\left(\mathrm{Tr}_B\left[|\psi_i\rangle\langle\psi_j|\right]\mathrm{Tr}_B\left[|\psi_j\rangle\langle\psi_i|\right]\right)^{\alpha}\right]} = C_\alpha d_A^{1-\alpha} d_B^{-\alpha} + O\left(d_B^{-\alpha-1}\right). \tag{5.28}$$

The trace norm is the $\alpha \to 1/2$ limit, such that

$$\overline{\left|\mathrm{Tr}_B\left[|\psi_i\rangle\langle\psi_j|\right]\right|_1} = \frac{8}{3\pi}\sqrt{\frac{d_A}{d_B}} + O\left(\frac{d_A}{d_B}\right)^{3/2}. \tag{5.29}$$

When d_A/d_B is large, the cyclic permutation is an allowed permutation and will dominate, leading to

$$\overline{\mathrm{Tr}\left[\left(\mathrm{Tr}_B\left[|\psi_i\rangle\langle\psi_j|\right]\mathrm{Tr}_B\left[|\psi_j\rangle\langle\psi_i|\right]\right)^{\alpha}\right]} = d_B^{1-2\alpha} + O\left(d_A^{-1}\right). \tag{5.30}$$

Taking the $\alpha \to 1/2$ limit tells us that $\left|\mathrm{Tr}_B\left[|\psi_i\rangle\langle\psi_j|\right]\right|_1$ is exponentially close to one.

For finite energy eigenstates, this translates to

$$\overline{\mathrm{Tr}\left[\left(\mathrm{Tr}_B\left[|\psi_i\rangle\langle\psi_j|\right]\mathrm{Tr}_B\left[|\psi_j\rangle\langle\psi_i|\right]\right)^{\alpha}\right]} = \mathcal{N}^{-2\alpha}\int d\mathcal{E}e^{S_A(\mathcal{E})+S_B(E-\mathcal{E})}G_\alpha(\mathcal{E}), \tag{5.31}$$

where $G_\alpha(\mathcal{E})$ is now given by

$$G_\alpha(\mathcal{E}) = \begin{cases} C_\alpha e^{\alpha S_A(\mathcal{E})+(\alpha-1)S_B(E-\mathcal{E})}, & e^{S_A(\mathcal{E})-S_B(E-\mathcal{E})} \ll 1 \\ e^{(2\alpha-1)S_A(\mathcal{E})}, & e^{S_A(\mathcal{E})-S_B(E-\mathcal{E})} \gg 1 \end{cases}. \tag{5.32}$$

At $\alpha = 1/2$, for sufficiently small f, the saddle point will fall in the $e^{S_A(\mathcal{E})-S_B(E-\mathcal{E})} \ll 1$ regime leading to a saddle point equation for the numerator of

$$3s'\left(\frac{\mathcal{E}_1}{Vf}\right) = s'\left(\frac{E-\mathcal{E}_1}{V(1-f)}\right). \tag{5.33}$$

Due to the factor of 3, $\mathcal{E}_1$ will be larger than $\mathcal{E}_2$. We then find

$$\overline{\left|\mathrm{Tr}_B\left[|\psi_i\rangle\langle\psi_j|\right]\right|_1} \simeq \frac{8e^{\frac{3}{2}S_A(\mathcal{E}_1)+\frac{1}{2}S_B(E-\mathcal{E}_1)}}{3\pi e^{S_A(\mathcal{E}_2)+S_B(E-\mathcal{E}_2)}}. \tag{5.34}$$

This is exponentially small because $e^{S_A(\mathcal{E}_1)+S_B(E-\mathcal{E}_1)} \ll e^{S_A(\mathcal{E}_2)+S_B(E-\mathcal{E}_2)}$ and $S_A(\mathcal{E}_1) \ll S_B(E-\mathcal{E}_1)$. For sufficiently large f, the saddle point will fall in the other regime such that the numerator and denominator are identical at $\alpha = 1/2$, leading to $\overline{\left|\mathrm{Tr}_B\left[|\psi_i\rangle\langle\psi_j|\right]\right|_1} \simeq 1$ at leading order.

5.1.2 Holographic states

We could now simply posit that because holographic systems are believed to be chaotic, their eigenstates will also have a spatial decomposition according to (5.4) and thus, will obey the

subsystem ETH for $f < 1/2$. However, this line of reasoning is somewhat unsatisfying because it is not constructive. Instead, we implement a gravitational calculation using the fixed-area states, to evaluate the PREE in normal states without any areas fixed. Our strategy follows Ref. [52] in manipulating the gravitational path integral into a form identical to (5.7), deriving the validity of using (5.4) for holographic eigenstates when computing the PRRE. Due to the similarities with Ref. [52], we keep the derivation brief, referring the interested reader to the original literature.

To begin, we make the assumption that black hole microstates can be represented as a random superposition of energy eigenstates in a microcanonical energy window $I_E = [E - \Delta, E + \Delta]$

$$|E, \hat{c}\rangle \propto \sum_{E_i \in I_E} c^i e^{-\frac{\beta E_i}{2}} |E_i\rangle \,, \tag{5.35}$$

where β is an effective temperature. These states are believed to be holographically dual to black hole geometries with end-of-world branes specifying the microstate lying behind the horizon [75, 104, 14, 39, 135], similar to the PSSY model.

The corresponding density matrix is represented as the path integral on a strip of width β with boundary conditions determined by $\hat{c}$. We will consider two microstates in the same energy window, corresponding to two independent sets of Gaussian random variables $\hat{c}$ and $\hat{d}$. As argued in Ref. [52], after disorder averaging, the random variables match up the boundary conditions of the strips according to the same Wick contractions previously discussed for Haar random states. Therefore, the path integral is given by a sum over all allowed Wick contractions

$$\mathrm{Tr}\left[\rho_A^\alpha \sigma_A^m\right] = \frac{\mathcal{Z}_{\alpha,m}}{\mathcal{Z}_{1,0}^\alpha \mathcal{Z}_{0,1}^m}, \quad \mathcal{Z}_{\alpha,m} \simeq \sum_{\mathcal{M}_i} \int \mathcal{D}\phi e^{-I_E(\mathcal{M}_i,\phi)}, \tag{5.36}$$

where $\mathcal{M}_i$ are the replica manifolds. Using the holographic dictionary, this is a sum over bulk

geometries with asymptotic boundary conditions $\mathcal{M}_i$. In these bulk geometries, the EOW branes have disappeared due to the disorder averaging. This is an alternative way to see the reduction of bulk saddles from $S_{\alpha+m}$ to $S_\alpha \times S_m$. In general, solving the bulk equations of motion to evaluate the path integrals on shell is very difficult. We use the "double-defect" construction of Ref. [52] to separate the action into a bulk contribution $I_{bulk}(g, \phi, E)$ and actions, $I_{brane}(\Sigma_1)$ and $I_{brane}(\Sigma_2)$, for cosmic branes, Σ_1 and Σ_2, that are located at the two extremal surfaces

$$\mathcal{Z}_{\alpha,m} = \sum_{\mathcal{M}_i} \int \mathcal{D}\phi \mathcal{D}g \mathcal{D}\Sigma_1 \mathcal{D}\Sigma_2 e^{-(\alpha+m) I_{bulk}(g,\phi,E) - \frac{\alpha+m-k_i}{4G} I_{brane}(\Sigma_1) - \frac{k_i-1}{4G} I_{brane}(\Sigma_2)}, \quad (5.37)$$

where k_i plays the role of the number of cycles in the Wick contraction corresponding to $\mathcal{M}_i$. The sum over $\mathcal{M}_i$ may be done prior to the path integral and we also take $m \to 1-\alpha$ to arrive at

$$\mathcal{Z}_{\alpha,m} = \int \mathcal{D}\phi \mathcal{D}g \mathcal{D}\Sigma_1 \mathcal{D}\Sigma_2 e^{-I_{bulk}(g,\phi,E)} G_\alpha(\Sigma_1, \Sigma_2), \quad (5.38)$$

where

$$G_\alpha(\Sigma_1, \Sigma_2) = \begin{cases} {}_2F_1\left(1-\alpha, -\alpha; 2; e^{\frac{\Delta I_{brane}}{4G}}\right) {}_2F_1\left(\alpha-1, \alpha; 2; e^{\frac{\Delta I_{brane}}{4G}}\right), \\ \qquad\qquad \Delta I_{brane} < 0 \\ e^{-\frac{\Delta I_{brane}}{4G}} {}_2F_1\left(1-\alpha, -\alpha; 2; e^{-\frac{\Delta I_{brane}}{4G}}\right) {}_2F_1\left(\alpha-1, \alpha; 2; e^{-\frac{\Delta I_{brane}}{4G}}\right), \\ \qquad\qquad \Delta I_{brane} > 0 \end{cases}, \quad (5.39)$$

and $\Delta I_{brane} := I_{brane}(\Sigma_1) - I_{brane}(\Sigma_2)$. Next, we make use of the fixed-area basis by

postponing the integrals over the areas of the branes until the very end, such that the integral is rewritten as

$$\mathcal{Z}_{\alpha,1-\alpha} = \int dA_1 dA_2 P(A_1, A_2) G_\alpha(A_1, A_2), \tag{5.40}$$

where $G_\alpha(A_1, A_2)$ is identical to $G_\alpha(\Sigma_1, \Sigma_2)$ with $I_{brane}(\Sigma_1) \leftrightarrow A_1$ and $I_{brane}(\Sigma_2) \leftrightarrow A_2$ and $P(A_1, A_2)$ is the (unnormalized) probability of being in the state with areas A_1 and A_2

$$P(A_1, A_2) := \int \mathcal{D}g\mathcal{D}\phi\Big|_{A_{\Sigma_1}=A_1, A_{\Sigma_2}=A_2} e^{-I_{bulk}(g,\phi,E)}. \tag{5.41}$$

For high-energy eigenstates, $P(A_1, A_2)$ will localize to a trajectory where A_2 is a function of A_1 [52]

$$P(A_1, A_2) \simeq \delta_{A_2, A_2(A_1)} \int \mathcal{D}g\mathcal{D}\phi\Big|_{A_{\Sigma_1}=A_1} e^{-I_{bulk}(g,\phi,E)}. \tag{5.42}$$

Finally, to compare with (5.7), we want to change the integration variable from A_1 to the energy density in region A, $\mathcal{E}$. Using the Bekenstein-Hawking formula [76], we write the areas in terms of entropy densities

$$\frac{A_1}{4G} = fVs\left(\frac{\mathcal{E}(A_1, A_2)}{fV}\right) + A_\infty, \quad \frac{A_2}{4G} = (1-f)Vs\left(E - \frac{\mathcal{E}(A_1, A_2)}{(1-f)V}\right) + A_\infty, \tag{5.43}$$

where A_∞ is the divergent piece of the Ryu-Takayanagi surface which approximately cancels in all expressions because we are in the high-energy limit where the surfaces are approximately purely radial until they reach the horizon and subsequently tightly wrap the horizon. $\mathcal{E}(A_1, A_2)$ is the ADM energy which is a function of the horizon area $A_{BH} \simeq A_1 + A_2 - 2A_\infty$. In a saddle point approximation, the probability then becomes [52]

$$P(\mathcal{E}) \simeq e^{S_A(\mathcal{E}) + S_B(E-\mathcal{E})}. \tag{5.44}$$

In total, we find

$$\frac{\mathcal{Z}_{\alpha,1-\alpha}}{\mathcal{Z}_{1,0}^{\alpha}\mathcal{Z}_{0,1}^{1-\alpha}} \simeq \frac{\int d\mathcal{E} e^{S_A(\mathcal{E})+S_B(E-\mathcal{E})} G_\alpha(\mathcal{E})}{\int d\mathcal{E} e^{S_A(\mathcal{E})+S_B(E-\mathcal{E})}}. \tag{5.45}$$

It should now be evident that this formula is identical to (5.7), so we conclude that the PRRE for holographic theories is exponentially small in the entropy when $f < 1/2$ and subsystem eigenstate thermalization will hold.

5.1.3 Numerics

We provide a preliminary numerical study of relative entropy between mid-spectrum eigenstates of integrable and chaotic spin chains of length N with Hamiltonian

$$H = -\sum_{i=1}^{N} \left(Z_i Z_{i+1} + h_x X_i + h_z Z_i\right), \tag{5.46}$$

where X and Z are Pauli spin operators. We take $h_x = 1$, $h_z = 0$ for the integrable limit and $h_x = -1.05$, $h_z = 0.5$ for the chaotic regime as in Ref. [20]. We also numerically study the Sachdev-Ye-Kitaev model [103] with Hamiltonian

$$H = \sum_{j<k<l<m}^{N} J_{ijkl}\chi_j\chi_k\chi_l\chi m, \quad \overline{J_{ijkl}^2} = \frac{6}{(N-3)(N-2)(N-1)} J^2, \tag{5.47}$$

where the χ_i's are Majorana fermions and J_{ijkl} is a Gaussian random variable. The comparison between numerical data and (5.22) is shown in Fig. 5.1. The eigenstates are chosen randomly from the middle of the spectrum and so are effectively at infinite temperature. The SYK model matches very well. This may be expected because the Hamiltonian is a random matrix and the SYK model is known to be maximally chaotic. The chaotic spin chain eigenstates have relative entropy close to, but noticeably larger than, random mixed states. This is reasonable because these eigenstates are not truly random and therefore should be

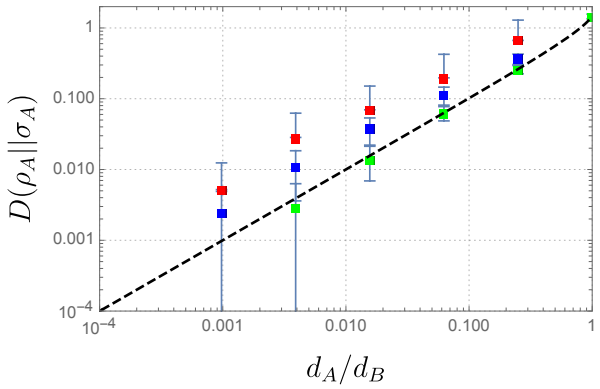


Figure 5.1: The relative entropy between 10^3 random pairs of mid-spectrum eigenstates. The blue (red) data points are for the chaotic (integrable) spin chain with 12 spins and the dashed line is (5.22). The green data points are for the SYK model with 20 Majorana fermions. We have omitted the lower error bars for the red data points for clarity, as they are very large and get in the way of the other data.

more easily distinguishable. We expect this to be a finite size effect that dissolves in the thermodynamic limit. Meanwhile, the integrable eigenstates are even more distinguishable, which is consistent with their violation of the eigenstate thermalization hypothesis. Moreover, the variance in relative entropy from eigenstate to eigenstate is much larger for the integrable spin chain.

5.2 Dynamics

We have, so far, discussed relative entropies for high-energy energy eigenstates. In nature, energy eigenstates in their pure form are overwhelmingly rare. The more natural form of thermalization is an approach to thermal equilibrium from nontrivial out-of-equilibrium dynamics. We now approach this using the equilibrium approximation.

5.2.1 Relative entropy

Consider a system evolved from two possible initial states specified respectively by density operators ρ_0, σ_0. We will assume that the support of ρ_0 lies inside that of σ_0 so that the

relative entropy

$$D(\rho_0||\sigma_0) = \mathrm{Tr}\rho_0 \log \rho_0 - \mathrm{Tr}\rho_0 \log \sigma_0 \tag{5.48}$$

is finite. Suppose $\rho = U\rho_0 U^\dagger$ and $\sigma = U\sigma_0 U^\dagger$ can be approximated at the macroscopic level by the same equilibrium density operator $\rho^{(\mathrm{eq})}$. We are interested in calculating the relative entropy

$$\begin{aligned} D(\rho_A||\sigma_A) &= \mathrm{Tr}_A \rho_A \log \rho_A - \mathrm{Tr}_A \rho_A \log \sigma_A & (5.49)\\ &= \lim_{n\to 1} \frac{1}{n-1} \left(\log \mathrm{Tr}_A \left[\rho_A^n\right] - \log \mathrm{Tr}_A \left[\rho_A \sigma_A^{n-1}\right]\right) & (5.50) \end{aligned}$$

between the reduced density operators ρ_A, σ_A of some subsystem A. The calculation of the first (entropy) term is reviewed in Appendix A, see e.g (A.28). Here we discuss how to use the equilibrium approximation to compute the second term, which depends on two distinct density matrices.

We rewrite the second term in (5.50) as a transition amplitude in the replica Hilbert space

$$\mathcal{D}_{n,A} = \mathrm{Tr}_A \left[\rho_A \sigma_A^{n-1}\right] = \langle \eta_A \otimes e_{\bar{A}}| \left(U \otimes U^\dagger\right)^{\otimes n} |\rho_0 \otimes \sigma_0^{\otimes(n-1)}, e\rangle \,, \tag{5.51}$$

where the state $|\rho \otimes \sigma^{\otimes(n-1)}, \tau\rangle$ for a permutation τ is defined as

$$\left\langle i_1 \bar{i}'_1 i_2 \bar{i}'_2 \cdots i_n \bar{i}'_n | \rho \otimes \sigma^{\otimes(n-1)}, \tau \right\rangle = \rho_{i_1 i'_{\tau(1)}} \sigma_{i_2 i'_{\tau(2)}} \cdots \sigma_{i_n i'_{\tau(n)}} \tag{5.52}$$

which is inhomogeneous in the replicas. Applying the equilibrium approximation, we find

$$\begin{aligned}
\mathcal{D}_{n,A} &\approx \frac{1}{Z_2^n}\sum_{\tau\in\mathcal{S}_n}\langle\eta_A\otimes e_{\bar{A}}|\,\mathcal{I}_\alpha,\tau\rangle\langle\mathcal{I}_\alpha,\tau\,|\rho_0\otimes\sigma_0^{\otimes(n-1)},e\rangle && (5.53)\\
&= \frac{1}{Z_2^n}\sum_{\tau\in\mathcal{S}_n}\langle\eta_A\otimes e_{\bar{A}}|\,\mathcal{I}_\alpha,\tau\rangle \\
&\qquad\times\mathrm{Tr}\left[\mathcal{I}\rho_0\left(\mathcal{I}_\alpha\sigma_0\right)^{m_1-1}\right]\mathrm{Tr}\left[\left(\mathcal{I}_\alpha\sigma_0\right)^{m_2}\right]\cdots\mathrm{Tr}\left[\left(\mathcal{I}\sigma_0\right)^{m_{k(\tau)}}\right] && (5.54)\\
&= \frac{1}{Z_1^n}\sum_{\tau\in\mathcal{S}_n}\langle\eta_A\otimes e_{\bar{A}}|\,\mathcal{I}_\alpha,\tau\rangle\frac{\mathrm{Tr}(\rho_0\sigma_0^{m_1-1})}{\mathrm{Tr}(\sigma_0^{m_1})}\prod_{i=1}^{k(\tau)}\mathrm{Tr}(\sigma_0^{m_i}) && (5.55)
\end{aligned}$$

where m_i is the number of elements in i-th cycle of τ, the $i=1$ cycle is taken to be the one containing the first copy of the Hilbert space, and we have used self-consistency conditions (derived similarly as (A.27))

$$\mathrm{Tr}\left[\mathcal{I}\rho_0\left(\mathcal{I}\sigma_0\right)^{n-1}\right]\approx\frac{Z_2^n}{Z_1^n}\mathrm{Tr}(\rho_0\sigma_0^{n-1}),\quad \mathrm{Tr}\left[\left(\mathcal{I}\sigma_0\right)^{n}\right]\approx\frac{Z_2^n}{Z_1^n}\mathrm{Tr}(\sigma_0^{n})\ . \tag{5.56}$$

We now make some general comments on the structure of (5.55). In (5.55), we can divide τ's into those with $m_1=1$, and those with $m_1>1$. Denoting the two sets respectively as $\mathcal{S}_{n,1}$ and $\mathcal{S}_{n,2}$, we have

$$\mathcal{D}_{n,A}=\sum_{\tau\in\mathcal{S}_{n,1}}\mathcal{Z}_{n,A}(\tau)+\sum_{\tau\in\mathcal{S}_{n,2}}\frac{\mathrm{Tr}(\rho_0\sigma_0^{m_1-1})}{\mathrm{Tr}(\sigma_0^{m_1})}\mathcal{Z}_{n,A}(\tau) \tag{5.57}$$

where

$$\mathcal{Z}_{n,A}(\tau)=\frac{1}{Z_1^n}\langle\eta_A\otimes e_{\bar{A}}|\,\mathcal{I}_\alpha,\tau\rangle\prod_{i=1}^{k(\tau)}\mathrm{Tr}(\sigma_0^{m_i}) \tag{5.58}$$

is the contribution of τ to the Renyi partition function for A with initial state σ_0.

From the discussion around (A.32), we then conclude that when A is small $\tau=e$ domi-

nates,[2] giving

$$\mathcal{D}_{n,A} = \mathcal{Z}^{(\text{eq})}_{n,A}, \quad \text{for } S^{(\text{eq})}_{n,A} \ll S^{(\text{eq})}_{n,\bar{A}} + S_n(\sigma_0) \ . \tag{5.59}$$

For A to be sufficiently large we expect the first term is dominated by η_{n-1} which cyclicly permutes $2, \cdots n$ and gives a contribution $\mathcal{Z}_{n,A}(\eta_{n-1})$, while the dominant permutation for the second term is η, giving a contribution $\text{Tr}(\rho_0\sigma_0^{n-1})\mathcal{Z}^{(\text{eq})}_{n,\bar{A}}$. From our general discussion $\mathcal{Z}_{n,A}(\eta_{n-1})$ is smaller than $\mathcal{Z}^{(\text{eq})}_{n,\bar{A}}$ by at least a factor Z_1^{-1}. Thus for the case $\text{Tr}(\rho_0\sigma_0^{n-1})$ is not too small (i.e. much larger than Z_1^{-1}) we then have

$$\mathcal{D}_{n,A} \approx \text{Tr}(\rho_0\sigma_0^{n-1})\mathcal{Z}^{(\text{eq})}_{n,\bar{A}}, \quad \text{for } S^{(\text{eq})}_{n,A} \gg S^{(\text{eq})}_{n,\bar{A}} + S_n(\sigma_0) \ . \tag{5.60}$$

Now combining the above discussion with (A.32), and assuming that we can analytically continue to $n = 1$, we find (5.50) can be written as

$$D(\rho_A||\sigma_A) \simeq \begin{cases} 0, & S^{(\text{eq})}_A \ll S^{(\text{eq})}_{\bar{A}} + S(\rho_0) \\ S^{(\text{eq})}_A - S^{(\text{eq})}_{\bar{A}} - S(\rho_0), & S^{(\text{eq})}_{\bar{A}} + S(\rho_0) \ll S^{(\text{eq})}_A \ll S^{(\text{eq})}_{\bar{A}} + S(\sigma_0) \\ D(\rho_0||\sigma_0), & S^{(\text{eq})}_A \gg S^{(\text{eq})}_{\bar{A}} + S(\sigma_0) \end{cases} . \tag{5.61}$$

where we have assumed $S(\sigma_0) \gg S(\rho_0)$. The above expressions are intuitively reasonable. When subregion A is sufficiently small, the two density matrices are entirely indistinguishable, a manifestation of thermalization in an isolated quantum system. Once we move beyond the first regime of (5.61), the state σ_0 becomes important. In particular, the relative entropy rises from 0 to $D(\rho_0||\sigma_0)$ where it plateaus; as we gain information, the density matrices become more and more distinguishable. The relative entropy never increases beyond $D(\rho_0||\sigma_0)$ due to the monotonicity of relative entropy under quantum channels.

Less conservatively, if we trust the analytic continuations of dominant permutations, we

2. Note $\text{Tr}(\rho_0\sigma_0^{m-1}) < \text{Tr}(\sigma_0^m)$.

find the following sharper version of (5.61):

$$D(\rho_A||\sigma_A) \simeq \begin{cases} 0, & S_A^{(\mathrm{eq})} < S_{\bar{A}}^{(\mathrm{eq})} + S(\rho_0) \\ S_A^{(\mathrm{eq})} - S_{\bar{A}}^{(\mathrm{eq})} - S(\rho_0), & S_{\bar{A}}^{(\mathrm{eq})} + S(\rho_0) < S_A^{(\mathrm{eq})} < S_{\bar{A}}^{(\mathrm{eq})} + S(\rho_0) + D(\rho_0||\sigma_0) \\ D(\rho_0||\sigma_0), & S_A^{(\mathrm{eq})} > S_{\bar{A}}^{(\mathrm{eq})} + S(\rho_0) + D(\rho_0||\sigma_0) \end{cases} . \tag{5.62}$$

We also found numerical evidence for this equation in small spin chains and it is consistent with an infinite temperature result derived exactly in Section 4.4. It would be interesting to test this equation more generally.

CHAPTER 6

HOPES FOR THE (NEAR) FUTURE

"I got a head full of ideas that are drivin' me insane."

— Maggie's Farm

To wrap up, I discuss a few current lines of research that occupy my mind.

6.1 Pseudorandom Eigenstate Thermalization

As previously discussed, there are various forms of the eigenstate thermalization hypothesis (ETH), starting with the work of Deutsch and Srednicki [43, 179]. We will refer to this early version of the ETH as the local ETH (lETH) because it can be expressed in terms of expectation values of local operators.

Proposition 1. *A Hamiltonian obeys the local eigenstate thermalization hypothesis if for all "simple," few-body operators and energy eigenstates*

$$\langle E_i | \mathcal{O}_\alpha | E_j \rangle = f_{\mathcal{O}_\alpha}(E)\delta_{ij} + R_{ij}e^{-S(E)/2}, \quad E := \frac{E_i + E_j}{2}, \tag{6.1}$$

where f is a smooth function of the energy (and other conserved quantities), S is the thermodynamic entropy, and R is an $O(1)$ pseudorandom matrix.

One can imagine that there is more to thermalization than expectation values of local operators. Stronger versions of eigenstate thermalization were later formulated for finite subsystems, A. In these formulations, the full density matrix of the subregion is "close" to a universal, thermal density matrix, ρ_{univ} [69, 53]. Following Ref. [53], we define subsystem ETH (sETH) as a statement about the trace distance between reduced density matrices.

Proposition 2. *A Hamiltonian obeys the subsystem eigenstate thermalization hypothesis for subsystem fractions $f_A \leq f_A^*$ if for sufficiently high energy eigenstates*

$$T(\rho_{A,i}, \rho_{univ}(E)) := \frac{1}{2} \left| \rho_{A,i} - \rho_{univ}(E) \right|_1 = O\left(e^{-S(E)/2}\right), \quad \rho_{A,i} := Tr_B \left[|E_i\rangle \langle E_i| \right], \quad (6.2)$$

where B is the spatial complement of A.

The trace distance is a distinguishability measure and characterizes the optimal error probability for distinguishing two states using a quantum measurement $\{M, \mathbb{1} - M\}$[85, 88]

$$\min_{0 \leq M \leq \mathbb{1}} \left[\mathrm{Tr}\left[(\mathbb{1} - M)\rho \right] + \mathrm{Tr}\left[M\sigma \right] \right] = 1 - T(\rho, \sigma). \quad (6.3)$$

There is a constructive formula for the explicit optimal measurement where $M = \Pi_+ + \Pi_0$. Π_+ is the projection onto the positive part of $\rho - \sigma$ and Π_0 is the projection onto the zero eigenspace. This gives an operational meaning to the subsystem ETH; states obeying the subsystem ETH are almost impossible to distinguish from thermal states with only an exponentially small improvement of identification over a completely random guess ($M \propto \mathbb{1}$).

The subsystem ETH necessarily breaks down when A is larger than half the total system size. A simple way to see this is by invoking the Fannes–Audenaert inequality [59, 17]

$$|S_{vN}(\rho) - S_{vN}(\sigma)| \leq 2T \log(d_A) - 2T \log 2T, \quad (6.4)$$

where d_A is the dimension of the subsystem Hilbert space and S_{vN} is the von Neumann entropy. The von Neumann entropy of A is equal to the von Neumann entropy of B in the eigenstate because it is pure. In contrast, the thermal entropy is always proportional to the volume of A ($O(\log(d_A))$). The difference is therefore extensive, leading to $O(1)$ trace distance. Therefore, if one has access to more than half the system, one can perform the Holevo-Helstrom measurement to determine if the state is an eigenstate or the thermal state.

From the above analysis, it seems that states lose all notions of thermality when one has access to larger than half the system. However, it seems like there should still be global notions of thermality for high energy eigenstates and states at late times following equilibration. The only loophole in the above argument is if an observer was not allowed to make the Holevo-Helstrom measurement. We will restrict a class of measurements including that of Holevo-Helstrom using the notion of complexity, only allowing sufficiently simple measurement. This brings us to *quantum pseudorandomness*.

Following Ref. [102], we say that a quantum state $|\Psi\rangle_{AB}$ is pseudorandom on subsystem A if there exists an $\alpha > 0$ such that

$$\left|\Pr(\mathcal{M}(\rho_A)) - \Pr\left(\mathcal{M}\left(\frac{\mathbb{1}}{d_A}\right)\right)\right| \leq d_B^{-\alpha}, \tag{6.5}$$

for any two-outcome measurement, $\mathcal{M}$, of polynomial complexity in $\log(d_B)$.

The Hypothesis For finite energy states, one only has access to a subset of Hilbert space and we posit a modified definition of quantum pseudorandomness at finite energy

$$|\Pr(\mathcal{M}(\rho_A)) - \Pr\left(\mathcal{M}\left(\rho_{univ}\right)\right)| \leq e^{-\alpha(1-f_A)S(E)}, \tag{6.6}$$

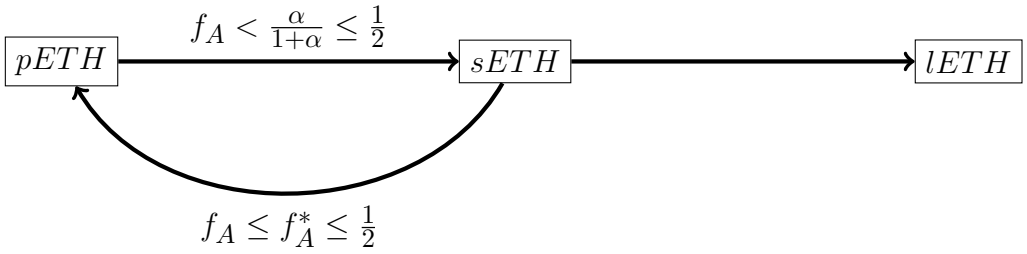
where f_A is the fraction of the system in subregion A and the measurements are restricted to those with finite energy.

We propose that the ETH can be further generalized to a *pseudorandom eigenstate thermalization hypothesis*, holding for much larger subsystems than the subsystem ETH.

Proposition 3. *A Hamiltonian obeys the pseudorandom eigenstate thermalization hypothesis for subsystems of fraction f_A if all energy eigenstates are pseudorandom on A*

$$\left|Pr(\mathcal{M}(\rho_{A,i})) - Pr\left(\mathcal{M}\left(\rho_{univ}\right)\right)\right| \leq e^{-\alpha(1-f_A)S(E)}. \tag{6.7}$$

Relations to local and subsystem ETH From its definition, it is not immediately clear how the pseudorandom ETH is related to the local and subsystem versions, if at all. We now clarify the relationships. We summarize the connections in the following diagram



.

pETH implies sETH Consider strings of Pauli operators, P_i, of which there are 4^N of, forming a complete basis. In the basis where each Pauli operator has eigenvalues ± 1, we can consider this a two-outcome measurement, $\mathcal{M}_i$, such that

$$\mathrm{Tr}\left[(\rho - \sigma)P_i\right] = 2\left(\Pr(\mathcal{M}_i(\rho)) - \Pr(\mathcal{M}_i(\sigma))\right). \tag{6.8}$$

We can evaluate the two-norm from the completeness and orthogonality of the Pauli operators

$$\begin{aligned} |\rho_A - \sigma_A|_2^2 &:= \mathrm{Tr}\left[(\rho_A - \sigma_A)^2\right] \\ &= 2^{-N_A} \sum_i |\mathrm{Tr}\left[(\rho_A - \sigma_A)P_i\right]|^2 \\ &= 2^{2-N_A} \sum_i |\Pr(\mathcal{M}_i(\rho)) - \Pr(\mathcal{M}_i(\sigma))|^2. \end{aligned} \tag{6.9}$$

We can now use the assumption of pETH to bound the two-norm

$$|\rho_A - \sigma_A|_2^2 \leq 2^{N_A+2} e^{-2\alpha(1-f_A)S(E)}. \tag{6.10}$$

The one-norm is bounded by the two norm as

$$|\rho_A - \sigma_A|_1 \leq \sqrt{\text{rank}(\rho_A - \sigma_A)}|\rho_A - \sigma_A|_2 \leq 2^{N_A+1} e^{-\alpha(1-f_A)S(E)}, \tag{6.11}$$

so the trace distance is bounded as

$$T(\rho_A, \sigma_A) \leq 2^{N_A} e^{-\alpha(1-f_A)S(E)}. \tag{6.12}$$

Therefore, the trace distance will be exponentially small and subsystem ETH will hold for subsystems $N_A \log 2 < \alpha(1 - f_A)S(E)$. The high energy limit of this result is $N_A < \alpha N_B$. This means that the pETH implies the sETH for $f_A < \frac{\alpha}{1+\alpha} \leq \frac{1}{2}$. α must therefore be no larger than one. As we will soon see, it may be possible for α to be larger than one when $f_A > \frac{1}{2}$, but not for $f_A < \frac{1}{2}$.

sETH implies pETH An alternative way to define the trace distance is through an optimal measurement

$$T(\rho, \sigma) := \max_{0 \leq \Lambda \leq \mathbb{1}} \left|\text{Tr}\left[\Lambda(\rho - \sigma)\right]\right|. \tag{6.13}$$

Therefore, if we restrict to polynomially complex measurements, Λ_P, we find an inequality

$$T(\rho, \sigma) \geq \max_{0 \leq \Lambda_P \leq \mathbb{1}} \left|\text{Tr}\left[\Lambda_P(\rho - \sigma)\right]\right|. \tag{6.14}$$

This immediately implies that whenever $f_A \leq f_A^*$, the pseudorandom ETH must hold. Because of the results of the prior subsection, we conclude that the sETH and pETH are equivalent in a finite range.

sETH implies lETH This was proven in Refs. [53]. The smooth ETH function is given by

$$f_{\mathcal{O}_\alpha}(E) = \mathrm{Tr}\left[\rho_{\mathrm{univ}}(E)\mathcal{O}_\alpha\right]. \tag{6.15}$$

The Cauchy-Schwartz inequality implies

$$\mathrm{Tr}\left[(\rho_{\mathrm{univ}}(E) - \rho_{A,i})\mathcal{O}_\alpha\right] \leq \sqrt{4\mathrm{Tr}\left[(\rho_{\mathrm{univ}}(E) + \rho_{A,i})\mathcal{O}_\alpha^2\right] T(\rho_{\mathrm{univ}}(E), \rho_{A,i})}. \tag{6.16}$$

Assuming the sETH, the difference in the expectation value from the smooth ETH function is exponentially small.

pETH for typical states Consider a Haar random state $|\Psi\rangle$ which should be pseudorandom on any subspace. Here, we characterize the extent of pseudorandomness by evaluating expectation values of Pauli operators supported on A, comparing to the infinite temperature state

$$\left|\mathrm{Tr}\left[\left(|\Psi\rangle\langle\Psi| - \frac{\mathbb{1}}{2^N}\right) P_i\right]\right| \leq 2^{1-\alpha N_B}. \tag{6.17}$$

Nontrivial Pauli operators are traceless, so the above condition reduces to

$$|\langle\Psi| P_i |\Psi\rangle| \leq 2^{1-\alpha N_B}. \tag{6.18}$$

The ensemble average of this correlation function is zero, but we need to understand the statistics. The variance is

$$\begin{aligned}\overline{\langle\Psi|P_i|\Psi\rangle^2} &= \overline{\mathrm{Tr}\left[|\Psi\rangle\langle\Psi|^{\otimes 2} P_i^{\otimes 2}\right]} \\ &= \mathrm{Tr}\left[\frac{\sum_{\tau\in S_2} g_\tau}{\sum_{\tau\in S_2}\mathrm{Tr}\left[g_\tau\right]} P_i^{\otimes 2}\right] \\ &= \frac{\mathrm{Tr}\left[P_i\right]^2 + \mathrm{Tr}\left[P_i^2\right]}{2^{2N}+2^N} \\ &= \frac{2^N}{2^{2N}+2^N}.\end{aligned} \tag{6.19}$$

The higher cumulants are further suppressed in total Hilbert space dimension, so the distribution is asymptotically normal with variance 2^{-N}.

For fun, let's compute the higher moments

$$\begin{aligned}\overline{\langle\Psi|P_i|\Psi\rangle^n} &= \mathrm{Tr}\left[\frac{\sum_{\tau\in S_n} g_\tau}{\sum_{\tau\in S_n}\mathrm{Tr}\left[g_\tau\right]} P_i^{\otimes n}\right] \\ &\simeq \left(\frac{n}{2}\right)!! 2^{-nN/2}\end{aligned} \tag{6.20}$$

This function is only nontrivial for even n. The permutations in the numerator that dominate are those composed of $n/2$ 2-cycles, hence the double factorial.

With probability $4^{-N_A}\epsilon$, $|\langle\Psi|P_i|\Psi\rangle|$ will be greater than X, where X is

$$2\int_X^\infty dx \frac{2^{N/2}}{\sqrt{2\pi}} e^{-\frac{1}{2}\left(2^{N/2}x\right)^2} = 4^{-N_A}\epsilon. \tag{6.21}$$

Therefore

$$X = 2^{\frac{1}{2}-\frac{N}{2}}\mathrm{erfc}^{-1}\left(4^{-N_A}\epsilon\right) \to C 2^{-\frac{N}{2}}\sqrt{N_A}, \tag{6.22}$$

where C is a constant. We can then use the union bound to state that with high probability,

this bounds all Pauli expectation values. This places an upper bound on α of $\frac{1}{2}$ because $2^{-\frac{1}{2}(1-f_A)N} \geq 2^{-\frac{N}{2}}$ for all f_A.

It is straightforward to show that typical states obey subsystem ETH for $f_A < \frac{1}{2}$. A quick way to show this is by starting with Page's formula (for $f_A < \frac{1}{2}$)

$$S_{vN}(A) = \log(d_A) - \frac{d_A}{2d_B}. \tag{6.23}$$

The quantum relative entropy between a state and the maximally mixed state can be expressed in terms of the entropy

$$D\left(\rho_A || \frac{\mathbb{1}}{d_A}\right) = \log(d_A) - S_{vN}(A) = \frac{d_A}{2d_B}. \tag{6.24}$$

The quantum Pinsker inequality upper bounds the trace distance by the square root of the relative entropy, so the trace distance is exponentially small when $f_A < \frac{1}{2}$

$$T(\rho, \rho_{\text{univ}}) \leq \sqrt{\frac{d_A}{d_B}}. \tag{6.25}$$

Therefore, sETH implies pseudorandomness for $f_A < \frac{1}{2}$ with a lower bound of $\alpha \geq \frac{1}{2}$. Therefore, we find that $\alpha = \frac{1}{2}$ and the Pauli measurements are near optimal.

Late-time states Consider starting in the standard basis state $|0\rangle^{\otimes(N_A+N_B)}$ and applying unitary $U(t)$

$$|\Psi(t)\rangle = U(t)\,|0\rangle^{\otimes(N_A+N_B)}\,. \tag{6.26}$$

Next, we trace over N_B and wish to determine if we have a pseudorandom state. The natural protocol for determining this is by tensoring the resultant state with the maximally mixed state on B, applying inverse time evolution, then projecting onto the basis state i.e. the

adjoint channel. This is only a polynomially complex measurement because tensoring with the identity matrix is simple and the unitary evolution is assumed to to be for exponential times. The majority of thermalization occurs in linear time. The probability of this two-outcome measurment is

$$\begin{aligned}\Pr(\mathcal{M}(\rho_A(t))) &= \langle 0|^{\otimes(N_A+N_B)} U^\dagger(t) \mathrm{Tr}_B \left[U(t) |0\rangle^{\otimes(N_A+N_B)} \langle 0|^{\otimes(N_A+N_B)} U^\dagger(t) \right] \otimes \frac{\mathbb{1}_B}{d_B} U(t) |0\rangle^{\otimes(N_A+N_B)} \\ &= \mathrm{Tr}\left[|\Psi(t)\rangle \langle \Psi(t)| \mathrm{Tr}_B \left[|\Psi(t)\rangle \langle \Psi(t)| \right] \otimes \frac{\mathbb{1}_B}{d_B} \right] \\ &= \frac{1}{d_B} \mathrm{Tr}\left[\mathrm{Tr}_B \left[|\Psi(t)\rangle \langle \Psi(t)| \right]^2 \right]. \end{aligned} \tag{6.27}$$

This is simply the purity of A. The purity is approximately $d_A^{-1} + d_B^{-1}$, so

$$\Pr(\mathcal{M}(\rho_A(t))) \simeq \frac{1}{d_A d_B} + \frac{1}{d_B^2}. \tag{6.28}$$

Applying the same measurement protocol to the maximally mixed state on A (the universal density matrix at infinite temperature), we find

$$\Pr\left(\mathcal{M}\left(\frac{\mathbb{1}}{d_A} \right) \right) = \langle 0|^{\otimes(N_A+N_B)} U^\dagger(t) \left(\frac{\mathbb{1}_A}{d_A} \otimes \frac{\mathbb{1}_B}{d_B} \right) U(t) |0\rangle^{\otimes(N_A+N_B)} = \frac{1}{d_A d_B}. \tag{6.29}$$

We expect this to be close to the optimal measurement when $d_A > d_B$, implying that the equilibrated pure state at infinite temperature is pseudorandom with $\alpha = 2$ for $d_A > d_B$. More rigorously, this is an upper bound on α.

Note that the second moment of the unitary was used. Therefore, if $U(t)$ is a 2-design, $|\Psi(t)\rangle$ will appear fully Haar random at this level of measurement. Approximate 2-designs can be made from local Hamiltonian circuits in polynomial time. The key difference between this case and the Haar random case is that the backwards evolution by $U(t)$ is not polynomial complex for truly Haar random states.

Finite temperature At finite temperature, we want to check distinguishability from universal density matrices, not the maximally mixed state. As a measurement protocol, it is again natural to choose the adjoint channel. However, an even better choice for measurement is tensoring with the universal density matrix on B instead of the maximally mixed state.

$$\Pr(\mathcal{M}(\rho_A(t))) = \mathrm{Tr}\left[|\Psi(t)\rangle\langle\Psi(t)|\,\mathrm{Tr}_B\left[|\Psi(t)\rangle\langle\Psi(t)|\right]\otimes \mathrm{Tr}_A\left[\rho_{\mathrm{univ}}\right]\right]. \tag{6.30}$$

Here,

$$\mathrm{Tr}_A\left[\rho_{\mathrm{univ}}\right] = \frac{\mathcal{I}_B}{Z_{1,B}}. \tag{6.31}$$

Before evaluating this for equilibrated pure states, we take a slight detour into a certain perspective on the formalism.

For Haar random states, ensemble averages of copies involve sums over the permutation group

$$\overline{|\Psi\rangle\langle\Psi|^{\otimes\alpha}} = \frac{\sum_{\tau\in S_\alpha} g_\tau}{\sum_{\tau\in S_\alpha}\mathrm{Tr}\left[g_\tau\right]}, \tag{6.32}$$

where g_τ is the matrix representation of permutation τ. For example, if $|\Psi\rangle$ lives in a two-dimensional Hilbert space and $\alpha = 2$, we have

$$g_{\mathbb{1}} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}, \quad g_{SWAP} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}. \tag{6.33}$$

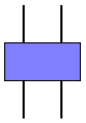
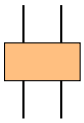
The equilibrium approximation requires no averaging, yet we still have a sum over permu-

tations. Appended to the permutation is an equilibrium density matrix

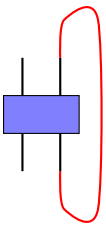
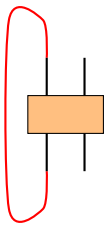
$$|\Psi\rangle\langle\Psi|^{\otimes\alpha} \simeq \frac{\sum_{\tau\in S_\alpha} g_\tau \mathcal{I}_\alpha^{\otimes\alpha}}{\sum_{\tau\in S_\alpha} \mathrm{Tr}\left[g_\tau \mathcal{I}_\alpha^{\otimes\alpha}\right]} \simeq \frac{\sum_{\tau\in S_\alpha} \left(g_{\tau,A}\mathcal{I}_A^{\otimes\alpha}\right)\left(g_{\tau,B}\mathcal{I}_B^{\otimes\alpha}\right)}{Z_{1,A}^\alpha Z_{1,B}^\alpha}, \tag{6.34}$$

where we will only take the identity permutation in the denominator due to its dominance in the thermodynamic limit.

Let us represent the relevant operators diagrammatically

$$\rho := |\Psi\rangle\langle\Psi| = \frac{1}{Z_{1,A}Z_{1,B}}$$  , $\mathcal{I} =$  . (6.35)

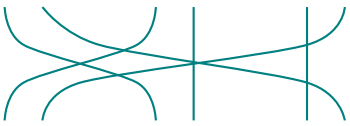
We can take partial traces as

$$\rho_A = \mathrm{Tr}_B[\rho] = \frac{1}{Z_{1,A}Z_{1,B}}$$  , $\mathcal{I}_B = \mathrm{Tr}_A[\mathcal{I}] =$  (6.36)

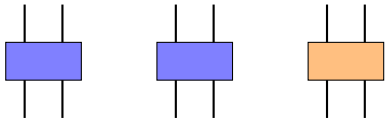
The measurement probability can then be considered as the expectation value of a certain permutation, τ_3, in this the tensor product state

$$\mathrm{Tr}\left[|\Psi(t)\rangle\langle\Psi(t)|\,\mathrm{Tr}_B\left[|\Psi(t)\rangle\langle\Psi(t)|\right]\otimes\mathrm{Tr}_A\left[\rho_{\mathrm{univ}}\right]\right] = \frac{1}{Z_{1,A}Z_{1,B}}\mathrm{Tr}\left[(\rho\otimes\rho\otimes\mathcal{I})\tau_3\right]. \tag{6.37}$$

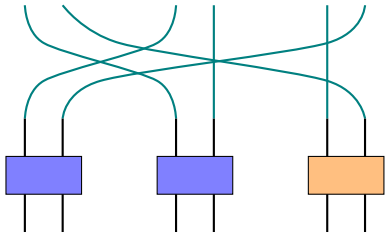
The relevant permutation is a swap between the first and second copies on A and a swap between the first and third copies on B

$\tau_3 =$  . (6.38)

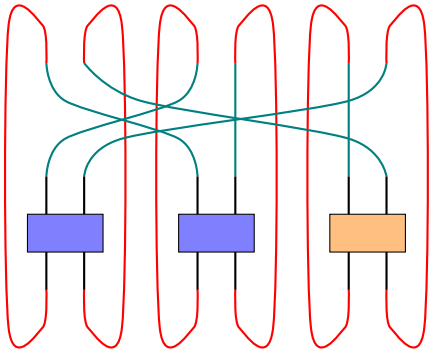
We then need the tensor product

$$\rho \otimes \rho \otimes \mathcal{I} = \frac{1}{(Z_{1,A} Z_{1,B})^2}$$  . (6.39)

Multiplying these, we have

$$(\rho \otimes \rho \otimes \mathcal{I})\tau_3 = \frac{1}{(Z_{1,A} Z_{1,B})^2}$$  . (6.40)

Taking the trace, we have

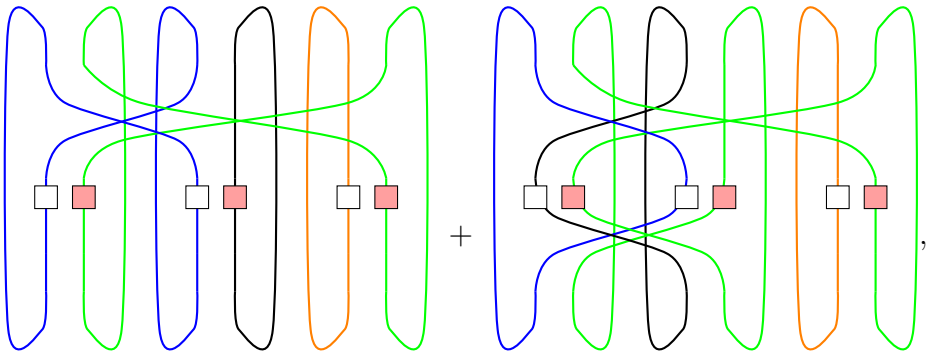
$$\mathrm{Tr}\left[(\rho \otimes \rho \otimes \mathcal{I})\tau_3\right] = \frac{1}{(Z_{1,A} Z_{1,B})^2}$$  . (6.41)

To evaluate this diagram, we recall that the two factors of ρ will lead to a sum over the

identity and sway permutations on these sites

$$\mathrm{Tr}\left[(\rho\otimes\rho\otimes\mathcal{I})\tau_3\right] = \tag{6.42}$$

$$\frac{1}{(Z_{1,A}Z_{1,B})^2}$$



where the white boxes are $\mathcal{I}_A$'s and the pink boxes are $\mathcal{I}_B$'s, leading to

$$\begin{aligned}\mathrm{Tr}\left[|\Psi(t)\rangle\langle\Psi(t)|\,\mathrm{Tr}_B\left[|\Psi(t)\rangle\langle\Psi(t)|\right]\otimes\mathrm{Tr}_A\left[\rho_{\mathrm{univ}}\right]\right] &= \frac{1}{Z_{1,B}}\left(\frac{Z_{2,A}Z_{1,B}Z_{2,B}+Z_{1,A}^2Z_{3,B}}{(Z_{1,A}Z_{1,B})^2}\right)\\ &= e^{-(S_2(A)+S_2(B))}+e^{-2S_3(B)}.\end{aligned} \tag{6.43}$$

Meanwhile, applying the same protocol to the equilibrium density matrix, we have only a single term

$$\Pr\left(\mathcal{M}\left(\rho_{univ}\right)\right) = \mathrm{Tr}\left[|\Psi(t)\rangle\langle\Psi(t)|\,\frac{\mathcal{I}_A}{Z_{1,A}}\otimes\frac{\mathcal{I}_B}{Z_{1,B}}\right] \tag{6.44}$$

$$\simeq \frac{Z_{2,A}Z_{2,B}}{(Z_{1,A}Z_{1,B})^2} = e^{-(S_2(A)+S_2(B))}. \tag{6.45}$$

The difference is $e^{-2S_3(B)}$, supporting the proposal for pseudorandomness at finite energy. It would be interesting to check these predictions numerically. It would also be interesting to apply the ideas to energy eigenstates rather than late-time states.

6.2 Distinguishing evaporating black holes

There was a clear shortcoming in the analysis of distinguishability in Hawking radiation in Chapter 4 due to the simplicity model. Namely, everything in the PSSY model is Euclidean. There is no real time evolution as there is in true black hole evaporation, only a tuning of parameters describing the number of degrees of freedom behind the horizon. While it is reasonable to believe this model is still sufficient for the qualitative mechanisms of resolving the black hole information problem, it should really be considered a simple example of Wheeler's bag of gold puzzle. We now discuss two models where genuine Lorentzian time evolution in the black hole evaporation process may be studied.

The first model consists of a SYK quantum dot coupled to a free fermionic bath. This model is purely quantum mechanical and therefore manifestly unitary. Nevertheless, it is useful due to SYK's relation to Jackiw-Teitelboim gravity at low energies.

As a model of a pure black hole microstate, I consider the class of states put forward by Kourkoulou and Maldacena (KM) defined by the conditions [104]

$$(\chi^{2k-1} - is_k\chi^{2k})\,|B_s\rangle = 0, \tag{6.46}$$

where the χ's are the fundamental Majorana fermions and the s_k's are signs. Each choice of $N/2$ signs defines a basis state in the Hilbert space. These states are reminiscient of conformal boundary states in two-dimensional conformal field theory. In a similar fashion, we smear these these states using Euclidean evolution to achieve an overcomplete basis of low energy states at effective temperature β

$$|KM_s\rangle = e^{-\beta H_{SYK}/2}\,|B_s\rangle\,. \tag{6.47}$$

These pure states appear thermal when subject to sufficiently simple probes. Importantly, they are not eigenstates of the Hamiltonian and will have non-trivial time evolution. They

may be qualitatively described by pure state black holes in AdS_2 with end-of-world branes located behind the horizon, specifying the particular set of s_k's. By evolving these states in real time, we can allow these black hole states to evaporate. Moreover, we may couple the SYK system to a non-gravitational bath that collects the radiation coming from the black hole. One tractable approach is to have this bath consist of an infinite chain of N free fermions, ψ, with Hamiltonian

$$H_\psi = i \sum_{x,i} \frac{\Lambda}{2} \psi_i(x) \psi_i(x+1). \tag{6.48}$$

The coupling between the SYK system and free fermion bath can be taken as a simple hopping at the origin of the bath as in Ref. [33]

$$H_{int} = \sum_i iV\sqrt{\Lambda} \chi_i \psi_i(0). \tag{6.49}$$

The advantage of this coupled set-up, including the KM states, is that it has a large-N limit that can be studied efficiently using a path integral. We can then evaluate distinguishability measures as a function of real time in the bath degrees of freedom with the two states corresponding to two different initial KM states for the SYK system. This should be tractable at all times by numerically solving the Schwinger-Dyson equations and may even be analytically tractable at very early and very late times using approximations. This should elucidate the universality of our preliminary conclusions based on the PSSY model and may lead to the study of new, interesting Lorentzian saddles in gravity. Taking a step back from gravitational motivations, this coupled system has many variations that may be interesting to study from a many-body physics and thermalization perspective.

6.3 Experimental Signatures

It is an exciting time to be working on quantum informational aspects of many-body physics and gravity, not only because of the tremendous theoretical progress, but the rapid development of experiments focused on quantum computation and quantum simulation. It is now possible for experimentalists to engineer many-body Hamiltonians in e.g. cold atom arrays and observe non-trivial dynamics in these systems using high-precision measurements, opening up a world of opportunities.

Of particular relevance to this thesis, are experimental measurements of quantum information theoretic quantities. There are various methods for conducting these measurements. One direction is full quantum state tomography, where repeated preparation and measurement of a quantum state allows one to reconstruct the entire density matrix, after which, any quantity can be directly computed. The issue with this approach is that it is highly inefficient and is not scalable to the many-body systems we are ultimately interested in. In particular, we hope these experiments are able to be more powerful than exact simulations that one can perform on a classical computer. Other approaches involve replicas. In theory (and sometimes in practice), one can prepare the same state twice in parallel, leading to the global state on the doubled Hilbert space $\mathcal{H}_1 \otimes \mathcal{H}_2$

$$|\Psi\rangle = |\psi\rangle_1 \otimes |\psi\rangle_2 \,. \tag{6.50}$$

One can then measure the operator that swaps the two Hilbert spaces on subregion A and is the identity on region B, leading to the purity

$$\mathrm{Tr}\left[\rho_A^2\right] = \langle\Psi| \, SWAP_A \otimes \mathbb{1}_B \, |\Psi\rangle \,. \tag{6.51}$$

While this is more efficient than full quantum state tomography, it introduces new challenges such as the larger Hilbert space and the necessity of preparing two copies of the system at

a time.

A novel approach was later devised using randomized measurements that is both efficient and avoids the above challenges [28]. This is the technique we will explore presently. We first choose a simple computational basis for N qudits to make measurements in. These should be product states of single-site computational bases $|s_i\rangle$ where i runs from 1 to d. A random product unitary is applied on region A prior to measurement

$$U_A = \bigotimes_{j=1}^{N_A} u^{(j)}, \tag{6.52}$$

where the $u^{(j)}$'s are $d \times d$ unitary matrices independent drawn from the circular unitary ensemble (CUE) i.e. Haar measure on $U(d)$. The probability of measuring basis state $|\mathbf{s}\rangle_A$ on A is $P(\mathbf{s}_A) = \mathrm{Tr}_A \left[U_A \rho_A U_A^\dagger |\mathbf{s}_A\rangle \langle \mathbf{s}_A| \right]$. Using the properties of random averages of $U(d)$, one can show that the purity of ρ_A is expressible using the classical probability distributions

$$\mathrm{Tr}\left[\rho_A^2\right] = d^{N_A} \sum_{\mathbf{s}_A \tilde{\mathbf{s}}_A} (-d)^{D[\mathbf{s}_A, \tilde{\mathbf{s}}_A]} \, \overline{P(\mathbf{s}_A) P(\tilde{\mathbf{s}}_A)}, \tag{6.53}$$

where D is the Hamming distance between the strings and the over line represents an ensemble averaging over the unitary group. It is crucial that $\overline{P(\mathbf{s}_A) P(\tilde{\mathbf{s}}_A)}$ is not equivalent to $\overline{P(\mathbf{s}_A)}\ \overline{P(\tilde{\mathbf{s}}_A)}$.

We would like to apply the same approach to distinguishability measures. In particular, we would like to understand subsystem thermalization, the relaxation of subsystems to universal thermal ensembles. The simplest distinguishability measure to compute using randomized measurements is the Hilbert-Schmidt fidelity [121]

$$F_{HS}(\rho_A, \sigma_A) = \frac{\mathrm{Tr}\left[\rho_A \sigma_A\right]}{\min\left[\mathrm{Tr}\left[\rho_A^2\right], \mathrm{Tr}\left[\sigma_A^2\right]\right]}. \tag{6.54}$$

This fidelity satisfies Jozsa's axioms, though, to our knowledge, does not (yet) have an

explicit operation meaning in terms of hypothesis testing. Nevertheless, we will use it as a proxy for the fidelity and hope to explore its operational meaning in quantum hypothesis testing.

We have already explained how to determine the two terms in the denominator using randomized measurements. The new term in the numerator is quite similar. Denoting the probability for state σ_A as $Q(\mathbf{s}_A)$, an analogous derivation leads to

$$\mathrm{Tr}\left[\rho_A \sigma_A\right] = d^{N_A} \sum_{\mathbf{s}_A \tilde{\mathbf{s}}_A} (-d)^{D[\mathbf{s}_A, \tilde{\mathbf{s}}_A]} \overline{P(\mathbf{s}_A) Q(\tilde{\mathbf{s}}_A)}. \tag{6.55}$$

Indeed, this was investigated in Ref. [55].

It is important to evaluate F_{HS} between quenched states of equivalent energy but orthogonal initial states as a function of time and subsystem size. For thermalizing systems, we expect F_{HS} to be close to one for subsystems less than half the total system size and close to zero for subsystems larger than half the total system size. A further direction is understanding higher moments with the hope of efficiently measuring quantities with explicit operational interpretations, such as the relative entropy. This is likely most tractable for large d.

In conclusion, there are many exciting directions in many-body and gravitational physics leveraging notions of distinguishability. I am optimistic that meaningful progress can be made.

APPENDIX A

EQUILIBRIUM APPROXIMATION

In this appendix, we review the equilibrium approximation introduced in [126]. We first discuss the formulation for a pure state, and then for a mixed state.

A.1 Pure states

We consider a system evolving from a far-from-equilibrium pure state $\rho_0 = |\Psi_0\rangle\langle\Psi_0|$ to a state $\rho = |\Psi\rangle\langle\Psi|$ with $|\Psi\rangle = U|\Psi_0\rangle$, which is in equilibrium at a macroscopic level. We assume that macroscopic physical properties of equilibrated pure state ρ can be approximated by an equilibrium density operator

$$\rho^{(\mathrm{eq})} = \frac{\mathcal{I}_\alpha}{Z(\alpha)}, \quad Z(\alpha) = \mathrm{Tr}\,\mathcal{I}_\alpha, \tag{A.1}$$

where α collectively denotes macroscopic parameters for the equilibrium state.

Consider the n-th Renyi entropy of the equilibrated pure state with respect to a subsystem A

$$\mathcal{Z}_{n,A} = e^{-(n-1)S_{n,A}} = \mathrm{Tr}_A \rho_A^n = \mathrm{Tr}_A \left(\mathrm{Tr}_{\bar{A}} U\rho_0 U^\dagger\right)^n = \langle \eta_A \otimes e_{\bar{A}}|(U\otimes U^\dagger)^n|\rho_0, e\rangle\,, \tag{A.2}$$

where in the last equality we have written it as an amplitude in the replica space $(\mathcal{H}\otimes\mathcal{H})^n$, with various notations defined as follows. For any operator $\mathcal{O}$ acting on $\mathcal{H}$, the state $|\mathcal{O},\sigma\rangle \in (\mathcal{H}\otimes\mathcal{H})^n$, where σ is an element of the permutation group $\mathcal{S}_n$ of n objects, is defined as

$$\left\langle i_1\bar{i}'_1 i_2\bar{i}'_2\cdots i_n\bar{i}'_n|\mathcal{O},\sigma\right\rangle = \mathcal{O}_{i_1 i'_{\sigma(1)}}\mathcal{O}_{i_2 i'_{\sigma(2)}}\cdots\mathcal{O}_{i_n i'_{\sigma(n)}}, \quad \mathcal{O}_{ij} = \langle i|\mathcal{O}|j\rangle\ . \tag{A.3}$$

Here $\{|i_1\bar{i}'_1 i_2\bar{i}'_2\cdots i_n\bar{i}'_n\rangle\}$ is a basis for $(\mathcal{H}\otimes\mathcal{H})^n$ and $\sigma(i)$ denotes the image of i under

σ. When $\mathcal{O}$ is given by the identity operator, we will denote the states obtained this way simply as $|\sigma\rangle$. When the system is divided into subsystems, we can similarly define states by associating different permutations to different subsystems. For example, suppose $\mathcal{H} = \mathcal{H}_A \otimes \mathcal{H}_{\bar{A}}$, $|\mathcal{O}, \tau_A \otimes \sigma_{\bar{A}}\rangle$ with $\tau, \sigma \in \mathcal{S}_n$ is defined as

$$\left\langle i_{1_a} i_{1_b} \bar{i}'_{1_a} \bar{i}'_{1_b} \cdots i_{n_a} i_{n_b} \bar{i}'_{n_a} \bar{i}'_{n_b} | \mathcal{O}, \tau_A \otimes \sigma_{\bar{A}} \right\rangle = \mathcal{O}_{i_{1_a} i_{1_b}, i'_{\tau(1)a} i'_{\sigma(1)_b}} \cdots \mathcal{O}_{i_{na} i_{n_b}, i'_{\tau(n)a} i'_{\sigma(n)_b}} \tag{A.4}$$

where $|i_{k_a}\rangle, |\bar{i}'_{k_a}\rangle$, $|i_{k_b}\rangle, |\bar{i}'_{k_b}\rangle$ denote respectively basis vectors for subsystem A and $\bar{A}$ in the k-th replica of $\mathcal{H} \otimes \mathcal{H}$. In (A.2), $|\eta_A \otimes e_{\bar{A}}\rangle$ is a state associated with the identity operator, with e representing the identity permutation and η the cyclic permutation $(n, n-1, \cdots 1)$.

We can decompose the identity on the replica Hilbert space as

$$\mathbb{1} = P_\alpha + Q, \quad P_\alpha Q = Q P_\alpha = 0, \quad Q^2 = Q, \tag{A.5}$$

where P_α is the projector

$$P_\alpha = \frac{1}{Z_2^n} \sum_{\sigma,\tau} g^{\sigma\tau} |\mathcal{I}_\alpha, \sigma\rangle \langle \mathcal{I}_\alpha, \tau|, \quad g_{\tau\sigma} = \frac{\langle \mathcal{I}_\alpha, \tau | \mathcal{I}_\alpha, \sigma \rangle}{\sqrt{\langle \mathcal{I}_\alpha, \tau | \mathcal{I}_\alpha, \tau \rangle \langle \mathcal{I}_\alpha, \sigma | \mathcal{I}_\alpha, \sigma \rangle}}, \quad Z_n := \mathrm{Tr} \mathcal{I}_\alpha^n \ . \tag{A.6}$$

We will be interested in systems with a large number of degrees of freedom, i.e. $Z_1 \gg 1$. For such systems, inserting the identity twice in the last expression of (A.2) and neglecting the term involving Q, we find $\mathcal{Z}_n^{(A)}$ can be approximated as

$$\mathcal{Z}_{n,A} \approx [\mathcal{Z}_{n,A}]_{\text{eq approx}} := \frac{1}{Z_2^n} \sum_{\sigma,\tau} g^{\tau\sigma} \langle \eta_A \otimes e_{\bar{A}} | \mathcal{I}_\alpha, \tau \rangle \langle \mathcal{I}_\alpha, \sigma | \rho_0, e \rangle \tag{A.7}$$

$$= \frac{1}{Z_1^n} \sum_{\sigma,\tau} g^{\tau\sigma} \langle \eta_A \otimes e_{\bar{A}} | \mathcal{I}_\alpha, \tau \rangle \tag{A.8}$$

$$\approx \frac{1}{Z_1^n} \sum_{\tau \in \mathcal{S}_n} \langle \eta_A \otimes e_{\bar{A}} | \mathcal{I}_\alpha, \tau \rangle \ . \tag{A.9}$$

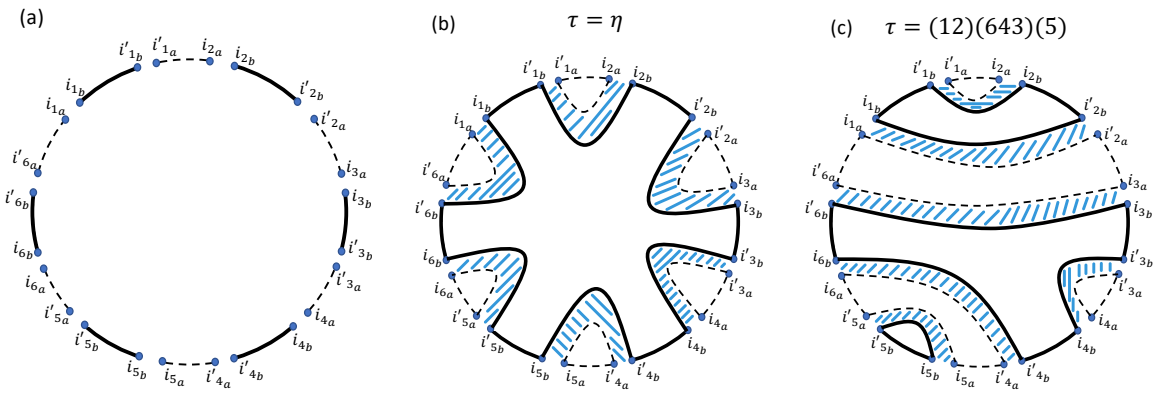


Figure A.1: (a) shows the common exterior lines of all diagrams for different terms in (A.9), and (b) and (c) show examples of diagrams for two choices of τ, for the case $n = 6$.

In going from (A.7) to (A.8), we use the fact that for (A.7) to be compatible with $\mathrm{Tr}\rho = 1$, $\mathcal{I}_\alpha$ should satisfy a consistency requirement

$$\mathrm{Tr}(\mathcal{I}_\alpha \rho_0) = \frac{Z_2}{Z_1} \ . \tag{A.10}$$

In going from (A.8) to (A.9), we use the fact that $g^{\tau\sigma}$ is approximately equal to the identity when Z_1 is large.

$\mathcal{Z}_{n,A}$, as given in (A.9), only depends on the equilibrium density operator $\mathcal{I}_\alpha$, but satisfies the unitarity constraint

$$\mathcal{Z}_{n,A} = \mathcal{Z}_{n,\bar{A}} \ . \tag{A.11}$$

The size of the terms we neglected in reaching (A.9) can be estimated from Δ, defined by

$$\Delta^2 := [(\mathcal{Z}_{n,A})^2]_{\text{eq approx}} - \left([\mathcal{Z}_{n,A}]_{\text{eq approx}}\right)^2 \ . \tag{A.12}$$

It was shown in Appendix B of [126] that

$$\frac{\Delta}{[\mathcal{Z}_{n,A}]_{\text{eq approx}}} \sim Z_1^{-1/2} \ll 1 \ . \tag{A.13}$$

Each term in the final expression in (A.9) can be given a diagrammatic representation, as shown in Fig. A.1. We can insert the identity to write

$$\begin{aligned} \left\langle \eta_A \otimes e_{\bar{A}} | \mathcal{I}_\alpha, \tau \right\rangle &= \sum_{i_1, i'_1, \ldots i_n, i'_n} \langle \eta_A \otimes e_{\bar{A}} | i_1 \bar{i'}_1 \ldots i_n \bar{i'}_n \rangle \left\langle i_1 \bar{i'}_1 \ldots i_n \bar{i'}_n | \mathcal{I}_\alpha, \tau \right\rangle , \\ |i_m\rangle = |i_{m_a}\rangle_A \, |i_{m_b}\rangle_{\bar{A}} , &\quad |\bar{i'}_m\rangle = |\bar{i'}_{m_a}\rangle_A \, |\bar{i'}_{m_b}\rangle_{\bar{A}} . \end{aligned} \tag{A.14}$$

The exterior of the diagram, which is the same for all τ, represents $\langle \eta_A \otimes e_{\bar{A}} | i_1 \bar{i'}_1 \ldots i_n \bar{i'}_n \rangle$ by connecting i_{m_a} with $i'_{\eta(m)_a}$ using dashed lines, and i_{m_b} with i'_{m_b} using solid lines, as shown in Fig. A.1(a). The interior of the diagram represents $\langle i_1 \bar{i'}_1 \ldots i_n \bar{i'}_n | \mathcal{I}_\alpha, \tau \rangle$, by connecting i_m with $i'_{\tau(m)}$, as shown for two examples in Fig. A.1(b) and (c). Roughly, each solid loop in the resulting diagram gives a power of $d_{\bar{A}}$ and each dashed loop gives a power of d_A, where d_A and $d_{\bar{A}}$ are respectively the effective Hilbert space dimensions of A and $\bar{A}$.[1] The number of solid and dashed loops in a diagram is respectively equal to $C(\tau)$ and $C(\eta^{-1} \circ \tau)$. We therefore find

$$\langle \eta_A \otimes e_{\bar{A}} | \mathcal{I}_\alpha, \tau \rangle \sim d_A^{C(\eta^{-1} \circ \tau)} d_{\bar{A}}^{C(\tau)} . \tag{A.15}$$

For any permutation τ, we have the inequality

$$C(\tau) + C(\eta^{-1} \circ \tau) \leq n + 1, \tag{A.16}$$

and the permutations for which this inequality is saturated are the ones associated with planar diagrams such as Fig. A.1(b).

From (A.15), when A is much smaller than $\bar{A}$, i.e. $d_A \ll d_{\bar{A}}$, the permutation $\tau = e$,

1. When $\mathcal{I}_\alpha$ can be factorized between A and $\bar{A}$, i.e. $\mathcal{I}_\alpha = \mathcal{I}_A \otimes \mathcal{I}_{\bar{A}}$, we can define the effective dimensions as $d_A = \mathrm{Tr}_A \mathcal{I}_A$ and $d_{\bar{A}} = \mathrm{Tr}_{\bar{A}} \mathcal{I}_{\bar{A}}$. When $\mathcal{I}_\alpha$ cannot be factorized, they can be estimated by counting the numbers of degrees of freedom of the subsystems.

which maximizes $k(\tau)$, dominates, and we have

$$\mathcal{Z}_{n,A} = \mathcal{Z}^{(\mathrm{eq})}_{n,A}, \tag{A.17}$$

where $\mathcal{Z}^{(\mathrm{eq})}_{n,A}$ is the Renyi partition function for A in the state $\rho^{(\mathrm{eq})}$. When A is much larger than $\bar{A}$, i.e. $d_A \gg d_{\bar{A}}$, $\tau = \eta$, which maximizes $C(\eta^{-1} \circ \tau)$, dominates

$$\mathcal{Z}_{n,A} = \mathcal{Z}^{(\mathrm{eq})}_{n,\bar{A}} \ . \tag{A.18}$$

Except for a crossover region around $d_A \sim d_{\bar{A}}$ where the behavior may be more complicated, we then have

$$S_{n,A} = \min\left(S^{(\mathrm{eq})}_{n,A}, S^{(\mathrm{eq})}_{n,\bar{A}}\right), \qquad n \geq 2 \tag{A.19}$$

where $S^{(\mathrm{eq})}_{n,A}$ denotes the n-th Renyi entropy for subsystem A in the equilibrium density operator $\rho^{(\mathrm{eq})}$.

When S_A can be obtained from $S_{n,A}$ by analytic continuation to $n = 1$, (A.19) implies

$$S_A = \min(S^{(\mathrm{eq})}_A, S^{(\mathrm{eq})}_{\bar{A}}), \tag{A.20}$$

where $S^{(\mathrm{eq})}_A$ is the entanglement entropy for subsystem A in $\rho^{(\mathrm{eq})}$. In cases where the system $A\bar{A}$ is inhomogeneous, in general (A.20) cannot be deduced by analytic continuation.

When $\mathcal{I}_\alpha$ can be factorized

$$\mathcal{I}_\alpha \approx \mathcal{I}_A \otimes \mathcal{I}_{\bar{A}}, \tag{A.21}$$

equations (A.9) can be written more explicitly in terms of partition functions of A and $\bar{A}$

$$Z_{m,A} = \mathrm{Tr}_A \mathcal{I}^m_A, \qquad Z_{m,\bar{A}} = \mathrm{Tr}_{\bar{A}} \mathcal{I}^m_{\bar{A}}, \tag{A.22}$$

as

$$\mathcal{Z}_{n,A} \approx \frac{1}{Z_1^n} \sum_{\tau} \left(Z_{m_1,A} \cdots Z_{m_l,A}\right) \left(Z_{n_1,\bar{A}} \cdots Z_{n_k,\bar{A}}\right) \ , \tag{A.23}$$

where k is the number of cycles of τ with $n_1, \cdots n_k$ the lengths of the corresponding cycles, and l is the number of cycles of $\tau\eta^{-1}$ with $m_1, \cdots m_l$ the lengths of the corresponding cycles.

A.2 Mixed states

The equilibrium approximation can also be applied to a system $A = A_1 \cup A_2$ in a mixed state ρ_A that is in macroscopic equilibrium, but is far from the thermal density operators [126].

Suppose the system starts with a far-from equilibrium mixed state $\rho_{0,A}$, which evolves under unitary evolution operator U_A to $\rho_A = U_A \rho_{0,A} U_A^\dagger$. All the moments of $\rho_{0,A}$ are also preserved by the time evolution

$$z_{n,A} = \mathrm{Tr}\rho_{0,A}^n = \mathrm{Tr}\rho_A^n = \mathrm{Tr}(U_A \rho_{0,A} U_A^\dagger)^n = \Big\langle \eta | (U_A \otimes U_A^\dagger)^n | \rho_{0,A}, e \Big\rangle, \quad n = 2, \cdots \ . \tag{A.24}$$

The statement that $\rho_{0,A}$ is far-from-equilibrium is imposed by requiring that the $n-$th Renyi entropy of $\rho_{0,A}$ is smaller than the equilibrium entropy of A. For example, this condition is satisfied if we have

$$z_{n,A} \sim Z_A^{-(n-1)f} \ , \quad 0 \leq f < 1 \, . \tag{A.25}$$

Assuming that ρ_A can be approximated by an equilibrium density operator $\rho_A^{(\mathrm{eq})} = \frac{1}{Z_A} \mathcal{I}_A$ and applying the equilibrium approximation by inserting the projector (A.6) in (A.24) and ignoring terms with $\sigma \neq \tau$, we have

$$z_{n,A} \approx \frac{1}{Z_{2,A}^n} \sum_{\tau} \langle \eta | \mathcal{I}_A, \tau \rangle \langle \mathcal{I}_A, \tau | \rho_{0,A}, e \rangle, \tag{A.26}$$

which can be further simplified to the following constraints on $\mathcal{I}_A$ under the out-of-equilibrium

assumption (A.25)

$$\mathrm{Tr}(\mathcal{I}_A\rho_{0,A})^n \approx z_{n,A}\frac{Z^n_{2,A}}{Z^n_A}, \quad Z_{n,A} = \mathrm{Tr}_A\mathcal{I}^n_A, \quad Z_A = Z_{1,A}\ . \tag{A.27}$$

The Renyi partition functions for A_1 can then be approximated as

$$\mathcal{Z}_{n,A_1} \approx [\mathcal{Z}_{n,A_1}]_{\text{eq approx}} = \frac{1}{Z^n_A}\sum_\tau \langle \eta_{A_1}\otimes e_{A_2}|\mathcal{I}_A,\tau\rangle \prod_{i=1}^{k(\tau)} \mathrm{Tr}_A\rho^{n_i}_{0,A}, \quad n=2,3,\cdots \tag{A.28}$$

where $k(\tau)$ is the number of cycles of τ with $n_1,\cdots n_{k(\tau)}$ the lengths of the cycles.

The above discussion can be further generalized by embedding A in a larger system $S = A\cup B$, with the total system S in an initial pure state $|\Psi_0\rangle$ evolved to $|\Psi\rangle = U|\Psi_0\rangle$ in macroscopic equilibrium.[2] Suppose $|\Psi\rangle$ can be approximated macroscopically by $\rho^{(\text{eq})} = \frac{1}{Z_1}\mathcal{I}$. We then have

$$\mathcal{Z}_{n,A_1} \approx \frac{1}{Z^n_1}\sum_\tau \langle \eta_{A_1}\otimes e_{A_2B}|\mathcal{I},\tau\rangle\ . \tag{A.29}$$

This generalizes (A.28) as under the evolution of U for the full system S, the evolution from the initial density operator $\rho_{0,A}$ to ρ_A is in general not unitary. To recover (A.28) we take $U = U_A\otimes U_B$ to be factorized between A and B, in which case the equilibrium density operator $\mathcal{I} = \mathcal{I}_A\otimes\mathcal{I}_B$ should also factorize, and (A.29) can be written as

$$\mathcal{Z}_{n,A_1} = \frac{1}{Z^n_A}\sum_\tau \langle \eta_{A_1}\otimes e_{A_2}|\mathcal{I}_A,\tau\rangle \prod_{i=1}^{k(\tau)} \hat{Z}_{n_i,B}, \quad \hat{Z}_{n,B} := \frac{1}{Z^n_B}\mathrm{Tr}_B\mathcal{I}^n_B\ . \tag{A.30}$$

Equations (A.28) and (A.30) are equal provided that we choose the initial state $\rho_{0,B}$ such that

$$\mathrm{Tr}_B\rho^m_{0,B} = \frac{1}{Z^m_B}\mathrm{Tr}_B\mathcal{I}^m_B\,. \tag{A.31}$$

Then since the initial state is pure, $z_{n,A}$ is also given by (A.31). The requirement that $\rho_{0,A}$

2. A and B in principle do not have to be in equilibrium with each other.

is out-of-equilibrium is then equivalent to the requirement that $Z_B \gg Z_A$.

The relation between (A.29) and (A.28) also gives a way to estimate which permutation dominates in (A.28). From (A.29), when A_1 is smaller (larger) than A_2B, the dominant contribution is $\tau = e$ ($\tau = \eta$). Translating these statements to the notation of (A.28), we conclude that

$$S_{n,A_1} = \begin{cases} S^{(\mathrm{eq})}_{n,A_1} & S^{(\mathrm{eq})}_{n,A_1} < S^{(\mathrm{eq})}_{n,A_2} + S_{n,A} \\ S^{(\mathrm{eq})}_{n,A_2} + S_{n,A} & S^{(\mathrm{eq})}_{n,A_1} > S^{(\mathrm{eq})}_{n,A_2} + S_{n,A} \end{cases} . \tag{A.32}$$

In applying (A.29) to explicit calculations, we will need to make assumptions regarding B in the equilibrium density operator $\rho^{(\mathrm{eq})}$ for the full system, which may be considered as specifying different universality classes for ρ_A.

REFERENCES

[1] Ibrahim Akal. Universality, intertwiners and black hole information. *arXiv e-prints*, page arXiv:2010.12565, October 2020.

[2] Chris Akers, Thomas Faulkner, Simon Lin, and Pratik Rath. Reflected entropy in random tensor networks. *arXiv e-prints*, page arXiv:2112.09122, December 2021.

[3] Chris Akers, Thomas Faulkner, Simon Lin, and Pratik Rath. The Page Curve for Reflected Entropy. *arXiv e-prints*, page arXiv:2201.11730, January 2022.

[4] Chris Akers and Geoff Penington. Leading order corrections to the quantum extremal surface prescription. *Journal of High Energy Physics*, 2021(4):62, April 2021.

[5] Chris Akers and Geoff Penington. Quantum minimal surfaces from quantum error correction. *arXiv e-prints*, page arXiv:2109.14618, September 2021.

[6] Chris Akers and Pratik Rath. Holographic Renyi entropy from quantum error correction. *Journal of High Energy Physics*, 2019(5):52, May 2019.

[7] Chris Akers and Pratik Rath. Entanglement wedge cross sections require tripartite entanglement. *Journal of High Energy Physics*, 2020(4):208, April 2020.

[8] Ahmed Almheiri, Xi Dong, and Daniel Harlow. Bulk locality and quantum error correction in AdS/CFT. *Journal of High Energy Physics*, 2015:163, April 2015.

[9] Ahmed Almheiri, Xi Dong, and Brian Swingle. Linearity of holographic entanglement entropy. *Journal of High Energy Physics*, 2017(2):74, February 2017.

[10] Ahmed Almheiri, Netta Engelhardt, Donald Marolf, and Henry Maxfield. The entropy of bulk quantum fields and the entanglement wedge of an evaporating black hole. *Journal of High Energy Physics*, 2019(12):63, December 2019.

[11] Ahmed Almheiri, Thomas Hartman, Juan Maldacena, Edgar Shaghoulian, and Amirhossein Tajdini. Replica wormholes and the entropy of Hawking radiation. *Journal of High Energy Physics*, 2020(5):13, May 2020.

[12] Ahmed Almheiri, Thomas Hartman, Juan Maldacena, Edgar Shaghoulian, and Amirhossein Tajdini. The entropy of Hawking radiation. *arXiv e-prints*, page arXiv:2006.06872, June 2020.

[13] Ahmed Almheiri, Raghu Mahajan, Juan Maldacena, and Ying Zhao. The Page curve of Hawking radiation from semiclassical geometry. *Journal of High Energy Physics*, 2020(3):149, March 2020.

[14] Ahmed Almheiri, Alexandros Mousatov, and Milind Shyani. Escaping the Interiors of Pure Boundary-State Black Holes. *arXiv e-prints*, page arXiv:1803.04434, March 2018.

[15] Drew Armstrong. Generalized Noncrossing Partitions and Combinatorics of Coxeter Groups. *arXiv Mathematics e-prints*, page math/0611106, November 2006.

[16] K. M. R. Audenaert, J. Calsamiglia, R. Muñoz-Tapia, E. Bagan, Ll. Masanes, A. Acin, and F. Verstraete. Discriminating States: The Quantum Chernoff Bound. *Physical Review Letters*, 98(16):160501, April 2007.

[17] Koenraad M. R. Audenaert. A sharp continuity estimate for the von Neumann entropy. *Journal of Physics A Mathematical General*, 40(28):8127–8136, July 2007.

[18] Koenraad M. R. Audenaert and Nilanjana Datta. α-z-Rényi relative entropies. *Journal of Mathematical Physics*, 56(2):022202, February 2015.

[19] Koenraad M. R. Audenaert and Milán Mosonyi. Upper bounds on the error probabilities and asymptotic error exponents in quantum multiple state discrimination. *Journal of Mathematical Physics*, 55(10):102201, October 2014.

[20] M. C. Bañuls, J. I. Cirac, and M. B. Hastings. Strong and Weak Thermalization of Infinite Nonintegrable Quantum Systems. *Physical Review Letters*, 106(5):050405, February 2011.

[21] Ning Bao and Hirosi Ooguri. Distinguishability of black hole microstates. *Physical Review D*, 96(6):066017, September 2017.

[22] Jacob D. Bekenstein. Black holes and entropy. *Phys. Rev. D*, 7:2333–2346, 1973.

[23] Bruno Bertini and Lorenzo Piroli. Scrambling in random unitary circuits: Exact results. *Physical Review B*, 102(6):064305, August 2020.

[24] Igor Bjelaković, Jean-Dominique Deuschel, Tyll Krüger, Ruedi Seiler, Rainer Siegmund-Schultze, and Arleta Szkoła. A Quantum Version of Sanov's Theorem. *Communications in Mathematical Physics*, 260(3):659–671, December 2005.

[25] David D. Blanco, Horacio Casini, Ling-Yan Hung, and Robert C. Myers. Relative entropy and holography. *Journal of High Energy Physics*, 2013:60, August 2013.

[26] R.P. Boas. *Entire Functions*. Number v. 5 in Entire Functions. Academic Press, 1954.

[27] E. Brézin and A. Zee. Universal relation between Green functions in random matrix theory. *Nuclear Physics B*, 453(3):531–551, February 1995.

[28] Tiff Brydges, Andreas Elben, Petar Jurcevic, Benoît Vermersch, Christine Maier, Ben P. Lanyon, Peter Zoller, Rainer Blatt, and Christian F. Roos. Probing Rényi entanglement entropy via randomized measurements. *Science*, 364(6437):260–263, April 2019.

[29] Pasquale Calabrese and John Cardy. Entanglement entropy and quantum field theory. *Journal of Statistical Mechanics: Theory and Experiment*, 2004(6):06002, June 2004.

[30] Pasquale Calabrese and John Cardy. Evolution of entanglement entropy in one-dimensional systems. *Journal of Statistical Mechanics: Theory and Experiment*, 2005(4):04010, April 2005.

[31] Chi-Fang Chen, Geoffrey Penington, and Grant Salton. Entanglement wedge reconstruction using the Petz map. *Journal of High Energy Physics*, 2020(1):168, January 2020.

[32] Yiming Chen and Henry W. Lin. Signatures of global symmetry violation in relative entropies and replica wormholes. *arXiv e-prints*, page arXiv:2011.06005, November 2020.

[33] Yiming Chen, Xiao-Liang Qi, and Pengfei Zhang. Replica wormhole and information retrieval in the SYK model coupled to Majorana chains. *Journal of High Energy Physics*, 2020(6):121, June 2020.

[34] Benoit Collins. Moments and Cumulants of Polynomial random variables on unitary groups, the Itzykson-Zuber integral and free probability. *arXiv e-prints*, pages math–ph/0205010, May 2002.

[35] Benoît Collins and Ion Nechita. Gaussianization and eigenvalue statistics for random quantum channels (III). *arXiv e-prints*, page arXiv:0910.1768, October 2009.

[36] Benoît Collins and Ion Nechita. Random matrix techniques in quantum information theory. *Journal of Mathematical Physics*, 57(1):015215, January 2016.

[37] Benoît Collins, Ion Nechita, and Karol Życzkowski. Random graph states, maximal flow and Fuss-Catalan distributions. *Journal of Physics A Mathematical General*, 43(27):275303, July 2010.

[38] Tom Cooney, Milán Mosonyi, and Mark M. Wilde. Strong converse exponents for a quantum channel discrimination problem and quantum-feedback-assisted communication. *arXiv e-prints*, page arXiv:1408.3373, August 2014.

[39] Sean Cooper, Moshe Rozali, Brian Swingle, Mark Van Raamsdonk, Christopher Waddell, and David Wakeham. Black hole microstate cosmology. *Journal of High Energy Physics*, 2019(7):65, July 2019.

[40] Thomas H Cormen, Charles E Leiserson, Ronald L Rivest, and Clifford Stein. *Introduction to algorithms*. MIT press, 2009.

[41] Jordan Cotler, Patrick Hayden, Geoffrey Penington, Grant Salton, Brian Swingle, and Michael Walter. Entanglement Wedge Reconstruction via Universal Recovery Channels. *arXiv e-prints*, page arXiv:1704.05839, April 2017.

[42] Luca D'Alessio, Yariv Kafri, Anatoli Polkovnikov, and Marcos Rigol. From quantum chaos and eigenstate thermalization to statistical mechanics and thermodynamics. *Advances in Physics*, 65(3):239–362, May 2016.

[43] J. M. Deutsch. Quantum statistical mechanics in a closed system. *Physical Review A*, 43(4):2046–2049, February 1991.

[44] J. M. Deutsch. Thermodynamic entropy of a many-body energy eigenstate. *New Journal of Physics*, 12(7):075021, July 2010.

[45] Joshua M. Deutsch. Eigenstate thermalization hypothesis. *Reports on Progress in Physics*, 81(8):082001, August 2018.

[46] Xi Dong. The gravity dual of Rényi entropy. *Nature Communications*, 7:12472, August 2016.

[47] Xi Dong, Daniel Harlow, and Donald Marolf. Flat entanglement spectra in fixed-area states of quantum gravity. *Journal of High Energy Physics*, 2019(10):240, October 2019.

[48] Xi Dong, Daniel Harlow, and Aron C. Wall. Reconstruction of Bulk Operators within the Entanglement Wedge in Gauge-Gravity Duality. *Physical Review Letters*, 117(2):021601, July 2016.

[49] Xi Dong and Aitor Lewkowycz. Entropy, extremality, euclidean variations, and the equations of motion. *Journal of High Energy Physics*, 2018(1):81, January 2018.

[50] Xi Dong, Sean McBride, and Wayne W. Weng. Replica Wormholes and Holographic Entanglement Negativity. *arXiv e-prints*, page arXiv:2110.11947, October 2021.

[51] Xi Dong, Xiao-Liang Qi, and Michael Walter. Holographic entanglement negativity and replica symmetry breaking. *Journal of High Energy Physics*, 2021(6):24, June 2021.

[52] Xi Dong and Huajia Wang. Enhanced corrections near holographic entanglement transitions: a chaotic case study. *Journal of High Energy Physics*, 2020(11):7, November 2020.

[53] Anatoly Dymarsky, Nima Lashkari, and Hong Liu. Subsystem eigenstate thermalization hypothesis. *Physical Review E*, 97(1):012140, January 2018.

[54] Jens Eisert and Martin B. Plenio. A comparison of entanglement measures. *Journal of Modern Optics*, 46(1):145–154, January 1999.

[55] Andreas Elben, Benoît Vermersch, Rick van Bijnen, Christian Kokail, Tiff Brydges, Christine Maier, Manoj K. Joshi, Rainer Blatt, Christian F. Roos, and Peter Zoller. Cross-Platform Verification of Intermediate Scale Quantum Devices. *Physical Review Letters*, 124(1):010504, January 2020.

[56] P. Elias, A. Feinstein, and C. Shannon. A note on the maximum flow through a network. *IRE Transactions on Information Theory*, 2(4):117–119, 1956.

[57] Peter Elias, Amiel Feinstein, and Claude Shannon. A note on the maximum flow through a network. *IRE Transactions on Information Theory*, 2(4):117–119, 1956.

[58] Netta Engelhardt and Aron C. Wall. Quantum extremal surfaces: holographic entanglement entropy beyond the classical regime. *Journal of High Energy Physics*, 2015:73, January 2015.

[59] M. Fannes. A continuity property of the entropy density for spin lattice systems. *Communications in Mathematical Physics*, 31(4):291 – 294, 1973.

[60] Mark Fannes. A continuity property of the entropy density for spin lattice systems. *Communications in Mathematical Physics*, 31(4):291–294, 1973.

[61] Thomas Faulkner. The holographic map as a conditional expectation. *arXiv e-prints*, page arXiv:2008.04810, August 2020.

[62] Thomas Faulkner and Stefan Hollands. Approximate recoverability and relative entropy II: 2-positive channels of general v. Neumann algebras. *arXiv e-prints*, page arXiv:2010.05513, October 2020.

[63] Thomas Faulkner, Stefan Hollands, Brian Swingle, and Yixu Wang. Approximate recovery and relative entropy I. general von Neumann subalgebras. *arXiv e-prints*, page arXiv:2006.08002, June 2020.

[64] Thomas Faulkner, Aitor Lewkowycz, and Juan Maldacena. Quantum corrections to holographic entanglement entropy. *Journal of High Energy Physics*, 2013:74, November 2013.

[65] L. R. Ford and D. R. Fulkerson. Maximal flow through a network. *Canadian Journal of Mathematics*, 8:399–404, 1956.

[66] Lester Randolph Ford and Delbert R Fulkerson. Maximal flow through a network. *Canadian journal of Mathematics*, 8:399–404, 1956.

[67] Christopher A. Fuchs and Jeroen van de Graaf. Cryptographic Distinguishability Measures for Quantum Mechanical States. *arXiv e-prints*, pages quant–ph/9712042, December 1997.

[68] Keiichiro Furuya, Nima Lashkari, and Shoy Ouseph. Monotonic multi-state quantum f-divergences. *arXiv e-prints*, page arXiv:2103.09893, March 2021.

[69] James R. Garrison and Tarun Grover. Does a Single Eigenstate Encode the Full Hamiltonian? *Physical Review X*, 8(2):021026, April 2018.

[70] Elliott Gesteau and Monica Jinwoo Kang. Nonperturbative gravity corrections to bulk reconstruction. *arXiv e-prints*, page arXiv:2112.12789, December 2021.

[71] Sevag Gharibian. Strong NP-Hardness of the Quantum Separability Problem. *arXiv e-prints*, page arXiv:0810.4507, October 2008.

[72] Wu-zhong Guo, Feng-Li Lin, and Jiaju Zhang. Distinguishing Black Hole Microstates using Holevo Information. *Physical Review Letters*, 121(25):251603, December 2018.

[73] Leonid Gurvits. Classical deterministic complexity of Edmonds' problem and Quantum Entanglement. *arXiv e-prints*, pages quant–ph/0303055, March 2003.

[74] Daniel Harlow. The Ryu-Takayanagi Formula from Quantum Error Correction. *Communications in Mathematical Physics*, 354(3):865–912, September 2017.

[75] Thomas Hartman and Juan Maldacena. Time evolution of entanglement entropy from black hole interiors. *Journal of High Energy Physics*, 2013:14, May 2013.

[76] S. W. Hawking. Particle creation by black holes. *Communications in Mathematical Physics*, 43(3):199 – 220, 1975.

[77] Masahito Hayashi. Optimal sequence of quantum measurements in the sense of Stein's lemma in quantum hypothesis testing. *Journal of Physics A Mathematical General*, 35(50):10759–10773, December 2002.

[78] Masahito Hayashi. *Quantum Information: An Introduction*. Springer, Berlin, Heidelberg, 2006.

[79] Masahito Hayashi. Error exponent in asymmetric quantum hypothesis testing and its application to classical-quantum channel coding. *Physical Review A*, 76(6):062301, December 2007.

[80] Masahito Hayashi and Marco Tomamichel. Correlation Detection and an Operational Interpretation of the Renyi Mutual Information. *arXiv e-prints*, page arXiv:1408.6894, August 2014.

[81] Patrick Hayden, Sepehr Nezami, Xiao-Liang Qi, Nathaniel Thomas, Michael Walter, and Zhao Yang. Holographic duality from random tensor networks. *Journal of High Energy Physics*, 2016(11):9, November 2016.

[82] Patrick Hayden, Onkar Parrikar, and Jonathan Sorce. The Markov gap for geometric reflected entropy. *arXiv e-prints*, page arXiv:2107.00009, June 2021.

[83] Patrick Hayden and Geoffrey Penington. Learning the Alpha-bits of black holes. *Journal of High Energy Physics*, 2019(12):7, December 2019.

[84] Patrick Hayden and John Preskill. Black holes as mirrors: quantum information in random subsystems. *Journal of High Energy Physics*, 2007(9):120, September 2007.

[85] Carl W. Helstrom. Quantum detection and estimation theory. *Journal of Statistical Physics*, 1(2):231–252, June 1969.

[86] Carl Wilhelm Helstrom. *Quantum detection and estimation theory*. Mathematics in science and engineering. Academic Press, New York, NY, 1976.

[87] Fumio Hiai and Dénes Petz. The proper formula for relative entropy and its asymptotics in quantum probability. *Communications in Mathematical Physics*, 143(1):99 – 114, 1991.

[88] A.S Holevo. Statistical decision theory for quantum systems. *Journal of Multivariate Analysis*, 3(4):337–394, 1973.

[89] Stefan Hollands and Ko Sanders. Entanglement measures and their properties in quantum field theory. *arXiv e-prints*, page arXiv:1702.04924, February 2017.

[90] Christoph Holzhey, Finn Larsen, and Frank Wilczek. Geometric and renormalized entropy in conformal field theory. *Nuclear Physics B*, 424(3):443–467, August 1994.

[91] Michał Horodecki, Paweł Horodecki, and Ryszard Horodecki. Separability of mixed states: necessary and sufficient conditions. *Physics Letters A*, 223(1):1–8, February 1996.

[92] Veronika E. Hubeny, Henry Maxfield, Mukund Rangamani, and Erik Tonni. Holographic entanglement plateaux. *Journal of High Energy Physics*, 2013:92, August 2013.

[93] Veronika E. Hubeny, Mukund Rangamani, and Tadashi Takayanagi. A covariant holographic entanglement entropy proposal. *Journal of High Energy Physics*, 2007(7):062, July 2007.

[94] Nicholas Hunter-Jones. Unitary designs from statistical mechanics in random quantum circuits. *arXiv e-prints*, page arXiv:1905.12053, May 2019.

[95] Daniel L. Jafferis, Aitor Lewkowycz, Juan Maldacena, and S. Josephine Suh. Relative entropy equals bulk relative entropy. *Journal of High Energy Physics*, 2016(6):4, June 2016.

[96] Cheryne Jonay, David A. Huse, and Adam Nahum. Coarse-grained dynamics of operator and state entanglement. *arXiv e-prints*, page arXiv:1803.00089, February 2018.

[97] Richard Jozsa. Fidelity for mixed quantum states. *Journal of Modern Optics*, 41(12):2315–2323, 1994.

[98] Marius Junge, Renato Renner, David Sutter, Mark M. Wilde, and Andreas Winter. Universal recovery maps and approximate sufficiency of quantum relative entropy. *arXiv e-prints*, page arXiv:1509.07127, September 2015.

[99] J. Jurkiewicz, G. Łukaszewski, and M. A. Nowak. Diagrammatic Approach to Fluctuations in the Wishart Ensemble. *Acta Physica Polonica B*, 39(4):799, April 2008.

[100] Sumeet Khatri and Mark M. Wilde. Principles of Quantum Communication Theory: A Modern Approach. *arXiv e-prints*, page arXiv:2011.04672, November 2020.

[101] A. S. Kholevo. On quasiequivalence of locally normal states. *Theoretical and Mathematical Physics*, 13(2):1071–1082, November 1972.

[102] Isaac Kim, Eugene Tang, and John Preskill. The ghost in the radiation: robust encodings of the black hole interior. *Journal of High Energy Physics*, 2020(6):31, June 2020.

[103] Alexei Kitaev. A simple model of quantum holography. KITP Program: Entanglement in Strongly-Correlated Quantum Matter, 2015.

[104] Ioanna Kourkoulou and Juan Maldacena. Pure states in the SYK model and nearly-AdS_2 gravity. *arXiv e-prints*, page arXiv:1707.02325, July 2017.

[105] G. Kreweras. Sur les partitions non croisees d'un cycle. *Discrete Mathematics*, 1(4):333 – 350, 1972.

[106] Jonah Kudler-Flam. Relative Entropy of Random States and Black Holes. *Physical Review Letters*, 126(17):171603, April 2021.

[107] Jonah Kudler-Flam, Yuya Kusuki, and Shinsei Ryu. Correlation measures and the entanglement wedge cross-section after quantum quenches in two-dimensional conformal field theories. *Journal of High Energy Physics*, 2020(4):74, April 2020.

[108] Jonah Kudler-Flam, Vladimir Narovlansky, and Shinsei Ryu. Distinguishing Random and Black Hole Microstates. *PRX Quantum*, 2(4):040340, November 2021.

[109] Jonah Kudler-Flam, Vladimir Narovlansky, and Shinsei Ryu. Negativity spectra in random tensor networks and holography. *Journal of High Energy Physics*, 2022(2):76, February 2022.

[110] Jonah Kudler-Flam, Masahiro Nozaki, Shinsei Ryu, and Mao Tian Tan. Quantum vs. classical information: operator negativity as a probe of scrambling. *Journal of High Energy Physics*, 2020(1):31, January 2020.

[111] Jonah Kudler-Flam, Masahiro Nozaki, Shinsei Ryu, and Mao Tian Tan. Entanglement of local operators and the butterfly effect. *Physical Review Research*, 3(3):033182, August 2021.

[112] Jonah Kudler-Flam and Pratik Rath. Large and Small Corrections to the JLMS Formula from Replica Wormholes. *arXiv e-prints*, page arXiv:2203.11954, March 2022.

[113] Santosh Kumar. Wishart and random density matrices: Analytical results for the mean-square Hilbert-Schmidt distance. *Physical Review A*, 102(1):012405, July 2020.

[114] Santosh Kumar and S. Sai Charan. Spectral statistics for the difference of two Wishart matrices. *Journal of Physics A Mathematical General*, 53(50):505202, November 2020.

[115] Aritra Laha, Agrim Aggarwal, and Santosh Kumar. Random density matrices: analytical results for mean root fidelity and mean square Bures distance. *arXiv e-prints*, page arXiv:2105.02743, May 2021.

[116] Nima Lashkari. Modular Hamiltonian for Excited States in Conformal Field Theory. *Physical Review Letters*, 117(4):041601, July 2016.

[117] Nima Lashkari, Anatoly Dymarsky, and Hong Liu. Eigenstate Thermalization Hypothesis in Conformal Field Theory. *arXiv e-prints*, page arXiv:1610.00302, October 2016.

[118] Felix Leditzky. Relative entropies and their use in quantum information theory. *arXiv e-prints*, page arXiv:1611.08802, November 2016.

[119] Aitor Lewkowycz and Juan Maldacena. Generalized gravitational entropy. *Journal of High Energy Physics*, 2013:90, August 2013.

[120] Ke Li. Discriminating quantum states: the multiple Chernoff distance. *arXiv e-prints*, page arXiv:1508.06624, August 2015.

[121] Yeong-Cherng Liang, Yu-Hao Yeh, Paulo E. M. F. Mendonça, Run Yan Teh, Margaret D. Reid, and Peter D. Drummond. Quantum fidelity measures for mixed states. *Reports on Progress in Physics*, 82(7):076001, July 2019.

[122] Elliott H Lieb. Convex trace functions and the wigner-yanase-dyson conjecture. *Advances in Mathematics*, 11(3):267–288, 1973.

[123] Göran Lindblad. Expectations and entropy inequalities for finite quantum systems. *Comm. Math. Phys.*, 39(2):111–119, 1974.

[124] Göran Lindblad. Completely positive maps and entropy inequalities. *Comm. Math. Phys.*, 40(2):147–151, 1975.

[125] Hong Liu and Shreya Vardhan. Void formation in operator growth, entanglement, and unitarity. *arXiv e-prints*, page arXiv:1912.08918, December 2019.

[126] Hong Liu and Shreya Vardhan. Entanglement entropies of equilibrated pure states in quantum many-body systems and gravity. *arXiv e-prints*, page arXiv:2008.01089, August 2020.

[127] Hong Liu and Shreya Vardhan. A dynamical mechanism for the Page curve from quantum chaos. *Journal of High Energy Physics*, 2021(3):88, March 2021.

[128] Tsung-Cheng Lu and Tarun Grover. Renyi Entropy of Chaotic Eigenstates. *arXiv e-prints*, page arXiv:1709.08784, September 2017.

[129] Donald Marolf, Shannon Wang, and Zhencheng Wang. Probing phase transitions of holographic entanglement entropy with fixed area states. *Journal of High Energy Physics*, 2020(12):84, December 2020.

[130] José Mejía, Camilo Zapata, and Alonso Botero. The difference between two random mixed quantum states: exact and asymptotic spectral analysis. *arXiv e-prints*, page arXiv:1511.07278, November 2015.

[131] Márk Mezei. Membrane theory of entanglement dynamics from holography. *Physical Review D*, 98(10):106025, November 2018.

[132] Márk Mezei and Julio Virrueta. Exploring the membrane theory of entanglement dynamics. *Journal of High Energy Physics*, 2020(2):13, February 2020.

[133] Ben Michel and Andrea Puhm. Corrections in the relative entropy of black hole microstates. *Journal of High Energy Physics*, 2018(7):179, July 2018.

[134] James A. Mingo and Roland Speicher. *Free Probability and Random Matrices*. Springer, 2017.

[135] Masamichi Miyaji, Tadashi Takayanagi, and Tomonori Ugajin. Spectrum of End of the World Branes in Holographic BCFTs. *arXiv e-prints*, page arXiv:2103.06893, March 2021.

[136] Wojciech Młotkowski, Maciej A. Nowak, Karol A. Penson, and Karol Życzkowski. Spectral density of generalized Wishart matrices and free multiplicative convolution. *Physical Review E*, 92(1):012121, July 2015.

[137] Ashley Montanaro. On the Distinguishability of Random Quantum States. *Communications in Mathematical Physics*, 273(3):619–636, August 2007.

[138] M. Mosonyi and F. Hiai. On the quantum Renyi relative entropies and related capacity formulas. *arXiv e-prints*, page arXiv:0912.1286, December 2009.

[139] Milán Mosonyi and Tomohiro Ogawa. Quantum Hypothesis Testing and the Operational Interpretation of the Quantum Rényi Relative Entropies. *Communications in Mathematical Physics*, 334(3):1617–1648, March 2015.

[140] Martin Müller-Lennert, Frédéric Dupuis, Oleg Szehr, Serge Fehr, and Marco Tomamichel. On quantum Rényi entropies: A new generalization and some properties. *Journal of Mathematical Physics*, 54(12):122203–122203, December 2013.

[141] Chaitanya Murthy and Mark Srednicki. Structure of chaotic eigenstates and their entanglement entropy. *Physical Review E*, 100(2):022131, August 2019.

[142] Hiroshi Nagaoka. The Converse Part of The Theorem for Quantum Hoeffding Bound. *arXiv e-prints*, pages quant–ph/0611289, November 2006.

[143] Adam Nahum, Jonathan Ruhman, Sagar Vijay, and Jeongwan Haah. Quantum Entanglement Growth under Random Unitary Dynamics. *Physical Review X*, 7(3):031016, July 2017.

[144] Adam Nahum, Sagar Vijay, and Jeongwan Haah. Operator Spreading in Random Unitary Circuits. *Physical Review X*, 8(2):021014, April 2018.

[145] Ion Nechita. Asymptotics of Random Density Matrices. *Annales Henri Poincaré*, 8(8):1521–1538, November 2007.

[146] Alexandru Nica and Roland Speicher. *Lectures on the Combinatorics of Free Probability*. London Mathematical Society Lecture Note Series. Cambridge University Press, 2006.

[147] Michael A. Nielsen and David Poulin. Algebraic and information-theoretic conditions for operator quantum error correction. *Physical Review A*, 75(6):064304, June 2007.

[148] Michael Nussbaum and Arleta Szkoła. The Chernoff lower bound for symmetric quantum hypothesis testing. *arXiv e-prints*, pages quant–ph/0607216, July 2006.

[149] Michael Nussbaum and Arleta Szkoła. Exponential error rates in multiple state discrimination on a quantum spin chain. *Journal of Mathematical Physics*, 51(7):072203–072203, July 2010.

[150] Michael Nussbaum and Arleta Szkoła. An asymptotic error bound for testing multiple quantum hypotheses. *arXiv e-prints*, page arXiv:1112.1529, December 2011.

[151] Michael Nussbaum and Arleta Szkoła. Asymptotically optimal discrimination between pure quantum states. In Wim van Dam, Vivien M. Kendon, and Simone Severini, editors, *Theory of Quantum Computation, Communication, and Cryptography*, pages 1–8, Berlin, Heidelberg, 2011. Springer Berlin Heidelberg.

[152] Tomohiro Ogawa and Hiroshi Nagaoka. *Strong Converse and Stein's Lemma in Quantum Hypothesis Testing*, pages 28–42. IEEE, 2005.

[153] M. Ohya and D. Petz. *Quantum Entropy and Its Use*. Theoretical and Mathematical Physics. Springer Berlin Heidelberg, 2004.

[154] Don N. Page. Average entropy of a subsystem. *Physical Review Letters*, 71(9):1291–1294, August 1993.

[155] Don N. Page. Information in black hole radiation. *Physical Review Letters*, 71(23):3743–3746, December 1993.

[156] Fernando Pastawski, Beni Yoshida, Daniel Harlow, and John Preskill. Holographic quantum error-correcting codes: toy models for the bulk/boundary correspondence. *Journal of High Energy Physics*, 2015:149, June 2015.

[157] Geoff Penington, Stephen H. Shenker, Douglas Stanford, and Zhenbin Yang. Replica wormholes and the black hole interior. *arXiv e-prints*, page arXiv:1911.11977, November 2019.

[158] Geoffrey Penington. Entanglement wedge reconstruction and the information paradox. *Journal of High Energy Physics*, 2020(9):2, September 2020.

[159] Karol A. Penson and Karol Życzkowski. Product of Ginibre matrices: Fuss-Catalan and Raney distributions. *Physical Review E*, 83(6):061118, June 2011.

[160] Asher Peres. Separability Criterion for Density Matrices. *Physical Review Letters*, 77(8):1413–1415, August 1996.

[161] Dénes Petz. Quasi-entropies for finite quantum systems. *Reports on Mathematical Physics*, 23(1):57–65, February 1986.

[162] Dénes Petz. Monotonicity of Quantum Relative Entropy Revisited. *Reviews in Mathematical Physics*, 15(1):79–91, January 2003.

[163] DÉNES PETZ. SUFFICIENCY OF CHANNELS OVER VON NEUMANN ALGEBRAS. *The Quarterly Journal of Mathematics*, 39(1):97–108, 03 1988.

[164] Dénes Petz. Sufficient subalgebras and the relative entropy of states of a von Neumann algebra. *Communications in Mathematical Physics*, 105(1):123 – 131, 1986.

[165] M. B. Plenio. Logarithmic Negativity: A Full Entanglement Monotone That is not Convex. *Physical Review Letters*, 95(9):090503, August 2005.

[166] John Preskill. Quantum Computing in the NISQ era and beyond. *arXiv e-prints*, page arXiv:1801.00862, January 2018.

[167] Zbigniew Puchała, Łukasz Pawela, and Karol Życzkowski. Distinguishability of generic quantum states. *Physical Review A*, 93(6):062112, June 2016.

[168] Xiao-Liang Qi, Zhou Shangnan, and Zhenbin Yang. Holevo information and ensemble theory of gravity. *Journal of High Energy Physics*, 2022(2):56, February 2022.

[169] Renato Renner. *Security of Quantum Key Distribution*. PhD thesis, -, December 2005.

[170] Alfred Renyi. On measures of entropy and information. *Proc. of the Fourth Berkeley Symp. on Math. Statist. and Prob.*, 1:547–561, 1961.

[171] Moshe Rozali, James Sully, Mark Van Raamsdonk, Christopher Waddell, and David Wakeham. Information radiation in BCFT models of black holes. *Journal of High Energy Physics*, 2020(5):4, May 2020.

[172] Shinsei Ryu and Tadashi Takayanagi. Aspects of holographic entanglement entropy. *Journal of High Energy Physics*, 2006(8):045, August 2006.

[173] Shinsei Ryu and Tadashi Takayanagi. Holographic Derivation of Entanglement Entropy from the anti de Sitter Space/Conformal Field Theory Correspondence. *Physical Review Letters*, 96(18):181602, May 2006.

[174] Robert Salzmann and Nilanjana Datta. Interpolating between symmetric and asymmetric hypothesis testing. *arXiv e-prints*, page arXiv:2104.09553, April 2021.

[175] Gábor Sárosi and Tomonori Ugajin. Relative entropy of excited states in conformal field theories of arbitrary dimensions. *Journal of High Energy Physics*, 2017(2):60, February 2017.

[176] Hassan Shapourian, Shang Liu, Jonah Kudler-Flam, and Ashvin Vishwanath. Entanglement Negativity Spectrum of Random Mixed States: A Diagrammatic Approach. *PRX Quantum*, 2(3):030347, September 2021.

[177] Rodica Simion. Noncrossing partitions. *Discrete Mathematics*, 217(1):367 – 409, 2000.

[178] Hans-Jürgen Sommers and Karol Zyczkowski. Statistical properties of random density matrices. *Journal of Physics A Mathematical General*, 37(35):8457–8466, September 2004.

[179] Mark Srednicki. Chaos and quantum thermalization. *Physical Review E*, 50(2):888–901, August 1994.

[180] Douglas Stanford and Edward Witten. Fermionic localization of the schwarzian theory. *Journal of High Energy Physics*, 2017(10):8, October 2017.

[181] A. Uhlmann. Relative entropy and the Wigner-Yanase-Dyson-Lieb concavity in an interpolation theory. *Communications in Mathematical Physics*, 54(1):21 – 32, 1977.

[182] Shreya Vardhan, Jonah Kudler-Flam, Hassan Shapourian, and Hong Liu. Bound entanglement in thermalized states and black hole radiation. *arXiv e-prints*, page arXiv:2110.02959, October 2021.

[183] Shreya Vardhan, Jonah Kudler-Flam, Hassan Shapourian, and Hong Liu. Mixed-state entanglement and information recovery in thermalized states and evaporating black holes. *arXiv e-prints*, page arXiv:2112.00020, November 2021.

[184] G. Vidal and R. F. Werner. Computable measure of entanglement. *Physical Review A*, 65(3):032314, March 2002.

[185] Lev Vidmar and Marcos Rigol. Entanglement Entropy of Eigenstates of Quantum Chaotic Hamiltonians. *Physical Review Letters*, 119(22):220603, December 2017.

[186] C. W. von Keyserlingk, Tibor Rakovszky, Frank Pollmann, and S. L. Sondhi. Operator Hydrodynamics, OTOCs, and Entanglement Growth in Systems without Conservation Laws. *Physical Review X*, 8(2):021013, April 2018.

[187] Huajia Wang and Tianci Zhou. Barrier from chaos: operator entanglement dynamics of the reduced density matrix. *Journal of High Energy Physics*, 2019(12):20, December 2019.

[188] Jinzhao Wang. The refined quantum extremal surface prescription from the asymptotic equipartition property. *arXiv e-prints*, page arXiv:2105.05892, May 2021.

[189] Eugene P. Wigner. Characteristic vectors of bordered matrices with infinite dimensions. *Annals of Mathematics*, 62(3):548–564, 1955.

[190] Mark M. Wilde. Recoverability for Holevo's just-as-good fidelity. *arXiv e-prints*, page arXiv:1801.02800, January 2018.

[191] Mark M. Wilde, Andreas Winter, and Dong Yang. Strong Converse for the Classical Capacity of Entanglement-Breaking and Hadamard Channels via a Sandwiched Rényi Relative Entropy. *Communications in Mathematical Physics*, 331(2):593–622, October 2014.

[192] Edward Witten. Notes on Some Entanglement Properties of Quantum Field Theory. *arXiv e-prints*, page arXiv:1803.04993, March 2018.

[193] Michael M. Wolf, Frank Verstraete, Matthew B. Hastings, and J. Ignacio Cirac. Area Laws in Quantum Systems: Mutual Information and Correlations. *Physical Review Letters*, 100(7):070502, February 2008.

[194] H. Yuen, R. Kennedy, and M. Lax. Optimum testing of multiple hypotheses in quantum detection theory. *IEEE Transactions on Information Theory*, 21(2):125–134, 1975.

[195] Jiaju Zhang, Paola Ruggiero, and Pasquale Calabrese. Subsystem Trace Distance in Quantum Field Theory. *Physical Review Letters*, 122(14):141602, April 2019.

[196] Tianci Zhou and Adam Nahum. Entanglement Membrane in Chaotic Many-Body Systems. *Physical Review X*, 10(3):031066, July 2020.

[197] Karol Życzkowski, Karol A. Penson, Ion Nechita, and Benoît Collins. Generating random density matrices. *Journal of Mathematical Physics*, 52(6):062201–062201, June 2011.

[198] Karol Zyczkowski and Hans-Jürgen Sommers. Induced measures in the space of mixed quantum states. *Journal of Physics A Mathematical General*, 34(35):7111–7125, September 2001.

[199] Karol Życzkowski and Hans-Jürgen Sommers. Average fidelity between random quantum states. *Physical Review A*, 71(3):032313, March 2005.