

Stabilizer complexity of Hawking radiation

Ritam Basu^{✉,*}, Onkar Parrikar[†], Suprakash Paul[‡], and Harshit Rajgadga[§]

*Department of Theoretical Physics, Tata Institute of Fundamental Research,
1 Homi Bhabha Road, Mumbai 400005, India*



(Received 4 March 2026; accepted 13 July 2026; published 18 August 2026)

We study the complexity of Hawking radiation for an evaporating black hole from the perspective of the stabilizer theory of quantum computation. Specifically, we calculate Wigner negativity—a magic monotone that can be interpreted as a measure of the stabilizer complexity, or equivalently, the complexity of classical simulation—in various toy models for evaporating black holes. We first calculate the Wigner negativity of Hawking radiation in the Penington-Shenker-Stanford-Yang (PSSY) model directly using the gravitational path integral, and show that the negativity is $O(1)$ before the Page transition, but becomes exponentially large past the Page transition. We also derive a universal, information theoretic formula for the negativity that interpolates between the two extremes. We then study the Wigner negativity of radiation in a dynamical model of black hole evaporation. In this case, the negativity shows a sharp spike at early times resulting from the coupling between the black hole and the radiation system, but at late times when the system settles down, we find that the negativity satisfies the same universal formula as in the PSSY model. Finally, we also propose a geometric formula for Wigner negativity in general holographic states using intuition from fixed area states and random tensor networks, and argue that a python’s lunch in the entanglement wedge implies a stabilizer complexity that is exponentially large in $\frac{1}{8G_N}$ times the difference between the areas corresponding to the outermost and minimal extremal surfaces.

DOI: [10.1103/49df-19n4](https://doi.org/10.1103/49df-19n4)

I. INTRODUCTION

The black hole information paradox is a sharp puzzle that highlights the mysterious nature of black holes. According to Hawking’s seminal work [1], a black hole emits thermal radiation, and so a black hole formed from the gravitational collapse of matter in a pure state apparently evaporates away leaving radiation in an approximately thermal state. This process seemingly violates the unitarity of quantum mechanics. In more technical terms, Hawking’s calculation suggests that the von Neumann entropy of the state of the radiation keeps growing with time until the end of evaporation, violating the basic fact that entropy of a bipartite quantum state is upper bounded by the dimension of the Hilbert space of the smaller subsystem, and should thus satisfy the Page curve [2]. Recent progress on the black hole information problem [3–8] (see also [9] for a

review and further references) has clarified how a unitary Page curve emerges in the calculation of the radiation entropy using generalizations of ideas developed in the AdS/CFT context. A heuristic way to understand why the radiation entropy keeps rising in Hawking’s picture is that the black hole emits more and more Hawking quanta with time, and each of these quanta is entangled with its Hawking partner behind the black hole horizon. So, the increasing radiation entropy stems from a continuous buildup of entanglement between the early radiation and the degrees of freedom in the black hole interior. A crucial insight in the recent resolution of the information paradox was that after the Page time, a new quantum extremal surface close to the black hole horizon becomes dominant. From the entanglement wedge reconstruction paradigm [10–16], this implies that the black hole interior degrees of freedom are no longer independent degrees of freedom entangled with the radiation, but are in fact encoded in the early radiation. A crucial role in this analysis is played by replica wormholes [7,8]—nontrivial Euclidean wormhole saddle points of gravity that dominate the Rényi entropy calculation after Page time and stop the growth of the radiation entropy predicted by Hawking’s calculation.

While the above resolution to the black hole information paradox is extremely elegant and satisfying in many ways, it highlights the essentially nonlocal nature of quantum gravity: degrees of freedom behind the black hole horizon,

*Contact author: ritam.basu@tifr.res.in

†Contact author: parrikar@theory.tifr.res.in

‡Contact author: suprakash.paul@tifr.res.in

§Contact author: harshit.rajgadga@tifr.res.in

Published by the American Physical Society under the terms of the Creative Commons Attribution 4.0 International license. Further distribution of this work must maintain attribution to the author(s) and the published article’s title, journal citation, and DOI. Funded by SCOAP³.

which are causally inaccessible to an observer at infinity, can nevertheless be encoded in—and thus manipulated via—the degrees of freedom at infinity. In local quantum field theory without gravity, this would be impossible; what makes it possible (at least, in principle) in gravity is the existence of gravitational wormholes.¹ Thus, if the semiclassical picture of the black hole interior makes sense, then it appears that quantum gravity must be fundamentally nonlocal, in that an observer with access to all the early radiation from a sufficiently old black hole can perform a unitary that manipulates a qubit deep inside the causally inaccessible black hole interior. Nevertheless, following the pioneering work of Harlow and Hayden [20], several authors have suggested that observers at infinity with *computationally bounded resources* should not be able to perform such causality violating operations [21–24] (see also [25,26]). Thus, the semiclassical notion of spacetime is protected by computational complexity.

A very similar situation arises in a more general context in AdS/CFT, where a fixed boundary subregion B in some states can have multiple corresponding homologous extremal surfaces in the bulk spacetime. If the outermost extremal surface (on the side of the asymptotic subregion B) does not have the smallest area, then the entanglement wedge is said to have a python’s lunch [21]. The python’s lunch region—i.e., the portion of the entanglement wedge behind the outermost extremal surface, is causally inaccessible from the asymptotic boundary, and in fact cannot be brought in causal contact with the asymptotic boundary by simple boundary operations [27]. Nevertheless, semiclassical degrees of freedom in the python’s lunch region are encoded in and can be manipulated from the boundary subregion B , much like in the case of the black hole interior in our previous discussion. Taking inspiration from the analogy between holography and tensor networks, it was conjectured in [21] (see also [27–31]) that for degrees of freedom in the python’s lunch region, the complexity of bulk reconstruction should scale as,

$$C \sim \exp \left[\frac{1}{2} \frac{(A(\gamma_{\text{bulge}}) - A(\gamma_{\text{out}}))}{4G_N} \right], \quad (1)$$

where γ_{out} is the outermost extremal surface and γ_{bulge} is the “bulge” surface (roughly corresponding to the maximal cross section of the python’s lunch, but see [21] for a more precise definition). The python’s lunch conjecture tries to make precise the idea that the semiclassical notion of spacetime in AdS/CFT is computationally (but not information theoretically) protected. It would be useful to have

¹Also, as emphasized in [17–19], the ADM Hamiltonian in gravity is a boundary term at infinity and well-defined observables in gravity are nonlocal on account of the need for gravitational dressing. These features already point to an essential nonlocality in gravity.

some concrete boundary calculations that shed light on this conjecture; part of the difficulty with this is the lack of a computable notion of complexity.

In this paper, we study the complexity of the Hawking radiation from the perspective of the stabilizer theory of quantum computation. The Gottesman-Knill theorem [32–34] identifies a class of circuits—called stabilizer circuits—that can generate arbitrary amounts of entanglement, and yet be efficiently simulated on classical computers. In the latter sense, stabilizer circuits are essentially classical, and do not have inherent quantum computational advantage. As one might expect, the stabilizer model of quantum computation is not universal. It was proposed by Bravyi and Kitaev [35] that one can get universal quantum computation by supplementing stabilizer circuits with some number of external “resource” states that provide the nonstabilizer-ness required for universal quantum computation. These authors called this quantum resource “magic.” In [36,37], a resource theory of magic was developed that centered around the construction of magic monotones—measures of magic that do not increase under stabilizer operations. One can think of a magic monotone as a measure of the “stabilizer complexity,” or the nonstabilizer content of a quantum circuit. Intuitively, we can also regard it as the complexity of classically simulating the quantum circuit on a classical computer. Remarkably, it was shown in [36] (building on the work of [34]) that a simple and computable magic monotone can be constructed out of the *Wigner function*. The Wigner function is a quasi-probability distribution that seeks to represent a quantum state as a distribution in phase space. For finite (prime) dimensional Hilbert spaces, the phase space is taken to be a lattice, and the corresponding Wigner function is called the discrete Wigner function [38–44]. The discrete Wigner function satisfies many of the properties of a standard probability distribution, except for one—it can take negative values. It was shown in [36] that the amount of negativity in the Wigner function [see (14) for the definition] is a magic monotone. In [45], more detailed arguments were given to show that the Wigner negativity in a circuit provides a measure of the complexity of classical simulation using quasiprobability methods.

Our goal in this paper is to study the stabilizer complexity of the state of the Hawking radiation in an evaporating black hole. More precisely, we will work with Wigner negativity as our magic monotone, and calculate the negativity of the radiation in toy models of black hole evaporation. Our main results are as follows:

- (i) In Sec. III, we consider the Penington-Shenker-Stanford-Yang (PSSY) model [7] for black hole evaporation in Jackiw-Teitelboim (JT) gravity coupled to an external bath. In this model, we compute the Wigner negativity directly by using the gravitational path integral and show that the negativity is $O(1)$ before the Page transition, but

grows exponentially large after the Page transition and is given by

$$\mathcal{N} \sim \sqrt{\frac{2}{\pi}} \exp \left[\frac{1}{2} (S_{\max} - S_2) \right], \quad (2)$$

where $S_{\max} = \log D$ is the entropy of the maximally mixed state on the radiation system, while S_2 is the second Rényi entropy of the radiation after the Page point. We also derive a more general formula that interpolates between the two extreme regimes. Importantly, gravity gives a universal, information theoretic formula for the Wigner negativity, essentially independent of the choice of computational basis for the bath subsystem.² In this calculation, an important role is played by two-boundary Euclidean wormholes (the $n = 2$ replica wormholes), which end up dominating past the Page point.

- (ii) In Sec. IV, we then turn to a dynamical model of black hole evaporation where the black hole is coupled to a bath and evolved in time by a joint Hamiltonian involving interaction terms between the black hole and the bath. With the simplifying assumption that the bath operators that couple to the black hole look like random matrices (picked from some unitarily invariant ensemble) with respect to the computational basis, we once again derive a universal formula for the Wigner negativity of the radiation system. This negativity shows a sharp spike (sort of like a “negativity shock”) immediately after the bath is coupled to the black hole, but then decays due to decoherence and settles down at late times to a time-independent value that precisely matches the gravity formula we obtained in the PSSY model.
- (iii) Finally, in Sec. V, we consider Wigner negativity in general holographic states satisfying the Ryu-Takayanagi formula for entanglement entropy. Here one must define the negativity with some care, because the boundary Hilbert space in AdS/CFT is infinite dimensional to begin with. Using intuition from fixed-area states and random tensor networks, we propose a geometric formula for the Wigner negativity: whenever there is a large gap between $\frac{1}{4G_N}$ times the areas of the outermost extremal surface γ_{out} and the true extremal surface γ_{min} (with minimal area), the Wigner negativity is given by

$$\mathcal{N} \sim \sqrt{\frac{2}{\pi}} \exp \left[\frac{1}{2} \frac{(A(\gamma_{\text{out}}) - A(\gamma_{\text{min}}))}{4G_N} \right], \quad (3)$$

²This essentially happens because the gravitational path integral computes an ensemble average over Hamiltonians and pseudorandom boundary conditions in the dual theory.

while the negativity is $O(1)$ when the gap vanishes. As a consequence of this formula, a python’s lunch region in the entanglement wedge corresponds to an exponentially large stabilizer complexity. Our formula for the negativity has a similar flavor to the python’s lunch conjecture, but also different—first it does not seem to involve the bulge surface, and second the Wigner negativity is not directly the same as the reconstruction complexity relevant for the python’s lunch conjecture.

II. PRELIMINARIES

A. The discrete Wigner function

For any finite dimensional quantum system of odd, prime dimension D ,³ there exists a natural quasiprobability representation called the *discrete Wigner function* [38–44]. Let $\{|k\rangle\}_{k=0}^{D-1}$ be an ordered, orthonormal basis for the Hilbert space. The essential idea is to interpret this as the “position” basis, and correspondingly construct a “phase space.” We define the *discrete phase space* $\mathcal{P}$ as the lattice $\mathbb{Z}_D \times \mathbb{Z}_D$ of size D^2 . With respect to the chosen basis, we define a set of D^2 operators $A(q, p)$ called phase-point operators, each labeled by a phase space point,

$$A(q, p) = \sum_{k, \ell=0}^{D-1} \hat{\delta}_{2q, k+\ell} e^{\frac{2\pi i}{D}(k-\ell)p} |k\rangle \langle \ell|, \quad (4)$$

where the hatted Kronecker delta $\hat{\delta}$ is the mod D version, i.e., it is one when $(k + \ell) = 2q \bmod D$, and zero otherwise. The discrete Wigner function for a density matrix ρ is now defined (in analogy with the continuous case [46]) as,

$$W_\rho(q, p) = \frac{1}{D} \text{Tr}(\rho A(q, p)). \quad (5)$$

The Wigner function attempts to represent the quantum state ρ as a probability distribution in phase space, much like is the case in classical mechanics. Indeed, it satisfies the following properties:

- (1) The discrete Wigner function is real and is unit normalized, i.e., $\sum_{q, p=0}^{D-1} W_\rho(q, p) = 1$, assuming $\text{Tr}(\rho) = 1$.
- (2) Summing over one of the directions, say either p or q , reduces the Wigner function to the probability

³Discrete Wigner function formalism is nicest when D is either an odd prime or a power of an odd prime, although most aspects of the formalism work more generally for odd D , and generalizations exist for even D as well. When D is an odd prime power, the Wigner function can be shown to be the unique quasiprobability distribution, which “geometrizes” the action of the Clifford group in terms of symplectic transformations on phase space [43]. For simplicity, we will take D to be an odd prime number in this work. Most of our results should also work when D is a power of an odd prime.

density along the other direction,

$$\begin{aligned} \sum_{p=0}^{D-1} W_\rho(q, p) &= \langle q | \rho | q \rangle, \\ \sum_{q=0}^{D-1} W_\rho(q, p) &= \langle p | \rho | p \rangle, \end{aligned} \quad (6)$$

where we have defined the “momentum eigenstates” as $|p\rangle = \frac{1}{\sqrt{D}} \sum_{q=0}^{D-1} e^{\frac{2\pi i p q}{D}} |q\rangle$.

(3) The time evolution of the discrete Wigner function follows a discrete version of the Moyal equation [38], which can be thought of as the quantum analog of the Liouville equation from classical mechanics. While the above properties seem to suggest that we should regard the Wigner function as a probability distribution in phase space, this interpretation fails for an important reason—the Wigner function can take negative values at some points in phase space. Indeed, had this not been the case, then quantum mechanics would reduce to classical probabilistic dynamics. The negativity of the Wigner function will play a central role in this work.

So far we have discussed Wigner functions for states, but Wigner functions can also be defined for quantum operations/channels [34,47]. For a quantum channel $\mathcal{E}(\rho) = \sum_m E_m \rho E_m^\dagger$ given in terms of a Kraus operator representation $\{E_m\}$, the Wigner function is defined as

$$W_{\mathcal{E}}(q, p | q', p') = \frac{1}{D} \sum_m \text{Tr}(A(q, p) E_m A(q', p') E_m^\dagger). \quad (7)$$

This can be thought of as a representation of the quantum channel as a quasistochastic matrix on phase space. Indeed, given an input state ρ with the Wigner function $W_\rho(q, p)$, the Wigner function of the output state $\mathcal{E}(\rho)$ satisfies

$$W_{\mathcal{E}(\rho)}(q, p) = \sum_{q', p'} W_{\mathcal{E}}(q, p | q', p') W_\rho(q', p'). \quad (8)$$

Similarly, for a positive operator valued measure $\{M_n\}$, the corresponding Wigner function is defined as,

$$W(n | q, p) = \text{Tr}(M_n A(q, p)). \quad (9)$$

This definition is designed to satisfy,

$$\text{Tr}(M_n \rho) = \sum_{q, p} W(n | q, p) W_\rho(q, p), \quad (10)$$

and turns the Positive operator-valued measure (POVM) into a quasiconditional probability distribution. In this way, the outcome of a general quantum circuit,

$$P_n = \text{Tr}(M_n \mathcal{E}_m \cdots \mathcal{E}_1(\rho)), \quad (11)$$

translates to quasistochastic evolution in terms of the Wigner function representations,

$$\begin{aligned} P_n &= \sum_{q, p} \sum_{q_1, p_1, \dots, q_m, p_m} \sum_{q', p'} W(n | q, p) \\ &\quad \times W_{\mathcal{E}_m}(q_m, p_m | q_{m-1}, p_{m-1}) \cdots \\ &\quad \times W_{\mathcal{E}_1}(q_1, p_1 | q', p') W_\rho(q', p'), \end{aligned} \quad (12)$$

with the important caveat that while P_n is always non-negative, the intermediate Wigner functions are not always everywhere non-negative. From the above discussion, it is natural to regard the set of states, operations, and measurements with non-negative Wigner functions as being *classical*, and the amount of negativity in the Wigner function as a measure of nonclassicality. One can make several arguments for this (see [48] for further discussion); here we will focus on one line of reasoning motivated by complexity considerations.

B. Wigner negativity as stabilizer complexity

A direct link between positivity of the Wigner function and stabilizer quantum computation is provided by the *Gottesman-Knill theorem* [32,33]. At a heuristic level, any quantum circuit that involves initializing to a Wigner positive state, quantum operations with positive Wigner functions, and measurements defined by POVMs with positive Wigner functions can be efficiently simulated on a classical computer [34,36]. The underlying intuition is that any quantum circuit constructed out of Wigner positive elements reduces to classical stochastic (Markovian) evolution in terms of the Wigner function representation—see Eq. (12), where now all the Wigner functions are genuine probability distributions. As long as each of these Wigner probability distributions can be efficiently sampled, then we can also efficiently sample the output probability distribution of the quantum circuit on a classical computer.⁴ Interestingly, one can generalize this argument to the case where circuit elements are not necessarily Wigner positive, and one finds that the amount of negativity in the circuit is directly connected to the classical simulation complexity [45,47].

From the above discussion, it is natural to interpret the negativity of the Wigner function as a measure of the complexity of classical simulation. One way to make this more concrete is in terms of the resource theory of stabilizer computation [36,37]. Pure states with positive Wigner

⁴Everywhere positive Wigner functions tend to be severely constrained and have very few free parameters, see for instance the characterization in [43] for pure Wigner positive states and Clifford unitaries. So, we expect that sampling from such distributions should be efficient, although this is not always guaranteed, especially for mixed states and quantum channels.

functions and convex combinations thereof are called *stabilizer states*.⁵ In the resource theory of stabilizer computation, stabilizer states are regarded as being “free of cost.” *Clifford unitaries*, which constitute a special subgroup of the unitary group that map stabilizer states to stabilizer states, are also regarded as being free.⁶ In addition, one is also allowed to bring in ancillas initialized to stabilizer states, perform joint Clifford unitaries, make measurements on the ancilla in the computational basis, and use conditioning on previous measurements or classical randomness; all these operations constitute “allowed” operations in stabilizer protocols. However, stabilizer protocols are *not* universal—one can think of them as generating a class of circuits that can be efficiently simulated classically. In order to achieve universal quantum computation, one needs access to additional *resource* states, sometimes referred to as “magic states” [35]. Much like entanglement plays the role of a resource in the theory of quantum communication with local operations and classical communication being treated as “free”, *magic* states play the role of a resource in the theory of stabilizer quantum computation, with stabilizer states and operations being treated as cheap [37]. Magic has found many applications in theoretical physics, such as in the context of conformal field theories [50], many-body systems [51–54], chaos and random circuits [55–62], and to some extent also in quantum gravity [48,63–65].

In order to quantify the notion of magic, one defines a *magic monotone* $\mathcal{M}$ as any real-valued, non-negative function on density matrices that vanishes on stabilizer states, and is, on average, nonincreasing under any stabilizer protocol [37]. By on average, we mean that if Λ is a stabilizer protocol potentially involving a measurement, and if under Λ we have $\rho \xrightarrow{\Lambda} \{p_i, \sigma_i\}$ (i.e., the outcome is the state σ_i with probability p_i), then

$$\mathcal{M}(\rho) \geq \sum_i p_i \mathcal{M}(\sigma_i). \quad (13)$$

Importantly, it may be the case that for some i , $\mathcal{M}(\rho) < \mathcal{M}(\sigma_i)$, as long as the corresponding probability is sufficiently small. It turns out that the *sum negativity* of the Wigner function, defined as,

$$\mathcal{N}_s(\rho) = \frac{1}{2} \sum_{q,p} (|W_\rho(q, p)| - 1), \quad (14)$$

⁵Usually stabilizer states are introduced as $+1$ eigenstates of a subgroup of Pauli operators, but the above definition turns out to be equivalent [43].

⁶In the discrete phase space formalism, Clifford unitaries correspond to symplectic affine transformations on the discrete phase space. Furthermore, for a system with n qudits, the most general Clifford unitary can be constructed out of $O(n \log d)$ elementary one and two-qudit gates [49].

is a magic monotone [37]. This gives Wigner negativity a direct operational meaning: if there exists a stabilizer protocol that converts an initial state ρ_0 to a final state ρ with probability p , then

$$p \leq \frac{\mathcal{N}_s(\rho_0)}{\mathcal{N}_s(\rho)}. \quad (15)$$

Thus, in order to obtain the desired state, one would need to repeat *any* such protocol at least $\frac{\mathcal{N}_s(\rho)}{\mathcal{N}_s(\rho_0)}$ number of times. Equivalently, one must start with a ρ_0 that comprises order $\log \mathcal{N}_s(\rho)$ number of copies of a resource state that has some $O(1)$ negativity in order to have a significant success probability for conversion to ρ . In this sense, one may regard Wigner negativity as a measure of “stabilizer complexity”, i.e., the minimal, unavoidable complexity in preparing ρ through a stabilizer protocol. Alternatively, one could think of Wigner negativity as a measure of the nonstabilizer content of the state, or “how far” the state is from being a stabilizer state. Since stabilizer circuits can be simulated efficiently on a classical computer, yet another interpretation of Wigner negativity is as a measure of the complexity of classical simulation using the Wigner quasiprobability representation (see [45] for more details). This last point requires some care, especially in the presence of a local tensor product structure—in that case, one can have states that have large Wigner negativity and yet are classically simulable using other (i.e., non-quasi-probability) methods, on account of low entanglement [66].⁷ Thus, Wigner negativity can be thought of as an operationally meaningful notion of complexity from these various points of view. We should emphasize that by complexity, we do not mean circuit complexity [67–77], although it would be interesting to study the connection to circuit complexity further.

In this paper we will focus on the quantity

$$\mathcal{N}(\rho) := 1 + 2\mathcal{N}_s(\rho) = \sum_{q,p} |W_\rho(q, p)|, \quad (16)$$

and we will refer to this quantity as the Wigner negativity. The negativity has several nice properties:

- (1) It is evident from the definition that $\mathcal{N} \geq 1$, where the lower bound is saturated if and only if the Wigner function is nowhere negative.
- (2) The negativity of a state ρ is also bounded above by $\sqrt{\frac{D}{e^{S_2}}}$, where S_2 is the second Rényi entropy of ρ . To see this, we note that

$$\sum_{q,p} W^2(q, p) = \frac{1}{D} \text{Tr}(\rho^2). \quad (17)$$

⁷We thank Alex May for bringing this to our attention.

Together with Jensen's inequality and the fact that $\text{Tr}\rho^2 = e^{-S_2}$, this gives us the desired bound,

$$\begin{aligned} \sum_{q,p} |W(q,p)| &\leq D \sqrt{\sum_{q,p} W^2(q,p)} \\ &= \sqrt{D} \sqrt{\text{Tr}\rho^2} = \sqrt{D e^{-S_2}}. \end{aligned} \quad (18)$$

- (3) The negativity is convex, i.e., for some set of probabilities $\{p_i\}$,

$$\mathcal{N}\left(\sum_i p_i \rho_i\right) \leq \sum_i p_i \mathcal{N}(\rho_i). \quad (19)$$

- (4) Finally, the negativity gives a lower bound on a natural measure of decoherence [78] and spreading [79–81], namely the *half-Rényi diagonal entropy*: given a density matrix ρ and an orthonormal basis $\mathcal{B} = \{|i\rangle\}$, the half-Rényi diagonal entropy is defined as

$$H^{(1/2)}(\rho) = 2 \log \left(\sum_i \lambda_i^{1/2} \right), \quad \lambda_i = \langle i | \rho | i \rangle. \quad (20)$$

Then, the Wigner negativity in the basis $\mathcal{B}$ satisfies the lower bound,

$$\log \mathcal{N}(\rho) \leq H^{(1/2)}(\rho). \quad (21)$$

Actually, the Wigner function transforms covariantly under any Clifford unitary, and so the negativity remains invariant under Clifford unitary transformations. This immediately implies the stronger bound,

$$\log \mathcal{N}(\rho) \leq \min_{U \in \text{Cliff}} H^{(1/2)}(U \rho U^\dagger), \quad (22)$$

where the minimization is over Clifford unitaries. One can interpret this bound as saying that a state with some amount of negativity in its Wigner function must have the above minimal and unavoidable spreading in the wave function along *any* direction in phase space.

III. WIGNER NEGATIVITY IN THE PSSY MODEL

In this Section, we will compute the Wigner negativity for Hawking radiation in a toy model of black hole evaporation due to PSSY [7]. The emphasis in this Section will be on computing the Wigner negativity using the *gravitational path integral*. As we will see, the answer obtained from the gravitational path integral will be completely information theoretic in nature (i.e., independent of the choice of computational basis for the radiation) and matches with the formula for Wigner negativity in Haar random states obtained in [55].

A. Setup

The Euclidean action of JT gravity with end of the world (EOW)-branes is given by,

$$I = I_{\text{JT}} + \mu \int_{\text{brane}} ds, \quad (23)$$

$$\begin{aligned} I_{\text{JT}} = & -\frac{S_0}{2\pi} \left[\frac{1}{2} \int_{\mathcal{M}} \sqrt{g} R + \int_{\partial\mathcal{M}} \sqrt{h} K \right] \\ & - \left[\frac{1}{2} \int_{\mathcal{M}} \sqrt{g} \phi (R + 2) + \int_{\partial\mathcal{M}} \sqrt{h} \phi (K - 1) \right], \end{aligned} \quad (24)$$

with the boundary conditions,

$$ds^2|_{\partial\mathcal{M}} = \frac{1}{\epsilon^2} d\tau^2, \quad \phi|_{\partial\mathcal{M}} = \frac{\phi_b}{\epsilon}, \quad \epsilon \rightarrow 0. \quad (25)$$

The boundary condition on the EOW-brane is given by

$$\partial_n \phi = \mu \geq 0, \quad K = 0. \quad (26)$$

We can regard S_0 as the entropy of the higher-dimensional extremal black hole for which the JT theory is the two-dimensional reduction, or it can be thought of as the ground state entropy of the JT system. The EOW-brane hosts semiclassical, intrinsic degrees of freedom, which we label as i, j etc., running from 1 to D . These intrinsic EOW-brane states are taken to be orthogonal in the bulk semiclassical description, i.e., these indices along the brane are contracted with the propagator δ_{ij} . We may regard these internal states of the EOW-brane as a toy model for bulk matter.

JT gravity in asymptotically AdS_2 spacetimes is dual to a quantum mechanical theory $\mathcal{B}$ on the asymptotic boundary—the path integral of JT gravity computes ensemble averaged quantities from the boundary point of view [82], where the averaging is over an ensemble of Hamiltonians. More precisely, JT gravity is dual to a double-scaled random matrix theory, with the leading contribution to the density of states given by,

$$\rho(s) = \frac{s}{2\pi^2} \sinh(2\pi s), \quad E = \frac{s^2}{2}. \quad (27)$$

The main object of interest for us in this paper is a stationary black hole solution in JT gravity with an EOW-brane capping off the geometry behind the black hole horizon. This black hole solution describes the semiclassical bulk dual to the following excited state in the dual quantum mechanical boundary theory:

$$|\psi_i\rangle = \sum_a 2^{\frac{1}{2}-\mu} \Gamma\left(\mu - \frac{1}{2} + i\sqrt{2E_a}\right) e^{-\frac{\beta}{2}E_a} C_{ia} |E_a\rangle, \quad (28)$$

where the coefficients C_{ia} are independent and identically distributed complex Gaussian random variables [7]. The

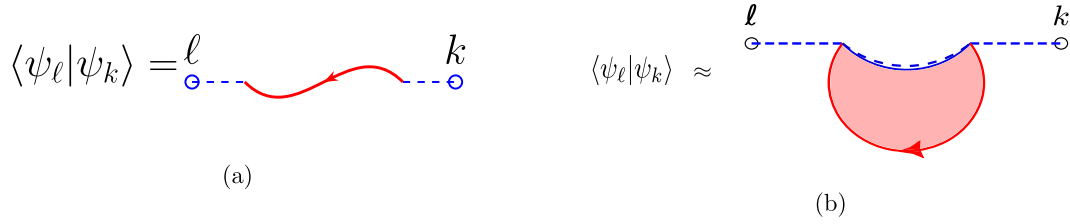


FIG. 1. (a) Boundary condition to evaluate the overlap $\langle \psi_\ell | \psi_k \rangle$: The solid line represents the asymptotic boundary of Euclidean AdS_2 with renormalized length β . The arrow indicates the direction of time from ket to bra. Indices attached to the dashed lines represent the states of the EOW-branes. The boundary condition suggests that EOW-branes of type ℓ or k should intersect the asymptotic AdS_2 boundary at the appropriate endpoints. (b) The bulk gravity path integral corresponding to the overlap, subject to the boundary conditions illustrated on the left.

states $|\psi_i\rangle$ are approximately orthogonal. The overlap $\langle \psi_\ell | \psi_k \rangle$, or more precisely, its ensemble average, is given by the gravity path integral (see Fig. 1),

$$\overline{\langle \psi_i | \psi_j \rangle} = e^{S_0} Z_1 \delta_{ij}, \quad (29)$$

where Z_1 is the disc partition function of JT gravity with one asymptotic boundary and one EOW-brane,

$$Z_1 = \int \rho(s) y(s), \quad \rho(s) = \frac{s}{2\pi^2} \sinh(2\pi s),$$

$$y(s) = e^{-\frac{\beta s^2}{2}} 2^{1-2\mu} \left| \Gamma\left(\mu - \frac{1}{2} + is\right) \right|^2. \quad (30)$$

More generally, the JT gravity path integral on the disc topology with n asymptotic boundaries of length β and n EOW-branes is given by $e^{S_0} Z_n$, where

$$Z_n = \int \rho(s) y(s)^n, \quad \rho(s) = \frac{s}{2\pi^2} \sinh(2\pi s),$$

$$y(s) = e^{-\frac{\beta s^2}{2}} 2^{1-2\mu} \left| \Gamma\left(\mu - \frac{1}{2} + is\right) \right|^2. \quad (31)$$

In order to model an evaporating scenario, one entangles the black hole system B with a quantum mechanical bath/reservoir R [7],

$$|\Psi\rangle = \frac{1}{\sqrt{D e^{S_0} Z_1}} \sum_{i=1}^D |\psi_i\rangle_B \otimes |i\rangle_R, \quad (32)$$

where the states $\{|i\rangle\}$ in R are orthonormal, and following [7], we have chosen the normalization factor to ensure that,

$$\overline{\langle \Psi | \Psi \rangle} = 1. \quad (33)$$

Equation (32) should be thought of as the late-time equilibrium state we get by coupling the black hole to a bath of Hilbert space dimension D through random interaction terms (without any conservation laws), and

waiting long enough till the system settles down. The reduced density matrix for the bath is given by,

$$\rho_R = \frac{1}{D e^{S_0} Z_1} \sum_{i,j=1}^D \langle \psi_i | \psi_j \rangle_B |j\rangle \langle i|_R. \quad (34)$$

For small D , one can treat the states $\{|\psi_i\rangle\}$ as being approximately orthogonal. This leads to a maximally mixed state on the bath whose entropy grows as $\log D$. However, when D becomes $O(e^{S_0})$, one can no longer treat the states $\{|\psi_i\rangle\}$ as being orthogonal. Indeed, it was shown in [7] that in the replica trick computation of the radiation entropy, a new gravitational saddle, i.e., the replica wormhole, makes the dominant contribution past the Page point [7]. This cuts off the naïve growth of the entropy with D , thus realizing the expected Page curve.

Our goal in this Section is to compute the Wigner negativity of the reduced density matrix ρ_R , viewed as a measure of the stabilizer complexity, or equivalently, the complexity of classically simulating the state of the Hawking radiation. In doing so, we will use the rules of the gravitational path integral. In order to compute the Wigner function, we must choose an orthonormal basis for the bath Hilbert space. Let us first consider the radiation basis $\{|i\rangle\}$ that appears in Eq. (32). The Wigner function with respect to this basis is given by⁸:

$$W(q, p) = \frac{1}{D^2 e^{S_0} Z_1} \sum_{k,\ell=0}^{D-1} \hat{\delta}_{2q,k+\ell} e^{-\frac{i2\pi p}{D}(k-\ell)} \langle \psi_\ell | \psi_k \rangle_B$$

$$= \frac{1}{D^2 e^{S_0} Z_1} \sum_{k,\ell=0}^{D-1} A_{k\ell} \langle \psi_\ell | \psi_k \rangle_B, \quad (35)$$

⁸Here we are taking D to be an odd prime for convenience, and no further structure on the radiation Hilbert space is assumed. If the radiation Hilbert space has some locality structure—for instance, if it consists of a tensor product of n local qudits (for some odd prime d)—then one can also use the Cartesian product phase space and its corresponding Wigner function [38,43]. Many of our conclusions will remain unaffected even in that case, with $D = d^n$.

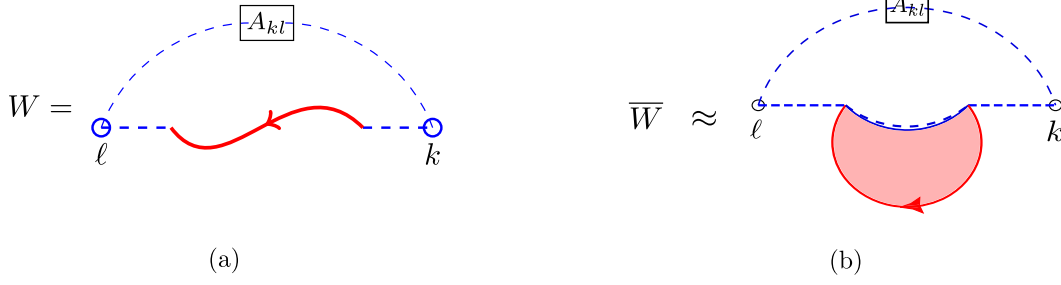


FIG. 2. (a) Boundary conditions for computing the Wigner function W . (b) The gravitational path integral with the given boundary conditions: the same EOW-brane gives δ_{kl} , and so the diagram evaluates to $e^{S_0} Z_1 \text{Tr}(A)$.

where $A_{k\ell} = \langle \ell | A(q, p) | k \rangle$. Any other basis $\{|\tilde{i}\rangle\}$ can be written in terms of the above radiation basis as $|\tilde{i}\rangle = \sum_{i'} U_{i'i'} |i'\rangle$, for some unitary matrix U . The corresponding Wigner function is given by

$$W_U(q, p) = \frac{1}{D^2 e^{S_0} Z_1} \sum_{k, \ell, k', \ell'} A_{k\ell}^{(U)} \langle \psi_\ell | \psi_k \rangle_B, \quad (36)$$

$$A_{k\ell}^{(U)} = U_{kk'} A_{k'\ell'} U_{\ell'\ell}^\dagger.$$

As we will see, the negativity of the above Wigner function computed using the gravitational path integral only involves traces of products of the form $\text{Tr}[(A^{(U)})^n]$, and so the unitary U cancels out. This means that the answer for the negativity is *independent* of the choice of basis. Two remarks:

- (1) In general, the Wigner negativity very much depends on the choice of basis. However, in our gravity calculation, the negativity seems to have a universal basis independent form, which can then only depend on information theoretic properties of the boundary state ρ_R . This essentially happens because the gravitational path integral computes an ensemble average over Hamiltonians and pseudorandom boundary conditions in the dual theory. As we will see later, this same universal (i.e., basis-independent) information-theoretic form for the Wigner negativity also appears more generally in chaotic systems with random matrix properties.
- (2) We have assumed that the basis for the bath is chosen independently, without prior knowledge of the microscopic details of the black hole. For instance, had we fine-tuned the basis in R to involve microscopic information—such as BH microstate energies, or the specific coefficients C_{ia} in Eq. (28)—then the above analysis would not be valid, because one would need to account for additional Euclidean wormholes emanating from the various U_{ij} matrices.

From Eq. (35), we see that the boundary conditions required to evaluate the Wigner function can be represented diagrammatically as in Fig. 2(a), where the phase point

operators are appended between the two terminal nodes. We can compute the ensemble average of the Wigner function $\overline{W}(q, p)$ by using the gravity path integral with these specified boundary conditions, as shown in Fig. 2(b). This just gives,

$$\begin{aligned} \overline{W}(q, p) &= \frac{1}{D^2 e^{S_0} Z_1} \sum_{k, \ell=0}^{D-1} A_{k\ell} e^{S_0} Z_1 \delta_{k,\ell} \\ &= \frac{1}{D^2} \text{Tr}(A) = \frac{1}{D^2}. \end{aligned} \quad (37)$$

So, on average, the Wigner function looks uniform and positive. However, this does not mean that the state of the Hawking radiation is simple—what we actually want to compute is the average of the Wigner negativity, and as we will see in the next Section, the negativity of the averaged Wigner function is not always the same as the averaged Wigner negativity. Indeed, we will find that after Page time, the Wigner function has large Wigner negativity; it is just that the averaging in Eq. (37) washes out this fine-grained information.

B. First look at Wigner negativity

Building on the above discussion, our goal is to compute the ensemble averaged Wigner negativity of the Hawking radiation using the gravitational path integral,

$$\bar{\mathcal{N}} = \sum_{q, p} |\overline{W}(q, p)|, \quad (38)$$

where note that the absolute value is inside the average. Indeed, the absolute value in the above formula creates trouble—while calculating the ensemble average of the Wigner function and its powers is straightforward using the gravity path integral, the absolute value is a nonanalytic function. In the next Section, we will present a systematic way to deal with this difficulty using an integral representation of the absolute value function. But to develop some rough intuition, we will first employ the following somewhat heuristic *replica trick*: we first evaluate the ensemble average over W^{2n} for integer n , and analytically

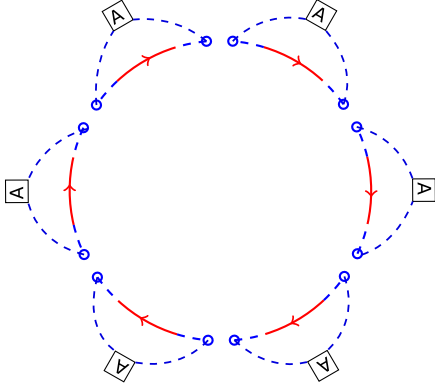


FIG. 3. Schematic diagram representing the boundary conditions used to evaluate W^{2n} , shown here for $n = 3$. Six copies of Fig. 2(a) are arranged in a circular layout, with each A denoting the insertion of a phase point operator. A gravitational path integral with these boundary conditions generally admits multiple saddle point solutions, corresponding to different ways of filling in the bulk region.

continue the result to $n = \frac{1}{2}$ in order to extract the average of the absolute value of W ,

$$\lim_{n \rightarrow \frac{1}{2}} \sum_{q,p} \overline{W^{2n}(q,p)}. \quad (39)$$

The point is that it is much easier to calculate W^{2n} using the gravitational path integral. In order to do so, we set up the appropriate boundary conditions corresponding to W^{2n} (see Fig. 3), and then let gravity fill in the bulk in all possible ways consistent with these boundary conditions. A key point is that for large values of n , multiple geometries contribute to $\overline{W^{2n}}$, and resumming the contributions from all such saddles becomes a somewhat nontrivial task. We will come back to this in the next Section, but to gain some initial insight, we will begin with a simplified analysis in

this Subsection. We will consider the two extreme limits $D \ll e^{S_0}$ and $D \gg e^{S_0}$, where the gravity path integral and subsequent analytic continuation can be performed straightforwardly. This approach will help capture the essential features of the Wigner negativity across the Page transition.

First, consider the regime $D \ll e^{S_0}$. In this limit, the dominant contribution comes from the completely disconnected diagram, shown in Fig. 4(a). All other diagrams come with lower powers of e^{S_0} , and are suppressed when $D \ll e^{S_0}$. This gives,

$$\overline{W^{2n}(q,p)} \approx (\overline{W(q,p)})^{2n} = \frac{1}{D^{4n}} \cdots (D \ll e^{S_0}). \quad (40)$$

Upon analytic continuation to $n = \frac{1}{2}$, we find that the averaged Wigner negativity is given by

$$\overline{\mathcal{N}} \approx 1, \cdots (D \ll e^{S_0}). \quad (41)$$

Thus, the state of the Hawking radiation well before the Page transition has vanishing stabilizer complexity [recall from Eq. (16) that $\tilde{\mathcal{N}}_s = \frac{1}{2}(\tilde{\mathcal{N}} - 1)$].

Next, we consider the opposite limit, where $D \gg e^{S_0}$. In this limit, the dominant diagrams are given by pairwise connected geometries, as illustrated in Fig. 4(b). The reason such pairwise connected diagrams become important is as follows: while such diagrams are subleading in e^{S_0} compared to the fully disconnected diagram, the EOW-brane index contractions for such diagrams give an enhancement at large D coming from the fact that $\text{Tr}(A^2) = D$ [see Eq. (50)]E. For $2n$ boundaries, such pairwise-connected configurations can be constructed in $\frac{(2n)!}{2^n n!}$ distinct ways. Contributions from other geometries (with three and higher boundary connections) are subleading and can be ignored. This leads to the following expression:

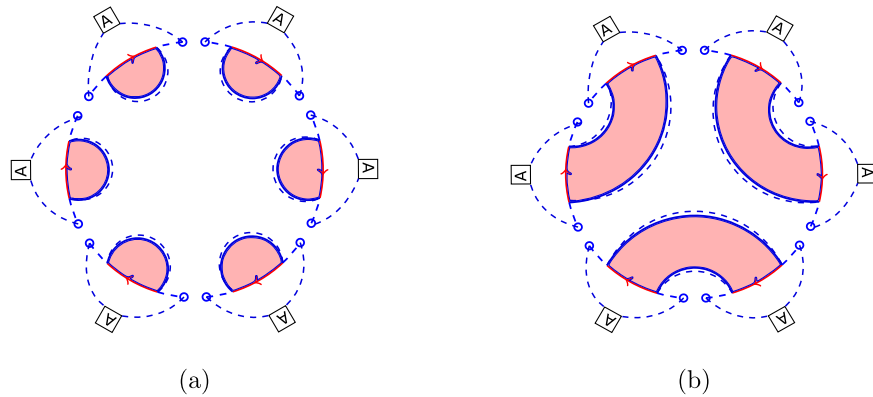


FIG. 4. Replica trick for $n = 3$. (a) In the regime $D \ll e^{S_0}$, the fully disconnected geometry dominates the gravitational path integral. Each disconnected piece contributes $e^{S_0} Z_1 \text{Tr}(A)$, leading to a total contribution of $(e^{S_0} Z_1)^6$. (b) In contrast, for $D \gg e^{S_0}$, the dominant saddle is a pairwise connected geometry. Each connected component connects two asymptotic boundaries. Each pair gives a factor $\text{Tr}(A^2) = D$ from the sum over EOW-brane indices, and the full diagram evaluates to $(D e^{S_0} Z_2)^3$.

$$\begin{aligned} \overline{W^{2n}(q, p)} &\approx \frac{(2n)!}{2^n n!} \frac{(e^{S_0} Z_2 \text{Tr}(A^2))^n}{(e^{S_0} Z_1 D^2)^{2n}} \\ &= \frac{(2n)!}{2^n n!} \left(\frac{Z_2}{Z_1^2}\right)^n \frac{1}{e^{nS_0} D^{3n}} \cdots (D \gg e^{S_0}). \end{aligned} \quad (42)$$

Analytically continuing this expression to $n \rightarrow \frac{1}{2}$, we find that for $D \gg e^{S_0}$, the averaged negativity is given by

$$\overline{\mathcal{N}} \approx \sqrt{\frac{2}{\pi}} \exp \left[\frac{1}{2} (S_{\max} - S_2) \right], \cdots (D \gg e^{S_0}), \quad (43)$$

where S_2 is the second Rényi entropy of the radiation post the Page transition

$$S_2 = S_0 - \log \left(\frac{Z_2}{Z_1^2} \right), \quad (44)$$

and $S_{\max} = \log D$ is the coarse-grained entropy, or equivalently the entropy of the maximally mixed state on R . Thus, we see that after the Page point, the Wigner negativity becomes exponentially large despite the averaged Wigner function being uniform and positive, as in Eq. (37). So, the state of the Hawking radiation post the Page transition has an exponentially large stabilizer complexity. It is interesting to note that this complexity is given universally—i.e., in a basis independent way—in terms of information theoretic quantities.

The above simple exercise already reveals the asymptotic behavior of the Wigner negativity as a function of the ratio D/e^{S_0} . However, when D and e^{S_0} are of comparable magnitude, one must, in principle, account for other diagrams beyond the fully disconnected and the pairwise connected ones in the gravitational path integral. This makes the analytic continuation nontrivial. To capture the intermediate behavior of the Wigner negativity in this regime, a more careful analysis is required, which we turn to in the next Subsection.

C. Integral representation method

A more systematic approach to deal with the absolute value in calculating the ensemble average of the Wigner negativity is to use the following integral representation:

$$\begin{aligned} |W| &= W \text{sgn}(W) = \lim_{\epsilon \rightarrow 0} \int_{-\infty}^{\infty} \frac{dz}{2\pi i} \frac{2z}{z^2 + \epsilon^2} W \exp(izW) \\ &= \lim_{\epsilon \rightarrow 0} \int_{-\infty}^{\infty} \frac{dz}{2\pi i} \frac{-2iz}{z^2 + \epsilon^2} \frac{d}{dz} \exp(izW). \end{aligned} \quad (45)$$

In order to avoid cumbersome notation, we will simply denote the Wigner function as W , leaving the dependence on (q, p) implicit, but it should be understood that the following analysis works pointwise in phase space. The

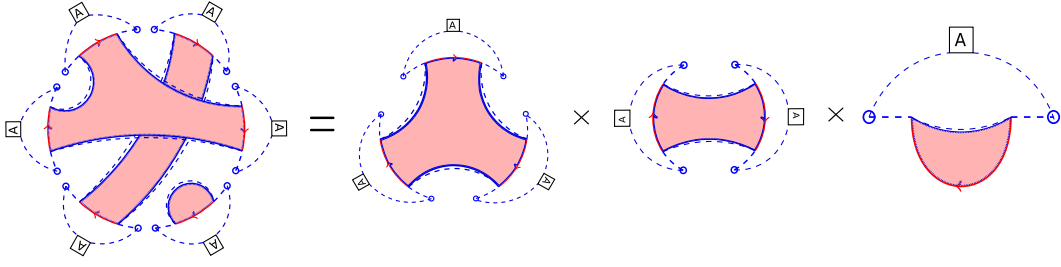
above integral representation is useful because it reduces the calculation of the ensemble average over the nonanalytic function $|W|$ to the ensemble average of an analytic function, namely e^{izW} . Expanding out the exponential and using Eq. (35), the ensemble average of e^{izW} is given by,

$$\begin{aligned} \overline{e^{izW}} &= \sum_{n=1}^{\infty} \frac{(iz)^n}{n!} \overline{W^n} \\ &= \sum_{n=1}^{\infty} \frac{1}{n!} \left(\frac{iz}{D^2 e^{S_0} Z_1} \right)^n \left(\sum_{k, \ell=0}^{D-1} A_{k\ell} \langle \psi_{\ell} | \psi_k \rangle_B \right)^n. \end{aligned} \quad (46)$$

So, we need to compute $\overline{\prod_{i=1}^n \langle \psi_{\ell_i} | \psi_{k_i} \rangle}$. Each overlap $\langle \psi_{\ell} | \psi_k \rangle$ corresponds to one asymptotic boundary. The bulk dual to the ensemble average corresponds to a sum over all possible ways in which the n boundaries can be filled in by a gravitational bulk. In the present case, it suffices to only consider bulk geometries topologically corresponding to multiple discs; geometries with handles are suppressed by powers of e^{-S_0} . Furthermore, since the index contractions are all determined by the structure of contractions of the EOW-branes, handles do not modify the invariant we get from contractions of the phase-point operators, and so, do not give any additional enhancements. For example, for $n = 3$, we have the following possibilities: For arbitrary n , the quantity $\overline{W^n}$ receives contributions from a series of such terms, corresponding to various bulk geometries with multiple discs. A key observation is that *any* such term can be decomposed into a product of irreducible building blocks. We illustrate this structure diagrammatically with a representative example in Fig. 6. For a configuration with n boundaries, any generic term can be expressed as a product of p_1 one-boundary contractions (i.e., p_1 discs, each with one asymptotic boundary and one EOW-brane), p_2 two-boundary contractions (i.e., p_2 discs, each with two asymptotic boundaries and two EOW-branes), p_3 three-boundary contractions (i.e., p_3 discs, each with three asymptotic boundaries and three EOW-branes), and so on, up to p_n n -boundary contractions, subject to the constraint $\sum_{j=1}^n j p_j = n$. Here, each j -boundary contraction is assumed to be irreducible, i.e., fully connected.

We can now organize the sum over saddles as a sum over products of “connected components” where each connected component corresponds to boundaries sharing a common gravitational bulk. For $n = 3$, this is shown in Fig. 5. After multiplying with the weights $A_{k\ell}$ and summing over all the indices, we get⁹:

⁹The combinatorial factor comes about as follows: there are a total of $\frac{n!}{p_1!(2p_2)!\cdots(np_n)!}$ ways of partitioning the n boundaries into p_1 single components, p_2 pairs, p_3 triples, and so on. Furthermore, the number of ways of grouping $j p_j$ boundaries into p_j groups of j boundaries each is given by $\frac{(j p_j)!}{p_j! (j!)^{p_j}}$.


 FIG. 5. All possible geometries (at disk level) contributing to the gravitational path integral for $\overline{W^3}$.

$$\overline{W^n} = \left(\frac{1}{D^2 e^{S_0} Z_1} \right)^n \sum_{p_1 \dots p_n} \delta_{\sum_r r p_r, n} \frac{n!}{p_1! \dots p_n!} \prod_{j=1}^n \left(\frac{D_j}{j!} \right)^{p_j}. \quad (47)$$

Here D_j is a JT gravity path integral corresponding to a fully connected diagram with j asymptotic boundaries and j EOW-branes,

$$D_j = (j-1)! e^{S_0} Z_j \text{Tr}(A^j), \quad (48)$$

where the $(j-1)!$ is a combinatorial factor that counts the number of possible ways to fully contract the j asymptotic boundaries. It is easy to evaluate $\text{Tr}(A^k)$ by repeatedly using the formula [38]

$$\begin{aligned} A(q_1, p_1) A(q_2, p_2) &= \sum_{q_3, p_3} \Gamma(q_1, p_1, q_2, p_2 | q_3, p_3) A(q_3, p_3), \\ \Gamma(q_1, p_1, q_2, p_2 | q_3, p_3) &= \frac{1}{D} \exp \left[\frac{4\pi i}{D} (p_1(q_3 - q_2) + p_2(q_1 - q_3) + p_3(q_2 - q_1)) \right], \end{aligned} \quad (49)$$

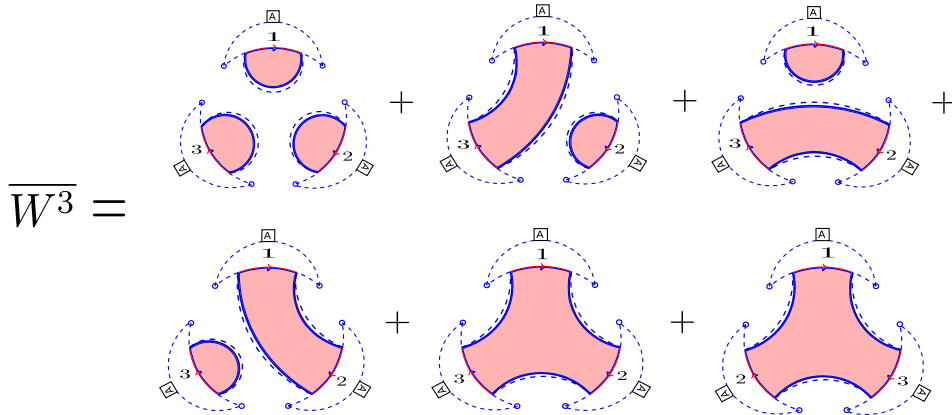
and this gives

$$\text{Tr}(A^k) = \begin{cases} 1 & \text{if } k \text{ is odd,} \\ D & \text{if } k \text{ is even.} \end{cases} \quad (50)$$

Since the only dependence on the phase space variables (q, p) is through $\text{Tr}(A^j)$, we see that $|\overline{W}|$ is independent of (q, p) .

Going back to Eq. (46), we can now perform the sum over n exactly. The sum over n effectively decouples the sums over p_j , allowing each to be evaluated independently, and we get

$$\begin{aligned} e^{iz\overline{W}} &= \sum_{n=0}^{\infty} \frac{(iz)^n}{n!} \overline{W^n} \\ &= \sum_{n=0}^{\infty} \left(\frac{iz}{D^2 e^{S_0} Z_1} \right)^n \frac{1}{n!} \sum_{p_1 \dots p_n} \delta_{\sum_r r p_r, n} \frac{n!}{p_1! \dots p_n!} \\ &\quad \times \prod_{j=1}^n \left(\frac{D_j}{j!} \right)^{p_j} \\ &= \prod_{j=1}^{\infty} \left\{ \sum_{p_j=0}^{\infty} \frac{1}{p_j!} \left[\left(\frac{iz}{D^2 e^{S_0} Z_1} \right)^j \frac{D_j}{j!} \right]^{p_j} \right\} \\ &= \exp \left(\sum_{j=1}^{\infty} \frac{(iz/D^2 e^{S_0} Z_1)^j}{j!} D_j \right). \end{aligned} \quad (51)$$


 FIG. 6. Generic saddle that arises in the computation of $\overline{W^6}$. It can be decomposed into three irreducible components: a three-boundary contraction, a two-boundary contraction, and a one-boundary contraction.

Using the explicit form of gravitational amplitudes D_j , we get

$$\overline{e^{izW}} = \exp \left[e^{S_0} \left\{ \sum_{m=1}^{\infty} \left(\frac{iz}{D^2 e^{S_0} Z_1} \right)^{2m-1} \frac{Z_{2m-1}}{(2m-1)} + D \sum_{m=1}^{\infty} \left(\frac{iz}{D^2 e^{S_0} Z_1} \right)^{2m} \frac{Z_{2m}}{2m} \right\} \right]. \quad (52)$$

Now, we can insert the above expression in Eq. (45) to calculate the ensemble averaged Wigner negativity,

$$|\overline{W}| = \lim_{\epsilon \rightarrow 0} \int_{-\infty}^{\infty} \frac{dz}{2\pi i} \frac{-2iz}{z^2 + \epsilon^2} \frac{d}{dz} (\overline{\exp(izW)}). \quad (53)$$

It is convenient to define the rescaled integration variable $t = \frac{z}{De^{S_2}}$, where S_2 is defined in Eq. (44). In terms of this rescaled integration variable, we get,

$$|\overline{W}| = \lim_{\epsilon \rightarrow 0} \frac{1}{De^{S_2}} \int_{-\infty}^{\infty} \frac{dt}{2\pi i} \frac{-2it}{t^2 + \epsilon^2} \frac{d}{dt} (\overline{\exp(iDe^{S_2}tW)}), \quad (54)$$

where $\epsilon = \frac{\epsilon}{De^{S_2}}$. To calculate the integral exactly, we need to know the partition functions Z_j for all $j > 0$. However, the integral simplifies if we take the large D and large $\exp(S_0)$ limit with the ratio e^{S_0}/D fixed. In this limit, we see that the exponent simplifies,

$$|\overline{W}| = \lim_{\epsilon \rightarrow 0} \frac{1}{De^{S_2}} \int_{-\infty}^{\infty} dt \frac{-2it}{t^2 + \epsilon^2} \frac{d}{dt} \left[\frac{e^{S_2}}{D} \left(it - \frac{t^2}{2} \right) \right] \times \exp \left[\frac{e^{S_2}}{D} \left\{ it - \frac{t^2}{2} \right\} + O\left(\frac{1}{D^2}\right) \right].$$

In this limit, terms in the exponential with powers of t^3 and higher are suppressed by a factor of $\frac{1}{D^2}$, so these terms can be dropped at large D . Thus, the integral reduces to,

$$|\overline{W}| = \lim_{\epsilon \rightarrow 0} \frac{1}{De^{S_2}} \int_{-\infty}^{\infty} dt \frac{-2it}{t^2 + \epsilon^2} \frac{d}{dt} \exp \left[\frac{e^{S_2}}{D} \left(it - \frac{t^2}{2} \right) \right]. \quad (55)$$

The above integral can be evaluated exactly (see Appendix A for details). This gives the following result for the ensemble average over the absolute value of the Wigner function,

$$|\overline{W}| = \frac{1}{D^2} \left(\frac{\exp(-r)}{\sqrt{\pi r}} + \text{erf}(\sqrt{r}) \right), \quad r = \frac{e^{S_2}}{2D}, \quad (56)$$

where S_2 is defined in Eq. (44) and

$$\text{erf}(z) = \frac{2}{\sqrt{\pi}} \int_0^z dt e^{-t^2}. \quad (57)$$

Since $|\overline{W}|$ does not depend on (q, p) , the ensemble averaged Wigner negativity is therefore given by,

$$\tilde{\mathcal{N}} = \text{erf}(\sqrt{r}) + \frac{\exp(-r)}{\sqrt{\pi r}}. \quad (58)$$

Note that when $r \rightarrow \infty$ (corresponding to $D \ll e^{S_2}$), $\tilde{\mathcal{N}} \sim 1$, while when $r \rightarrow 0$ (corresponding to $D \gg e^{S_2}$), then $\tilde{\mathcal{N}} \sim \sqrt{\frac{2D}{\pi e^{S_2}}}$, which is the result we obtained from the replica trick calculation in the previous Section. Equation (58) interpolates between these two extreme limits. Notably, the answer obtained from the gravitational path integral is completely information theoretic in nature, does not depend on the choice of computational basis, and is closely related to the formula for Wigner negativity in Haar random states [55]. In Appendix B, we argue that fluctuations around the ensemble average are suppressed in $\frac{1}{De^{S_0}}$. In Fig. 7, we compare our analytical formula for the Wigner negativity with direct numerical calculations for the microcanonical ensemble and find good agreement.

So, to summarize, the state of the Hawking radiation in the PSSY model has an $O(1)$ Wigner negativity before the Page transition, corresponding to low stabilizer complexity. Thus, before the Page transition, the state of the Hawking radiation is classically simulable. On the other hand, after the Page transition, the Hawking radiation has an exponentially large Wigner negativity, and thus an exponentially large stabilizer complexity. In this case, the state of the Hawking radiation is hard to simulate classically. Interestingly, the radiation system contains the black hole interior in its entanglement wedge after the Page transition, so one interpretation is that the large complexity is associated with the black hole interior. We will return to this point in Sec. V.

IV. A DYNAMICAL MODEL OF BLACK HOLE EVAPORATION

In this Section, we study Wigner negativity in a dynamical model of black hole evaporation. We consider two quantum systems B (for black hole) and R (for radiation) coupled to each other. The Hilbert space dimensions of B and R are e^S and D , respectively. The subsystem B should be thought of as the quantum system dual to the black hole and the subsystem R should be identified with radiation bath. Our goal is to couple the black hole (in some pure state) to a “simple” pure state of the bath and study the growth of Wigner negativity of the bath system in real time. The calculation in this section can be regarded as a time-dependent generalization of the calculations in [55] (see also [48]).

A. Ensemble averaged negativity

We choose a computational basis $\mathcal{B} = \{|i_0\rangle\}$ for the bath. We start with a product state of the total system of the form $|\chi_0\rangle_B \otimes |\psi_0\rangle_R$, where ψ_0 is a pure, stabilizer (i.e.,

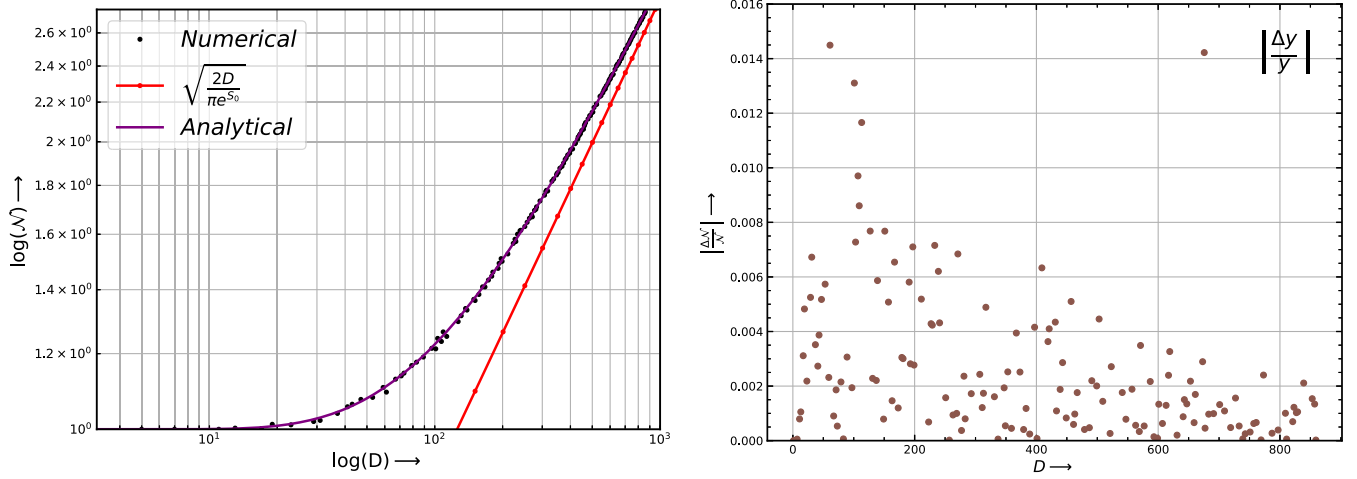


FIG. 7. (Left) Log-log plot of Wigner negativity for the microcanonical ensemble with the black hole Hilbert space dimension $e^{S_{\text{micro}}} = 80$. The black dots are numerically calculated values for the state in Eq. (28) for one particular choice of the C_{ia} , and where the energies are taken to be identical. (Right) Relative error between analytical formula and numerical data.

Wigner positive) state with respect to the computational basis, and χ_0 is a state of the B system which is dual to a black hole on the gravity side. The combined BR system is then evolved in time with some joint Hamiltonian H_{BR} . We will consider a joint Hamiltonian of the form,

$$H_{BR} = O_B \otimes O_R, \quad (59)$$

with the further assumption that O_R is a random matrix with respect to the computational basis on R , drawn from some unitarily invariant ensembles. We are interested in computing the growth of Wigner negativity of the time evolved density matrix of subsystem R ,

$$\rho_R(t) = \text{Tr}_B(e^{-iH_{BR}t} |\chi_0\rangle\langle\chi_0|_B \otimes |\psi_0\rangle\langle\psi_0|_R e^{iH_{BR}t}), \quad (60)$$

with respect to the computational basis. In general, it is hard to compute the growth of negativity for one particular choice of Hamiltonian. Our assumption that the operator O_R is a random matrix allows us to get some analytical handle on the problem using tools from random matrix theory. This way, instead of computing the Wigner negativity for one specific choice of Hamiltonian, we can compute the ensemble average,

$$\overline{\mathcal{N}(\rho_R(t))} = \int dH_{BR} \mu(H_{BR}) \mathcal{N}_{H_{BR}}(\rho_R(t)), \quad (61)$$

where by the average over H_{BR} , we really mean the average over O_R with some unitarily invariant measure μ .

In order to compute the negativity, we will follow the strategy explained in [48]. Let $\mathcal{B} = \{|i_0\rangle\}_{i=0}^{D-1}$ be the chosen computational basis on the bath. Without loss of generality, we can take the initial state ψ_0 of the bath to be the first basis vector $|0\rangle$ from the computational basis. The Wigner

function corresponding to $\rho_R(t)$ in the computational basis is given by,

$$W_{\rho_R, \mathcal{B}}(q, p) = \frac{1}{D} \sum_{k, \ell=0}^{D-1} e^{\frac{i2\pi}{D}(\ell-k)p} \hat{\delta}_{2q, k+\ell} \langle k_0 | \rho_R(t) | \ell_0 \rangle. \quad (62)$$

Now consider another basis: $U \cdot \mathcal{B} = \{|i\rangle\}_{i=0}^{D-1}$, where $|i\rangle = U|i_0\rangle$ with $U|\psi_0\rangle = |\psi_0\rangle$. This new basis is just a unitary rotation of the computational basis $\mathcal{B}$, which leaves the first basis vector $|\psi_0\rangle$ invariant. So, the unitary U acts on a $D-1$ dimensional subspace. We can write,

$$|i\rangle = \begin{cases} |\psi_0\rangle \cdots (i=0) \\ \sum_{j=1}^{D-1} U_{ij} |j_0\rangle \cdots (i \neq 0), \end{cases} \quad (63)$$

where $U_{ij} = \langle j_0 | U | i_0 \rangle$. Equivalently,

$$|i\rangle = \delta_{i,0} |\psi_0\rangle + \bar{\delta}_{i,0} \sum_{j=1}^{D-1} U_{ij} |j_0\rangle, \quad (64)$$

with the definition $\bar{\delta}_{m,n} = (1 - \delta_{m,n})$. The reason to introduce this rotated basis will become clear below. In the rotated basis, the Wigner function can be written as,

$$W_{\rho_R, U \cdot \mathcal{B}}(q, p) = \frac{1}{D} \sum_{k, \ell} e^{\frac{i2\pi}{D}(\ell-k)p} \hat{\delta}_{2q, k+\ell} \langle k_0 | U^\dagger \rho_R(t) U | \ell_0 \rangle. \quad (65)$$

Since U acts only on R and the partial trace is performed over subsystem B , we can write,

$$\begin{aligned}
U^\dagger \rho_R(t) U &= \text{Tr}_B(U^\dagger e^{-iH_{\text{BR}}t} |\chi_0\rangle \langle \chi_0|_B \otimes |\psi_0\rangle \langle \psi_0|_R e^{iH_{\text{BR}}t} U) \\
&= \text{Tr}_B(U^\dagger e^{-iH_{\text{BR}}t} |\chi_0\rangle \langle \chi_0|_B \otimes U |\psi_0\rangle \langle \psi_0|_R U^\dagger e^{iH_{\text{BR}}t} U) \\
&= \text{Tr}_B(U^\dagger e^{-iH_{\text{BR}}t} U |\chi_0\rangle \langle \chi_0|_B \otimes |\psi_0\rangle \langle \psi_0|_R U^\dagger e^{iH_{\text{BR}}t} U),
\end{aligned} \tag{66}$$

where in the second line we have used the fact $U|\psi_0\rangle = |\psi_0\rangle$. Consider now the ensemble averaged negativity with respect to the rotated basis $U \cdot \mathcal{B}$:

$$\overline{\mathcal{N}_{U \cdot \mathcal{B}}(\rho_R(t))} = \int dH_{\text{BR}} \mu(H_{\text{BR}}) \sum_{q,p} |W_{U \cdot \mathcal{B}}(q, p)|. \tag{67}$$

From Eq. (66), together with unitary invariance of the measure μ under $H_{\text{BR}} \rightarrow (1_B \otimes U) H_{\text{BR}} (1_B \otimes U^\dagger)$, we get,

$$\overline{\mathcal{N}_{U \cdot \mathcal{B}}(\rho_R(t))} = \overline{\mathcal{N}_{\mathcal{B}}(\rho_R(t))}. \tag{68}$$

So, the ensemble averaged negativity is invariant under unitary rotations of the computational basis that leave the initial state ψ_0 invariant. We can thus write the negativity in the computational basis in the following way:

$$\begin{aligned}
\overline{\mathcal{N}_{\mathcal{B}}(\rho_R(t))} &= \frac{1}{\text{Vol}_{U(D-1)}} \int dU \int dH_{\text{BR}} \mu(H_{\text{BR}}) \\
&\times \sum_{q,p} |W_{U \cdot \mathcal{B}}(q, p)|,
\end{aligned} \tag{69}$$

where we have by hand introduced an integral over U which cancels with the measure factor, given Eq. (68). We now switch the order of the integration and perform the Haar integration first,

$$\overline{\mathcal{N}_{\mathcal{B}}(\rho_R(t))} = \int dH_{\text{BR}} \mu(H_{\text{BR}}) \mathcal{N}_{\text{Haar}}, \tag{70}$$

where $\mathcal{N}_{\text{Haar}}$ is given by,

$$\mathcal{N}_{\text{Haar}} = \frac{1}{\text{Vol}_{U(D-1)}} \int dU \sum_{q,p} |W_{U \cdot \mathcal{B}}(q, p)|. \tag{71}$$

The quantity $\mathcal{N}_{\text{Haar}}$ can be interpreted as the Wigner negativity averaged over all possible basis transformations leaving the initial state fixed, and therefore it is an inherently information theoretic quantity. Our goal now is to compute $\mathcal{N}_{\text{Haar}}$ using the integral representation method discussed in the previous Section.

B. Integral representation method and Haar integration

As before, we will use the following integral representation to compute the Haar average over the absolute value of the Wigner function,

$$|W| = \lim_{\epsilon \rightarrow 0} \frac{1}{2\pi i} \int_{-\infty}^{\infty} dz \frac{2z}{z^2 + \epsilon^2} W e^{izW}. \tag{72}$$

From Eqs. (71) and (72), we get,

$$\begin{aligned}
\mathcal{N}_{\text{Haar}} &= \lim_{\epsilon \rightarrow 0} \frac{1}{2\pi i} \int_{-\infty}^{\infty} dz \frac{2z}{z^2 + \epsilon^2} \\
&\times \sum_{q,p} \frac{1}{\text{Vol}_{U(D-1)}} \int dU W_{U \cdot \mathcal{B}}(q, p) e^{izW_{U \cdot \mathcal{B}}(q, p)} \\
&= \sum_{q,p} \lim_{\epsilon \rightarrow 0} \frac{1}{2\pi i} \int_{-\infty}^{\infty} dz \frac{2z}{z^2 + \epsilon^2} \mathcal{I}'(q, p; z),
\end{aligned} \tag{73}$$

where

$$\mathcal{I}'(q, p; z) = \frac{1}{\text{Vol}_{U(D-1)}} \int dU W_{U \cdot \mathcal{B}}(q, p) e^{izW_{U \cdot \mathcal{B}}(q, p)}. \tag{74}$$

Our goal now is to evaluate $\mathcal{I}'(q, p; z)$ using the standard tools of Haar integration. To this end, we will expand out the exponential in powers of $W_{U \cdot \mathcal{B}}$, perform the average term by term, and re-sum the result in an appropriate limit.

In order to do this systematically, we will use the diagrammatic notation explained in [48]. We first write down the expression for the Wigner function in the basis $U \cdot \mathcal{B}$ [see Eq. (64)],

$$\begin{aligned}
W_{U \cdot \mathcal{B}}(q, p) &= \frac{1}{D} \sum_{i, i'=0}^{D-1} e^{\frac{i2\pi}{D}(i'-i)p} \hat{\delta}_{2q, i+i'} \langle i' | \rho_R(t) | i \rangle \\
&= \frac{1}{D} \sum_{i, i', j, j'=0}^{D-1} e^{\frac{i2\pi}{D}(i'-i)p} \hat{\delta}_{2q, i+i'} [\delta_{i',0} \delta_{i,0} S_0 \\
&\quad + \delta_{i',0} \bar{\delta}_{i,0} U_{ij} S_j + \bar{\delta}_{i',0} \delta_{i,0} U_{i'j'}^* S_{j'}^* \\
&\quad + \bar{\delta}_{i',0} \bar{\delta}_{i,0} U_{i'j'}^* U_{ij} S_{j'j}],
\end{aligned} \tag{75}$$

where we have defined the following quantities,

$$S_0 := \langle \psi_0 | \rho_R(t) | \psi_0 \rangle, \tag{76}$$

$$S_j := \langle \psi_0 | \rho_R(t) | j_0 \rangle, \tag{77}$$

$$S_{j'j} := \langle j'_0 | \rho_R(t) | j_0 \rangle. \tag{78}$$

It is convenient to diagrammatically represent each of the terms in Eq. (75) in order to facilitate the Haar integration over U . We will use the following notation:

$$\begin{aligned}
 \textcircled{\bullet} &= \frac{1}{D} \sum_{i,i'} e^{\frac{i2\pi}{D}(i'-i)p} \widehat{\delta}_{2q,i+i'} \delta_{i',0} \delta_{i,0} S_0 = \frac{1}{D} \delta_{q,0} S_0, \\
 \textcircled{\bullet} &= \frac{1}{D} \sum_{i,i',j} e^{\frac{i2\pi}{D}(i'-i)p} \widehat{\delta}_{2q,i+i'} \delta_{i',0} \bar{\delta}_{i,0} U_{ij} S_j, \\
 \textcircled{\bullet} &= \frac{1}{D} \sum_{i,i',j'} e^{\frac{i2\pi}{D}(i'-i)p} \widehat{\delta}_{2q,i+i'} \bar{\delta}_{i',0} \delta_{i,0} U_{i'j'}^* S_{j'}^*, \\
 \textcircled{\bullet} &= \frac{1}{D} \sum_{i,i',j,j'} e^{\frac{i2\pi}{D}(i'-i)p} \widehat{\delta}_{2q,i+i'} \bar{\delta}_{i',0} \bar{\delta}_{i,0} U_{i'j'}^* U_{ij} S_{j'j}.
 \end{aligned} \tag{79}$$

The convention is as follows: each copy of the Wigner function corresponds to a blob. For each blob, the empty dot corresponds to the unprimed index i (corresponding to the ket) and the shaded dot corresponds to the primed index i' (corresponding to the bra). No legs attached refers to the case when that particular index has been set to zero. The matrix elements U_{ij} are denoted by a couple of legs coming out of the (empty/shaded) dot. The solid line corresponds to the i index of the unitary matrix and the dotted line corresponds to the j index. Note that the dotted indices are appropriately contracted with either S_j , S_j^* or $S_{j'j}$. With this notation, Eq. (75) can be written as,

$$W_{U.B}(q, p) = \textcircled{\bullet} + \textcircled{\bullet} + \textcircled{\bullet} + \textcircled{\bullet}. \tag{80}$$

The utility of the diagrammatic notation is that it allows us to visualize Haar integration over W^n in terms of contractions of the various legs in these diagrams. As a quick warm-up, let us first compute the average over the Wigner function $W_{U.B}$. Using the formula,

$$\frac{1}{\text{Vol}_{U(D-1)}} \int dU U_{ij} U_{i'j'}^* = \frac{1}{D-1} \delta_{ii'} \delta_{jj'}, \tag{81}$$

this can be represented as,

$$\frac{1}{\text{Vol}_{U(D-1)}} \int dU W_{U.B} = \textcircled{\bullet} + \textcircled{\bullet} = \frac{1}{D} \delta_{q,0} S_0 + \frac{1}{D(D-1)} \bar{\delta}_{q,0} (1 - S_0), \tag{82}$$

where in the second equality, we have used $\sum_j S_{jj} = (1 - S_0)$. One can easily check that the averaged Wigner function is normalized, i.e., $\sum_{q,p} \overline{W_{U.B}} = 1$. Note that the quantity S_0 is the square of the *fidelity*¹⁰ between $\rho_R(t)$ and the initial state ψ_0 ,

$$S_0(t) = (F(\psi_0, \rho_R(t)))^2. \tag{83}$$

The fidelity is a measure of the distance between the two states, and in particular satisfies $0 \leq F(\rho, \sigma) \leq 1$. Thus, $0 \leq S_0 \leq 1$. From Eq. (82), we see that the Haar-averaged Wigner function, and consequently also the ensemble averaged

¹⁰Recall that the fidelity between two density matrices ρ and σ is defined as $F(\rho, \sigma) = \text{Tr}(\sqrt{\sqrt{\rho}\sigma\sqrt{\rho}})$. When one of the states is pure, this formula reduces to $F(\psi_0, \rho) = \sqrt{\langle \psi_0 | \rho | \psi_0 \rangle}$.

Wigner function $\bar{W}_B$ is non-negative everywhere. However, this does not mean that the Haar/ensemble averaged Wigner negativity vanishes, as we will now see.

C. Diagrammatics and resummation

In order to compute the negativity, we need to calculate the Haar integral $\mathcal{I}'(q, p; z)$ in Eq. (74). First note that,

$$\mathcal{I}'(q, p; z) = \frac{1}{\text{Vol}_{U(D-1)}} \int dU W_{U,B} e^{izW_{U,B}} = -i \frac{\partial}{\partial z} \mathcal{I}(q, p; z), \quad (84)$$

where

$$\mathcal{I}(q, p; z) = \frac{1}{\text{Vol}_{U(D-1)}} \int dU e^{izW_{U,B}}. \quad (85)$$

So, it will be sufficient for us to compute $\mathcal{I}(q, p; z)$ as a function of z . Using the diagrammatic expression for $W_{U,B}$ in Eq. (80), we can write,

$$\begin{aligned} \mathcal{I}(q, p; z) &= \frac{1}{\text{Vol}_{U(D-1)}} \int dU \exp \left\{ iz \left[\text{blob with no legs} + \text{blob with 1 leg} + \text{blob with 2 legs} + \text{blob with 3 legs} \right] \right\} \\ &= \exp \left[iz \text{blob with no legs} \right] \underbrace{\frac{1}{\text{Vol}_{U(D-1)}} \int dU \exp \left\{ iz \left[\text{blob with 1 leg} + \text{blob with 2 legs} + \text{blob with 3 legs} \right] \right\}}_{\mathcal{J}}. \end{aligned} \quad (86)$$

The first blob with no legs comes out of the integral since this piece is independent of U . To evaluate the rest of the integral $\mathcal{J}$, we can expand out the part of the exponential containing blobs with one pair of legs attached. Thus we get,

$$\begin{aligned} \mathcal{J} &= \frac{1}{\text{Vol}_{U(D-1)}} \int dU \exp \left[iz \text{blob with 1 leg} \right] \times \sum_{n=0}^{\infty} \frac{(iz)^n}{n!} \left[\text{blob with 2 legs} + \text{blob with 3 legs} \right]^n \\ &= \sum_{n=0}^{\infty} \frac{(iz)^n}{n!} \underbrace{\frac{1}{\text{Vol}_{U(D-1)}} \int dU \exp \left[iz \text{blob with 1 leg} \right]}_A \underbrace{\left[\text{blob with 2 legs} + \text{blob with 3 legs} \right]^n}_B. \end{aligned} \quad (87)$$

Note that we have not yet performed the Haar integration. We can now expand the exponential in A , after which the Haar integral is evaluated by combining the blobs in A and B in all possible ways and contracting the corresponding open legs. To keep the presentation concise, we quote only the final result of the Haar integration here. The details of the computation are deferred to Appendix C. As explained in the Appendix, the final result can be expressed in terms of fully connected n -blob contractions. There are two classes of such contractions. The first class consists of irreducible n -body contractions constructed from n blobs, where blobs of type B are always present, while blobs of type A may or may not be included. The open legs are then contracted in all possible ways such that the resulting diagram is fully connected. We denote the set of all such contractions by $\mathcal{D}'_n$. The second class consists of irreducible n -body contractions formed entirely from type- A blobs. Again, the open legs of the A blobs are contracted among themselves to produce a fully connected diagram. We denote this set by $\mathcal{D}_{n,0}$. We show a few examples below.

$$\mathcal{D}'_2 = \text{blob with 1 leg} \text{ --- } \text{blob with 2 legs} \quad (88)$$

$$\mathcal{D}'_3 = \text{diagram with three circles and arcs} \quad (89)$$

$$\mathcal{D}_{1,0} = \text{diagram with one circle and arc} \quad (90)$$

$$\mathcal{D}_{2,0} = \text{diagram 1} + \text{diagram 2} + \text{diagram 3} \quad (91)$$

We can now write down the final expression for $\mathcal{I}(q, p; z)$,

$$\mathcal{I}(q, p; z) = \exp\left(\sum_{n=1}^{\infty} (iz)^n \mathcal{D}_n\right), \quad (92)$$

where

$$\mathcal{D}_1 = \mathcal{D}_{1,0} + \text{diagram with one circle and dot} \quad (93)$$

$$\mathcal{D}_n = \frac{\mathcal{D}_{n,0}}{n!} + \mathcal{D}'_n \quad n = 2, 3, \dots \quad (94)$$

From Eqs. (73) and (C26), the Haar averaged Wigner negativity is given by,

$$\mathcal{N}_{\text{Haar}} = \lim_{\epsilon \rightarrow 0} \sum_{q,p} \frac{1}{2\pi i} \int_{-\infty}^{\infty} dz \frac{2z}{z^2 + \epsilon^2} \left(-i \frac{\partial}{\partial z}\right) e^{(\sum_{n=1}^{\infty} (iz)^n \mathcal{D}_n)}. \quad (95)$$

By redefining $s = \frac{2\mathcal{D}_2}{\mathcal{D}_1} z$, we can write this integral as

$$\begin{aligned} \mathcal{N}_{\text{Haar}} &= \lim_{\epsilon \rightarrow 0} \sum_{q,p} \frac{2\mathcal{D}_2}{2\pi i \mathcal{D}_1} \int_{-\infty}^{\infty} ds \frac{2s}{s^2 + (\frac{2\mathcal{D}_2 \epsilon}{\mathcal{D}_1})^2} \left(-i \frac{\partial}{\partial s}\right) e^{\sum_{n=1}^{\infty} (is)^n (\frac{\mathcal{D}_1}{2\mathcal{D}_2})^n \mathcal{D}_n} \\ &= \lim_{\epsilon \rightarrow 0} \sum_{q,p} \frac{2\mathcal{D}_2}{2\pi i \mathcal{D}_1} \int_{-\infty}^{\infty} ds \frac{2s}{s^2 + (\frac{2\mathcal{D}_2 \epsilon}{\mathcal{D}_1})^2} \left(-i \frac{\partial}{\partial s}\right) e^{\frac{\mathcal{D}_1^2}{4\mathcal{D}_2} (2is - s^2 - \frac{i\mathcal{D}_1 \mathcal{D}_3}{2\mathcal{D}_2^2} s^3 + \dots)}. \end{aligned} \quad (96)$$

In order to analyze this expression further, we now need to evaluate the diagrams $\mathcal{D}_n$ explicitly. Note that all the $\mathcal{D}_n$ can be expressed in terms of various information theoretic invariants built out of the reduced density matrix $\rho_R(t)$, its higher powers and $|\psi_0\rangle\langle\psi_0|$. Some of the first few invariants that will be relevant for us are (see Fig. 8),

$$S_0(t) = \langle\psi_0|\rho_R(t)|\psi_0\rangle, \quad (97)$$

$$g(t) = \langle\psi_0|\rho_R^2(t)|\psi_0\rangle, \quad (98)$$

$$S_2^R(t) = -\log \text{Tr} \rho_R^2(t), \quad (99)$$

where note that S_2^R is the second Rényi entropy.

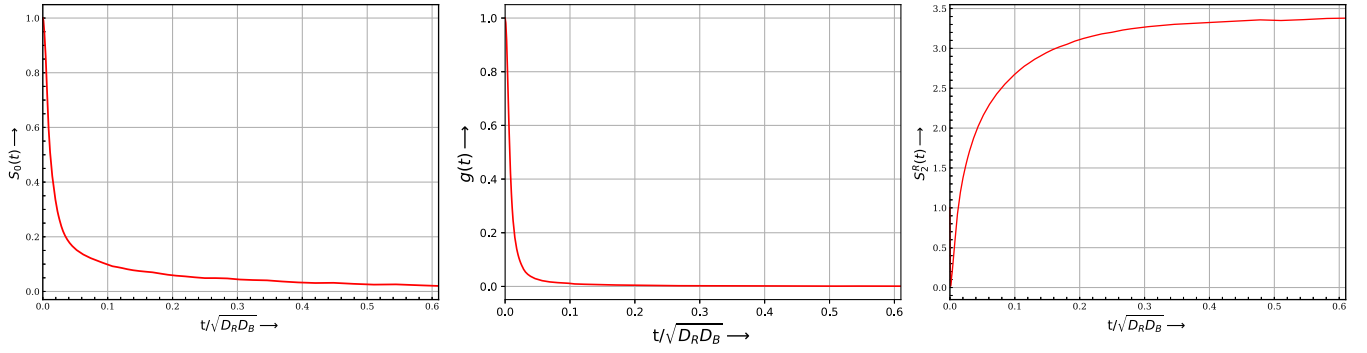


FIG. 8. Invariants S_0 , g , and S_2^R as functions of time for the situation where the operators O_B and O_R in the interaction term (59) are taken to be random matrices from the Gaussian Unitary Ensemble (GUE).

Since $\rho_R(t)$ evolves in time starting from a pure state, all the amplitudes $\mathcal{D}_n$ are time dependent. For instance, the first two amplitudes $\mathcal{D}_1$ and $\mathcal{D}_2$ are given by,

$$\mathcal{D}_1 = \text{diagram 1} + \text{diagram 2} \quad (100)$$

$$= \frac{1}{D^2} \left[(1 - S_0) \bar{\delta}_{q,0} + D S_0 \delta_{q,0} \right],$$

$$2 \mathcal{D}_2 = \text{diagram 3} + \text{diagram 4} + \text{diagram 5} + 2 \times \text{diagram 6} \quad (101)$$

$$= \frac{1}{D^4} (\alpha_0 \delta_{q,0} + \bar{\alpha}_0 \bar{\delta}_{q,0}),$$

where the quantities α_0 and $\bar{\alpha}_0$ are given by,

$$\alpha_0 = (D-1)(e^{-S_2^R} - 2g + S_0^2) - \left(1 - \frac{1}{D}\right)(1 - S_0)^2, \quad (102)$$

$$\bar{\alpha}_0 = \left(D - 2 - \frac{1}{D}\right)(e^{-S_2^R} - 2g + S_0^2) - \left(1 - \frac{2}{D}\right)(1 - S_0)^2 + 2D(g - S_0^2). \quad (103)$$

Similarly one can compute the higher point diagrams. In the above formulas, we have only kept the leading contribution in the Weingarten functions [see the discussion around Eq. (C6)]. We can further simplify the above

expressions at large D as,

$$\alpha_0 \simeq D(e^{-S_2^R} - 2g + S_0^2) - (1 - S_0)^2, \quad (104)$$

$$\bar{\alpha}_0 \simeq D(e^{-S_2^R} - S_0^2) - (1 - S_0)^2. \quad (105)$$

At small times, the terms proportional to D dominate, and so we can afford to drop the $O(1)$ terms. At late times, $e^{S_2^R}$ becomes large, and so we can work in the scaling regime where $D \rightarrow \infty$, $e^{S_2^R} \rightarrow \infty$ with $\frac{D}{e^{S_2^R}}$ fixed. In this regime, the terms we dropped above are again small in $e^{-S_2^R}$.

We will first focus on the case $q \neq 0$, since the negativity coming from $q = 0$ can at most contribute some $O(1)$ amount. For $q \neq 0$, the cubic and higher terms in the exponent of Eq. (96) are suppressed in the large D limit (see

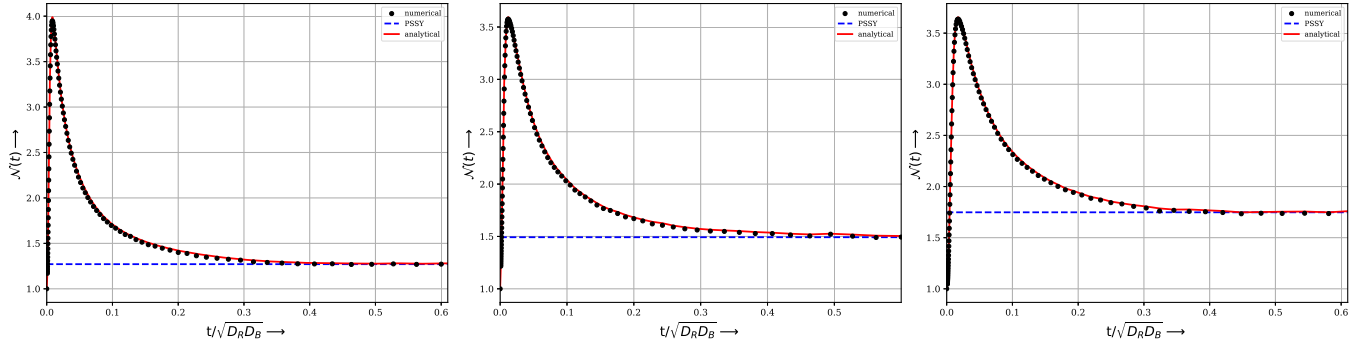


FIG. 9. Log-log plot of the Wigner negativity for the case where the operators O_R and O_B in equation (59) are chosen from the GUE. The radiation Hilbert space is taken to be $D_R = 101$, while the black hole Hilbert space dimension is taken to be $D_B = 200$ (left), 101 (center) and 50 (right) respectively.

Appendix D for more details), and so we can resort to the Gaussian approximation for the integral,

$$\mathcal{N}_{\text{Haar}}|_{q \neq 0} \simeq \lim_{\epsilon \rightarrow 0} \sum_{q,p} \frac{2D_2}{D_1} \int_{-\infty}^{\infty} \frac{ds}{2\pi i} \frac{2s}{s^2 + \left(\frac{2D_2\epsilon}{D_1}\right)^2} \left(-i \frac{\partial}{\partial s}\right) e^{\frac{D_1^2}{4D_2}(2is-s^2)}. \quad (106)$$

This is the same integral we encountered in the PSSY model, and gives

$$\mathcal{N}_{\text{Haar}}|_{q \neq 0} \simeq (1 - S_0) \left(\frac{1}{\sqrt{\pi r_0}} e^{-r_0} + \text{erf}(\sqrt{r_0}) \right), \quad (107)$$

where we have defined

$$r_0 = \frac{D_1^2}{4D_2} = \frac{(1 - S_0)^2}{2[D(e^{-S_2^R} - S_0^2) - (1 - S_0)^2]}. \quad (108)$$

We now consider the $q = 0$ case. Here, the Gaussian approximation fails at early times (see Appendix D), but becomes valid at later times when the scaling limit $D \rightarrow \infty$, $e^{S_2^R} \rightarrow \infty$, $\frac{D}{e^{S_2^R}}$ fixed becomes valid. We will not worry about the failure of the Gaussian approximation at early times too much, as this can at most give an $O(1)$ correction to the negativity. Performing the Gaussian integral as before, we then find

$$\mathcal{N}_{\text{Haar}}|_{q=0} \simeq S_0 \left(\frac{1}{\sqrt{\pi r_0}} e^{-r_0} + \text{erf}(\sqrt{r_0}) \right), \quad (109)$$

where r_0 is defined as,

$$r_0 = \frac{D^2 S_0^2}{2[D(e^{-S_2^R} - 2g + S_0^2) - (1 - S_0)^2]}. \quad (110)$$

The total negativity is then given by

$$\mathcal{N}_{\text{Haar}} = \mathcal{N}_{\text{Haar}}|_{q=0} + \mathcal{N}_{\text{Haar}}|_{q \neq 0}. \quad (111)$$

Finally, the ensemble averaged negativity can be obtained from the Haar averaged negativity using Eq. (70),

$$\overline{\mathcal{N}_B(t)} = \int dH_{\text{BR}} \mu(H_{\text{BR}}) \mathcal{N}_{\text{Haar}}.$$

In Fig. 9 we compare our analytical formula for the Wigner negativity with numerical calculations and find a good match. These figures show an interesting feature of negativity growth—immediately after the black hole is coupled to the radiation bath, the negativity shows a sharp spike and becomes $O(\sqrt{D})$. This instantaneous rise in the negativity happens due to the interactions between the black hole and the radiation bath, and can essentially be thought of as a “negativity shock.” But soon after, the negativity starts to decay—this happens because of *decoherence*. When the black hole dimension is much larger than that of the radiation bath, the black hole serves as a good environment and the bath decoheres to a classical density matrix. On the other hand, when the black hole is comparable or much smaller than the radiation bath, then the radiation system does not decohere completely, but only partially.

It is instructive to study our analytical formula in limiting cases. When the black hole dimension is taken to be small, then the above formula reduces to the one derived in [48] for the time evolution of Wigner negativity for pure states. Another interesting case is $t \rightarrow \infty$. At very late times, we expect that $S_0 \sim \frac{1}{D}$, $g \sim \frac{1}{D^2}$, and $e^{-S_2^R} \sim e^{-S_{2,\text{eq}}^R} \sim \frac{1}{D} + \frac{1}{e^{S_0}}$. In this limit, we get

$$\mathcal{N}_{\text{Haar}} \simeq \left(\frac{1}{\sqrt{\pi r}} e^{-r} + \text{erf}(\sqrt{r}) \right), \quad r = \frac{1}{2(De^{-S_{2,\text{eq}}^R} - 1)}. \quad (112)$$

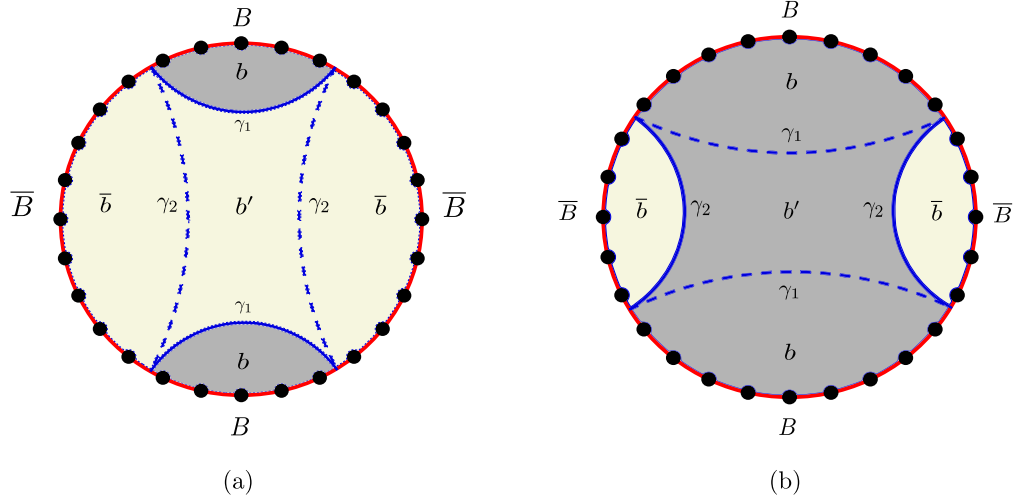


FIG. 10. Spatial slice of AdS_3 is shown, where the black dots on the boundary schematically represent the boundary degrees of freedom. Two competing extremal surfaces, γ_1 and γ_2 , enclose a region b' between them. (a) When $A(\gamma_1) < A(\gamma_2)$, the entanglement wedge of B does not include b' . (b) When $A(\gamma_1) > A(\gamma_2)$, the entanglement wedge of B includes b' . In this case, the region b' constitutes a python's lunch.

This formula agrees with the answer for the Wigner negativity we obtained in the PSSY model in the micro-canonical ensemble, and also with the negativity of Haar random states [55]. Thus, the gravitational result appears as the late time equilibrium value for the negativity in our dynamical model of black hole evaporation. This also explains why two-boundary wormholes were the only relevant wormholes in the gravity calculation—the Haar average over the Wigner negativity only depends on the second Rényi entropy, and the corresponding replica wormholes are two-boundary wormholes.

V. STABILIZER COMPLEXITY AND THE PYTHON'S LUNCH

In the context of AdS/CFT, the entanglement wedge reconstruction paradigm states that bulk degrees of freedom in the entanglement wedge of a given boundary subregion B are encoded in and can be reconstructed in terms of operators in B . The basic framework underlying this is the theory of quantum error correction [12–16]. When there exist multiple quantum extremal surfaces (QESs) satisfying the homology condition with respect to the same boundary subregion, the entanglement wedge reconstruction paradigm gets more interesting. It was shown in [27] that the reconstruction of bulk operators up to the outermost (i.e., closest to the boundary subregion) QES can be accomplished by “simple” operators in B , essentially combining ideas from Hamilton, Kabat, Lifschytz, and Lowe [83] together with forward and backward time evolution. However, for operators behind the outermost QES (but still within the entanglement wedge), no such simple reconstruction exists. In fact, based on the analogy between the time-reflection symmetric slice in AdS and tensor

networks, it was conjectured in [21] that reconstruction of operators behind the outermost QES should be exponentially complex. This conjecture is known as the Python's lunch conjecture, referring to the characteristic bulge in the entanglement wedge geometry. Part of the difficulty in proving the python's lunch conjecture is the lack of computable measures of complexity. Here, we will make some remarks about the python's lunch conjecture from the perspective of Wigner negativity. For the most part, we will confine our attention to the simplified setting of time reflection symmetric states, and where the entanglement wedge of the boundary subregion B has two locally minimal surfaces γ_1 and γ_2 (see Fig. 10 for an example).

Recall that in our calculation of Wigner negativity in the PSSY model, we observed that the negativity of the radiation state remains small before the Page time, but begins to grow exponentially past the Page time. One way to interpret this result is from the perspective of the Python's lunch conjecture. As the black hole evaporates past the Page time, the entanglement wedge of the radiation bath forms an island inside the black hole interior. This island region constitutes a python's lunch from the radiation system perspective, as it lies behind a subdominant QES, namely the empty surface. We could interpret the fact that the Wigner negativity becomes exponentially large beyond the Page time as a manifestation of the formation of a python's lunch region in the entanglement wedge.

If this interpretation is correct, then the same phenomenon should be present in more general holographic states. Consider a time-reflection symmetric, bipartite holographic state $\psi_{B,\bar{B}}$ for complementary boundary subregions B and $\bar{B}$ such that there are two minimal surfaces in the bulk geometry homologous to B (see Fig. 10 for an example).

Let γ_{out} be the outermost minimal surface and let γ_{min} be the minimal surface with the smaller entropy. We would like to compute the Wigner negativity of the reduced density matrix on B . One complication is that in general the boundary Hilbert space is infinite dimensional, so it is not immediately clear how to apply our previous finite-dimensional results in this setting. We will proceed by truncating the Hilbert space of B to a finite but large prime dimensional “microcanonical” subspace (more on this below). A second complication has to do with the choice of computational basis. While the Wigner negativity of the reduced density matrix on B will depend on a choice of basis, we can get a natural basis-independent quantity by averaging over the choice of basis with the Haar measure. This calculation is essentially identical to the late-time limit of the calculation carried out in the previous Section, and the answer is given by,

$$\mathcal{N}_{\text{Haar}} = \left(\frac{1}{\sqrt{\pi r}} e^{-r} + \text{erf}(\sqrt{r}) \right), \quad (113)$$

where

$$r = \frac{1}{2(De^{-S_2^B} - 1)}, \quad (114)$$

with S_2^B being the second Rényi entropy of the reduced density matrix on B . An irritating feature of this formula is that it involves the second Rényi entropy, which does not have a simple geometric characterization in terms of the original geometry dual to the state $\psi_{B,\bar{B}}$. Nevertheless, we will argue below that the correct formula for the negativity of a general holographic state should involve areas of extremal surfaces in the original geometry.

As a first pass, we will derive a lower bound on the negativity which only depends on areas of minimal surfaces in the original geometry. We proceed by making the following two assumptions: first, we will assume that the area of the outermost minimal surface calculates some kind of a *coarse-grained* entropy for the state ρ_B . The precise definition of this coarse-grained entropy is not important (see [84,85] for more details), but the only thing we will assume is that the choice of UV regulator in truncating the Hilbert space of B is consistent with this interpretation, in that

$$\frac{1}{4G_N} A(\gamma_{\text{out}}) \leq \log D. \quad (115)$$

Second, we will assume that the von Neumann entropy of the truncated state ρ_B is approximately (i.e., at leading order in $\frac{1}{G_N}$) equal to the area of the minimal area extremal surface γ_{min} . In this case, we have

$$e^{-S_2^B} \geq e^{-S_{vN}} \sim e^{-\frac{1}{4G_N} A(\gamma_{\text{min}})}. \quad (116)$$

These two inequalities imply that

$$De^{-S_2^B} \geq e^\Delta, \quad \Delta = \frac{1}{4G_N} (A(\gamma_{\text{out}}) - A(\gamma_{\text{min}})). \quad (117)$$

Note that when the gap Δ is large, the formula for the Haar averaged Wigner negativity reduces to,

$$\mathcal{N}_{\text{Haar}} \sim \sqrt{\frac{2}{\pi}} De^{-S_2^B} \geq \sqrt{\frac{2}{\pi}} e^{\frac{\Delta}{2}}. \quad (118)$$

Thus, the negativity becomes large, and scales at least exponentially in the gap between the areas of the outermost and minimal extremal surfaces. This line of reasoning leads us to expect that holographic states with a python’s lunch in their entanglement wedge will have a large stabilizer complexity for the corresponding reduced density matrix on the boundary subregion B with respect to a generic choice of computational basis. Of course, for this result to be genuinely interesting, we must also be able to show that when the gap is small, the negativity becomes small. More ambitiously, we would want to derive a formula for the negativity, and not just a lower bound.

We will now sketch out a proposal for a general formula for the Wigner negativity in holographic states. First, we have to be clear on what Hilbert space truncation we will use. Our proposal is that we should write the full state $\psi_{B,\bar{B}}$ in terms of *random tensor networks* (RTNs) [86–89], or equivalently, *fixed-area* states [90,91], and keep only the RTN/fixed-area state where the probability distribution over RTNs/areas is peaked. Roughly speaking, this can be thought of as truncating the boundary Hilbert space to a “microcanonical” window corresponding to fixed macroscopic variables (such as energy, area etc.). Note that while fixed-area states are defined from a bulk point of view as approximate eigenstates of the area operators, they behave quite similarly to RTNs (see Appendix E of [7]); in particular, they have a flat entanglement spectrum, and hence Rényi entropies independent of the Rényi parameter to leading order in $\frac{1}{G_N}$. This analogy between fixed-area/random tensor network states suggests that for such states it is reasonable to make the identifications,

$$D \sim e^{\frac{1}{4G_N} A(\gamma_{\text{out}})}, \quad e^{S_2^B} \sim e^{S_{vN}} \sim e^{\frac{1}{4G_N} A(\gamma_{\text{min}})}. \quad (119)$$

Thus, the Wigner negativity for such states should be given by

$$\mathcal{N} = \left(\frac{1}{\sqrt{\pi r}} e^{-r} + \text{erf}(\sqrt{r}) \right),$$

$$r = \exp \left[\frac{1}{4G_N} (A(\gamma_{\text{out}}) - A(\gamma_{\text{min}})) \right]. \quad (120)$$

In fact, for such RTNs, the negativity is automatically independent of choice of computational basis, and one does not need explicit Haar averaging (see [61] and Appendix E for more details).

Of course, general holographic states are superpositions of fixed area states/RTNs [90–92],

$$|\psi_{B\bar{B}}\rangle = \sum_{A_1, A_2} \sqrt{p(A_1, A_2)} |A_1, A_2\rangle, \quad (121)$$

where A_1 and A_2 are the areas of the two competing extremal surfaces. The probability distribution $p(A_1, A_2)$ is sharply peaked around values of A_1 and A_2 given by the bulk geometry dual to $\psi_{B, \bar{B}}$. As mentioned above, our proposal is that the negativity, at leading order in $\frac{1}{G_N}$, should simply be given by the formula (120) applied to the fixed-area state on which $p(A_1, A_2)$ is peaked. The motivation behind this proposal is as follows: in defining the Wigner negativity by averaging over choice of computational basis, we must decide what to average over. Equation (114) follows from a naïve Haar average over the entire Hilbert space, which does not make sense for infinite dimensional Hilbert spaces. On the other hand, it seems more physically motivated to Haar average over the microstates within each fixed (A_1, A_2) sector of the Hilbert space (or from the point of view of RTNs, within a fixed microcanonical sector with fixed macroscopic variables), while treating (A_1, A_2) as labels for classical superselection sectors (at leading order in $\frac{1}{G_N}$). Alternatively, we can regard this as a particular choice of UV truncation to make sense of the Wigner negativity, i.e., we compute the negativity of the fixed-area state/random tensor network that best approximates the original state $\psi_{B, \bar{B}}$. Thus, we are led to propose that the Wigner negativity of a general holographic state should be given by Eq. (120) at leading order in $\frac{1}{G_N}$. As an immediate consequence of our proposal, the presence of a python's lunch region in the entanglement wedge results in an exponentially large stabilizer complexity for the corresponding boundary state,

$$\mathcal{N} \sim \sqrt{\frac{2}{\pi}} \exp \left[\frac{1}{2} \frac{(A(\gamma_{\text{out}}) - A(\gamma_{\text{min}}))}{4G_N} \right]. \quad (122)$$

Interestingly, our proposed formula for the negativity only seems to involve the outermost and the minimal extremal surfaces, but not the bulge surface that plays a crucial role in the python's lunch conjecture. This is not directly a problem, because the python's lunch conjecture is a statement about the complexity of bulk reconstruction

within the python's lunch region, and not about the stabilizer complexity of the boundary density matrix. It would be interesting to use our techniques to explicitly calculate the stabilizer complexity of the Petz recovery channel that implements bulk reconstruction in the python's lunch [7,16,93]; we leave this to future work.

ACKNOWLEDGMENTS

We would like to thank Abhijit Gadde, Shiraz Minwalla, Shiroman Prakash, Sandip Trivedi, and Spenta Wadia for helpful discussions and Vijay Balasubramanian and Alex May for comments on an earlier version of the draft. We are grateful to Suman Nandi for his significant help with numerical computations. We acknowledge support from the Department of Atomic Energy, Government of India, under Project No. RTI 4002, and from the Infosys Endowment for the study of the Quantum Structure of Spacetime.

DATA AVAILABILITY

The data and code supporting the findings of this article are available in Ref. [94].

APPENDIX A: EVALUATING THE INTEGRAL

In this Section, we will evaluate the integral obtained in Eq. (55),

$$I = \lim_{\epsilon \rightarrow 0} \int_{-\infty}^{\infty} \frac{dt}{2\pi i} \frac{-2it}{t^2 + \epsilon^2} \frac{d}{dt} \exp[-r(-2it + t^2)], \quad r = \frac{e^{S_2}}{2D}. \quad (A1)$$

By completing squares in the exponent, we can write this integral as,

$$I = \lim_{\epsilon \rightarrow 0} e^{-r} \int_{-\infty}^{\infty} \frac{dt}{2\pi i} \frac{-2it}{t^2 + \epsilon^2} \frac{d}{dt} \exp[-r(t-i)^2], \quad r = \frac{e^{S_2}}{2D}. \quad (A2)$$

We pick up the pole at $t = i\epsilon$ and then deform the contour in the upper half plane to the contour $\gamma = \mathbb{R} + i$.¹¹ This gives,

$$I = 2r + e^{-r} \int_{-\infty}^{\infty} \frac{dx}{2\pi i} \frac{-2i}{(x+i)} \frac{d}{dx} e^{-rx^2}$$

$$= 2r + e^{-r} \int_{-\infty}^{\infty} \frac{dx}{2\pi i} \frac{-2i}{(x+i)^2} e^{-rx^2}, \quad (A3)$$

where in the second line we have integrated by parts. We can simplify the remaining integral in the following way:

¹¹The contour segments at infinity are exponentially suppressed and can be dropped.

$$\begin{aligned}
I &= 2r + \frac{1}{\sqrt{4\pi r}} e^{-r} \int_{-\infty}^{\infty} \frac{dx}{2\pi i} \frac{-2i}{(x+i)^2} \int_{-\infty}^{\infty} du e^{-\frac{1}{4r}u^2 - iux} \\
&= 2r + \frac{1}{\sqrt{4\pi r}} e^{-r} \int_{-\infty}^{\infty} du e^{-\frac{1}{4r}u^2} \int_{-\infty}^{\infty} \frac{dx}{2\pi i} \frac{-2i}{(x+i)^2} e^{-iux}.
\end{aligned} \tag{A4}$$

Using

$$\int_{-\infty}^{\infty} \frac{dx}{2\pi i} \frac{-2i}{(x+i)^2} e^{-iux} = 2u\theta(u)e^{-u}, \tag{A5}$$

we get

$$\begin{aligned}
I &= 2r + \frac{1}{\sqrt{4\pi r}} e^{-r} \int_0^{\infty} du 2ue^{-\frac{1}{4r}u^2 - u} \\
&= 2r + \frac{1}{\sqrt{\pi r}} \int_0^{\infty} du ue^{-\frac{1}{4r}(u+2r)^2} \\
&= 2r + \frac{1}{\sqrt{\pi r}} \int_{2r}^{\infty} dy ye^{-\frac{1}{4r}y^2} - \frac{2r}{\sqrt{\pi r}} \int_{2r}^{\infty} dy e^{-\frac{1}{4r}y^2}.
\end{aligned} \tag{A6}$$

The integral in the second term above is elementary and is given by

$$\frac{1}{\sqrt{\pi r}} \int_{2r}^{\infty} dy ye^{-\frac{1}{4r}y^2} = \sqrt{\frac{r}{\pi}} e^{-r}. \tag{A7}$$

The integral in the third term can be written as

$$-\frac{2r}{\sqrt{\pi r}} \int_0^{\infty} dy e^{-\frac{1}{4r}y^2} + \frac{2r}{\sqrt{\pi r}} \int_0^{2r} dy e^{-\frac{1}{4r}y^2}. \tag{A8}$$

The first of these terms cancels the pole contribution in I , while the second term can be written in terms of the error function. This leads to Eq. (58).

APPENDIX B: REMARKS ON FLUCTUATIONS IN THE PSSY MODEL

We wish to argue that the variance for the Wigner negativity in the PSSY model is suppressed both in D and in e^{S_0} . To quantify this, we consider the quantity $\overline{\mathcal{N}^2} - (\bar{\mathcal{N}})^2$, which can be evaluated using the gravitational path integral. For a rough estimate of its order of magnitude, we will use the replica trick introduced earlier, namely $|W| = \lim_{n \rightarrow 1/2} W^{2n}$. This leads us to the following expression:

$$\begin{aligned}
\overline{\mathcal{N}^2} - (\bar{\mathcal{N}})^2 &= \sum_{q,p} \sum_{q',p'} \lim_{m \rightarrow \frac{1}{2}} \lim_{n \rightarrow \frac{1}{2}} [\overline{W^{2m}(q,p) W^{2n}(q',p')} \\
&\quad - \overline{W^{2m}(q,p)} \times \overline{W^{2n}(q',p')}] .
\end{aligned} \tag{B1}$$

One can impose the boundary conditions associated with each term in the above expression and then fill in the bulk

by connecting the various boundaries in all possible ways. We observe that only those diagrams appearing in $\overline{W^{2m}(q,p) W^{2n}(q',p')}$ which connect across—that is, involve contractions between both $A(q,p)$ and $A(q',p')$ within at least one of their irreducible components—contribute to the variance. Importantly, any such diagram can only get enhanced provided $q = q'$ and $p = p'$. Thus, these diagrams must necessarily be suppressed in comparison with those that do not connect across.

Suppose there are $\mathcal{C}(m,n)$ such diagrams. Each diagram may be regarded as a product of α_1 one-boundary contractions, α_2 two-boundary contractions, and so on up to α_{2m+2n} $(2m+2n)$ -boundary contractions, subject to the constraint $\sum_i i\alpha_i = 2m+2n$. Such a diagram contributes at order

$$(e^{S_0})^{\sum_{i=1}^{2m+2n} \alpha_i} (D)^{\sum_{i=1}^{m+n} \alpha_{2i}}.$$

The fully disconnected configuration $\{\alpha_1 = 2m+2n, \alpha_i = 0 \text{ for } i \geq 2\}$ is excluded since all diagrams are required to connect across. One then finds upon inspection that the maximal power of e^{S_0} can be $2m+2n-1$, while the maximal power of D can be $m+n$. Now recall, that there are overall prefactors $(\frac{1}{D^2 e^{S_0}})^{2m} (\frac{1}{D^2 e^{S_0}})^{2n}$ originating from the definition of the Wigner function. Taking these into account, we may therefore write,

$$\begin{aligned}
\overline{\mathcal{N}^2} - (\bar{\mathcal{N}})^2 &= \sum_{q,p} \sum_{q',p'} \lim_{m \rightarrow \frac{1}{2}} \lim_{n \rightarrow \frac{1}{2}} \delta_{q,q'} \delta_{p,p'} \\
&\quad \times \frac{1}{D^{4m+4n}} \times \mathcal{P}(D, e^{S_0}).
\end{aligned} \tag{B2}$$

Here, $\mathcal{P}(D, e^{S_0})$ denotes a polynomial in $\frac{D}{e^{S_0}}$ and $\frac{1}{e^{S_0}}$, containing in total $\mathcal{C}(m,n)$ terms. It turns out that when $\frac{D}{e^{S_0}} < 1$, we have $\mathcal{P}(D, e^{S_0}) \leq \mathcal{C}(m,n) \frac{D}{e^{S_0}}$, where $\mathcal{C}(m,n)$ does not scale with D or e^{S_0} . Consequently, the entire sum over diagrams is approximately of order $\frac{1}{D^{4m+4n-1} e^{S_0}}$. Upon analytically continuing both m and n to $1/2$ and subsequently summing over q, q' and p, p' , one finds that the variance scales as $\frac{1}{D e^{S_0}}$. On the other hand, when $\frac{D}{e^{S_0}} > 1$, one instead finds $\mathcal{P}(D, e^{S_0}) \leq \mathcal{C}(m,n) (\frac{D}{e^{S_0}})^{n+m}$, so that the entire sum is approximately of order $\frac{1}{D^{3n+3m} (e^{S_0})^{n+m}}$. In this case as well, after the continuation $m, n \rightarrow 1/2$, the variance scales as $\frac{1}{D e^{S_0}}$. Thus, regardless of the relative size of D and e^{S_0} , the variance is suppressed as $O(\frac{1}{D e^{S_0}})$.

APPENDIX C: DETAILS OF THE CALCULATION OF $\mathcal{I}(q,p;z)$

In this Appendix, we present the details of the derivation of (92) quoted in the main text. Recall that we had,

$$\mathcal{I}(q, p; z) = \exp \left[iz \text{ (blob with dot) } \right] \underbrace{\frac{1}{\text{Vol}_{U(D-1)}} \int dU \exp \left\{ iz \left[\text{blob with dot} + \text{blob with dot} + \text{blob with dot} \right] \right\}}_{\mathcal{J}}.$$

Our task here is to systematically compute the integral $\mathcal{J}$. From (87), we have

$$\mathcal{J} = \sum_{n=0}^{\infty} \frac{(iz)^n}{n!} \frac{1}{\text{Vol}_{U(D-1)}} \int dU \exp \left[iz \underbrace{\left[\text{blob with dot} \right]}_A \underbrace{\left[\text{blob with dot} + \text{blob with dot} \right]}_B \right]^n. \quad (\text{C1})$$

Let us now focus on the n th term in the above sum in Eq. (C1). Firstly, note that for n odd, the integral vanishes. Further, when n is even, out of all the terms in the binomial expansion of B , only those contribute which have the same power of U and U^* , because the Haar integral is nonzero only when we have an equal number U s and U^* s in the integrand. So, from the expansion of B , we pick only those terms that have an equal number of U s and U^* s. For $n = 2p$, we can make the replacement,

$$\left[\text{blob with dot} + \text{blob with dot} \right]^{2p} \rightarrow \frac{(2p)!}{(p!)^2} \left(\text{blob with dot} \text{ blob with dot} \right)^p, \quad (\text{C2})$$

and so Eq. (C1) can be written as

$$\mathcal{J} = \sum_{p=0}^{\infty} \frac{(iz)^{2p}}{(p!)^2} \frac{1}{\text{Vol}_{U(D-1)}} \int dU \exp \left[iz \text{ (blob with dot) } \left(\text{blob with dot} \text{ blob with dot} \right)^p \right]. \quad (\text{C3})$$

We can now perform the U integral. In doing so, we must further expand out the exponential inside the integral. Any given term in the expansion now looks like,

$$\underbrace{\left[\text{blob with dot} \right]^m}_{A'} \underbrace{\left(\text{blob with dot} \right)}_{B_1} \underbrace{\left(\text{blob with dot} \right)}_{B_2}^p. \quad (\text{C4})$$

At this stage, we will need to use the following general formula for Haar integration:

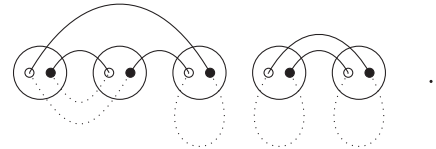
$$\begin{aligned} & \frac{1}{\text{Vol}_{U(D-1)}} \int dU U_{i_1 j_1} U_{i_2 j_2} \dots U_{i_n j_n} U_{i'_1 j'_1}^* U_{i'_2 j'_2}^* \dots U_{i'_n j'_n}^* \\ &= \sum_{\sigma, \tau \in S_n} \text{Wg}(\sigma\tau^{-1}, D-1) \delta_{i_1 i'_{\sigma(1)}} \delta_{i_2 i'_{\sigma(2)}} \dots \delta_{i_n i'_{\sigma(n)}} \\ & \quad \times \delta_{j_1 j'_{\tau(1)}} \delta_{j_2 j'_{\tau(2)}} \dots \delta_{j_n j'_{\tau(n)}}. \end{aligned} \quad (\text{C5})$$

In our diagrammatic approach, each term in Eq. (C5) can be represented by an appropriately contracted diagram. In

general, it is not necessary that the solid and dotted lines get contracted in the same way, i.e., the permutations σ and τ in S_n need not be the same. Consider the example of the following contraction which appears as one of the terms in Eq. (C5) for $n = 5$,

$$\text{Wg}(\sigma\tau^{-1}) \delta_{i_1, i'_3} \delta_{i_2, i'_1} \delta_{i_3, i'_2} \delta_{i_4, i'_5} \delta_{i_5, i'_4} \delta_{j_1, j'_2} \delta_{j_2, j'_1} \delta_{j_3, j'_3} \delta_{j_4, j'_4} \delta_{j_5, j'_5}$$

with $\sigma = (132)(45)$, $\tau = (12)(3)(4)(5)$. We can represent this term diagrammatically as,



From the diagram it appears as if this term factorizes into two irreducible pieces: one three-blob contraction and one two-blob contraction. One might think that for n blobs, any diagram can be built out of various such irreducible pieces: n_1 one-blob contractions, n_2 two-blob contraction, n_3

three-blob contractions and so on, such that $\sum_{p=1}^n p n_p = n$. But this is *not* quite true in general since the Weingarten functions do not factorize accordingly. In our example above, the first irreducible 3 – blob diagram by itself would be proportional to $\text{Wg}(21, D-1)$ and the 2 – blob piece would be proportional to $\text{Wg}(2, D-1)$. But for finite D ,

$$\text{Wg}(2^2 1, D-1) \neq \text{Wg}(21, D-1) \times \text{Wg}(2, D-1).$$

However, things simplify at large D , where the Weingarten functions have the asymptotic behavior,

$$\text{Wg}(\sigma, D) = \frac{1}{D^{n+|\sigma|}} \prod_i (-1)^{|C_i|-1} C_{|C_i|-1} + \mathcal{O}\left(\frac{1}{D^{n+|\sigma|+2}}\right), \quad (\text{C6})$$

where $|\sigma|$ is the smallest number of transpositions (pairwise exchanges) that σ is composed of, $\{C_i\}$ denote the various cycles in the permutation σ , $|C_i|$ is the length of the cycle C_i , and $C_n = \frac{(2n)!}{n!(n+1)!}$ are the Catalan numbers. In the large- D limit, one can check that the Weingarten functions indeed factorize, up to $\mathcal{O}(1/D^2)$ corrections.

To make progress, in our subsequent discussion we will work in the large D limit. In this case, the Haar average over $n(U, U^*)$ pairs can be decomposed into sums of products of various irreducible (i.e., fully connected) pieces with appropriate multiplicity. A word of caution before we proceed: note that the Weingarten functions factorize in the large D limit, up to $\mathcal{O}(1/D^2)$ corrections. If we were only interested in computing Rényi Wigner negativities, i.e., $\overline{W_B^{2n}}$, these corrections would be unimportant (for any fixed Rényi index n) and could be safely ignored in the $D \rightarrow \infty$ limit. But in the integral representation method, we are summing over all moments, and so one might worry that the $\mathcal{O}(1/D^2)$ terms could sum up to give a nontrivial contribution to the Wigner negativity. Recall that in the PSSY calculation, we encountered an analogous situation. There we were able to explicitly resum and check that the subleading corrections are not important. In the present calculation, we will simply assume that the $\mathcal{O}(1/D^2)$ corrections can be ignored, without any further justification. We will later compare our analytic results with numerical calculations and find good agreement, thus justifying the above assumption *a posteriori*.

Let us now go back to equation (C3). Using the notation in (C4), we can schematically express this equation as,

$$\mathcal{J} = \sum_{p=0}^{\infty} \sum_{m=0}^{\infty} \frac{(iz)^{2p}}{(p!)^2} \frac{(iz)^m}{m!} \overline{(A')^m (B_1^p B_2^p)}, \quad (\text{C7})$$

where the overline indicates Haar averaging. $\overline{(A')^m (B_1^p B_2^p)}$ can be expressed as a sum over all possible contracted

diagrams, where each diagram can be understood as a product of irreducible (i.e., fully contracted) pieces, as argued above. In the present case, we proceed as follows. For arbitrary m and p , the A' blobs can either contract among themselves or be sandwiched between the B_1 and B_2 blobs. Suppose n of the A' blobs are contracted with the B_1, B_2 blobs, while n' of them contract only among themselves, with $n + n' = m$. Then any generic term in $\overline{(A')^m (B_1^p B_2^p)}$ can be written as a product of two factors: one involving only the n' number of A' blobs, and the other consisting of the n A' blobs that contract with the p (B_1, B_2) blobs. For now, we leave the first part aside and focus on the second.

Observe that any fully connected diagram containing a B_1 blob must also contain a B_2 blob. Thus, (B_1, B_2) always appear in pairs in any fully contracted diagram, and a generic fully contracted diagram may contain multiple such pairs. Hence, the second part decomposes into a product of irreducible components in the following manner: p_1 irreducible pieces each containing one (B_1, B_2) pair, p_2 irreducible pieces each containing two (B_1, B_2) pairs, and so on, up to p_p irreducible pieces each containing p number of pairs, subject to the constraint $\sum_i p_i = p$. For a given set of p B_1 's and p B_2 's, we can first build the possible pairings in $p!$ ways. Then any particular configuration $\{p_i\}$ can be realized in $\frac{p!}{p_1! p_2! \dots p_p! (1!)^{p_1} (2!)^{p_2} \dots (p!)^{p_p}}$ different ways. Thus, the combinatorial factor associated with distributing p pairs of (B_1, B_2) into various irreducible components is given by,

$$\frac{(p!)^2}{\prod_{i=1}^p p_i! (i!)^{p_i}}.$$

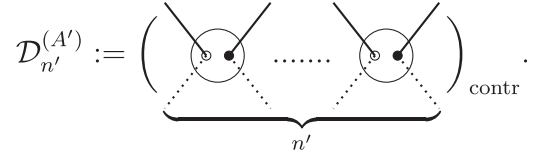
We now address the matter of distributing the n copies of A' blobs among the irreducible structures introduced above. For a fixed configuration $\{p_i\}$, this proceeds as follows. Consider first the p_1 irreducible pieces of the first kind [each containing a single (B_1, B_2) pair]. Suppose that in the j th such piece, $n_j^{(1)}$ copies of A' are contracted, with $j = 1, 2, \dots, p_1$. Here the superscript indicates the type of irreducible component, while the subscript labels the individual copy within that type. We denote the entire j th piece of the first kind by $\mathcal{D}_{n_j^{(1)}}^{(1)}$. In a similar manner, for the p_2 irreducible pieces of the second kind [each involving two (B_1, B_2) pairs], we may contract $n_j^{(2)}$ copies of A' into the j th such piece, denoting the resulting structure by $\mathcal{D}_{n_j^{(2)}}^{(2)}$. The same prescription applies to higher kinds of irreducible components. Note that each $\mathcal{D}_{n_j^{(i)}}^{(i)}$ represents not a single object, but rather a sum of all possible fully connected diagrams involving i number of (B_1, B_2) pairs and $n_j^{(i)}$ number of A' blobs. In the above way of distributing A'

blobs, the consistency condition is that the total number of A' insertions equals n , namely $\sum_{i=1}^p \sum_{j=1}^{p_i} n_j^{(i)} = n$. Thus, for any fixed choice of $\{p_i\}$, the above distribution of n A' blobs among the irreducible components can be found in

$$\frac{n!}{\prod_{i=1}^p \prod_{j=1}^{p_i} n_j^{(i)}!}$$

number of ways. For each fixed configuration $\{p_i\}$, one must sum over all possible distributions $\{n_j^{(i)}\}$ of the A' blobs across the corresponding irreducible components. Finally, one should sum over every admissible configuration $\{p_i\}$.

We now turn to the factor arising in any generic term of $(A')^m (B_1^p B_2^p)$, which involves only the n' number of A' blobs. These blobs can be selected in $\frac{m!}{n'!m!}$ distinct ways. Importantly, the n' A' blobs are *not* necessarily fully connected among themselves; rather, they may form all possible factorizable configurations. We will discuss this below. For the moment, we collectively denote the entire set of such configurations as



One can now readily write down an expression for $(A')^m (B_1^p B_2^p)$ as a sum of products of irreducible diagrams,

$$\begin{aligned} (A')^m (B_1^p B_2^p) &= \sum'_{\{p_i\}} \sum'_{\{n_j^{(i)}, n'\}} \frac{(p!)^2}{\prod_{i=1}^p p_i! (i!)^{p_i}} \\ &\times \frac{m!}{n'! \times \prod_{i=1}^p \prod_{j=1}^{p_i} n_j^{(i)}!} \prod_{i=1}^p \prod_{j=1}^{p_i} \mathcal{D}_{n_j^{(i)}}^{(i)} \times \mathcal{D}_{n'}^{(A')}. \end{aligned} \quad (\text{C8})$$

Here, the primes on the summations indicate that these sums are constrained, subject to the conditions outlined above. Substituting this expression into Eq. (C7), the first observation is that the overall summation over m effectively decouples the sums over $\{n_j^{(i)}, n'\}$. In other words, one can independently sum over $\{n_j^{(i)}\}$ and n' , with each ranging from 0 to ∞ . Consequently, one may write,

$$\begin{aligned} \mathcal{J} &= \sum_{p=0}^{\infty} \sum'_{\{p_i\}} \frac{(iz)^{2p}}{\prod_{i=1}^p p_i! (i!)^{p_i}} \times \prod_{k=1}^p \left(\sum_{n^{(k)}=0}^{\infty} \frac{(iz)^{n^{(k)}}}{n^{(k)}!} \mathcal{D}_{n^{(k)}}^{(k)} \right)^{p_k} \times \left(\sum_{n'=0}^{\infty} \frac{(iz)^{n'}}{n'!} \mathcal{D}_{n'}^{(A')} \right) \\ &= \mathcal{J}_0 \sum_{p=0}^{\infty} \sum'_{\{p_i\}} \frac{(iz)^{2p}}{\prod_{i=1}^p p_i! (i!)^{p_i}} \times \prod_{k=1}^p \left(\sum_{n^{(k)}=0}^{\infty} \frac{(iz)^{n^{(k)}}}{n^{(k)}!} \mathcal{D}_{n^{(k)}}^{(k)} \right)^{p_k}, \end{aligned} \quad (\text{C9})$$

where we have defined $\mathcal{J}_0$ as,

$$\mathcal{J}_0 = \sum_{n'=0}^{\infty} \frac{(iz)^{n'}}{n'!} \mathcal{D}_{n'}^{(A')}. \quad (\text{C10})$$

We will treat this piece separately later. To obtain the first line of equation (C9), we have used the constraints $\sum_{i=1}^p \sum_{j=1}^{p_i} n_j^{(i)} = n$ and $n + n' = m$. We then distributed the factor $(iz)^m$ among the various groups according to the configuration $\{n_j^{(i)}, n'\}$. Now we further observe that the sums over $\{p_i\}$ also decouple for the same reason as before. In this step, we redistribute the factor $(iz)^{2p}$ in accordance with the constraint $\sum_i i p_i = p$, and thereby rewrite Eq. (C9) as,

$$\begin{aligned} \mathcal{J} &= \mathcal{J}_0 \sum_{p_1=0}^{\infty} \frac{1}{p_1!} \left(\frac{(iz)^2 \mathcal{S}_1}{1!} \right)^{p_1} \sum_{p_2=0}^{\infty} \frac{1}{p_2!} \left(\frac{(iz)^4 \mathcal{S}_2}{2!} \right)^{p_2} \times \dots \\ &= \mathcal{J}_0 e^{\frac{(iz)^2 \mathcal{S}_1}{1!}} e^{\frac{(iz)^4 \mathcal{S}_2}{2!}} e^{\frac{(iz)^6 \mathcal{S}_3}{3!}} \dots \\ &= \mathcal{J}_0 \exp \left[\sum_{k=1}^{\infty} \frac{(iz)^{2k} \mathcal{S}_k}{k!} \right], \end{aligned} \quad (\text{C11})$$

where we have defined $\mathcal{S}_k$ as

$$\mathcal{S}_k = \sum_{n^{(k)}=0}^{\infty} \frac{(iz)^{n^{(k)}}}{n^{(k)}!} \mathcal{D}_{n^{(k)}}^{(k)}. \quad (\text{C12})$$

Recall that $\mathcal{D}_{n^{(k)}}^{(k)}$ denotes the collection of irreducible diagrams containing k pairs of (B_1, B_2) together with $n^{(k)}$ insertions of A' blobs. Each such diagram therefore corresponds to a $(2k + n^{(k)})$ -body contraction. Our goal is to reorganize the sum appearing in the exponent of Eq. (C11) as a sum over n -body contractions, $\mathcal{D}'_n$. This can be expressed as follows:

$$\mathcal{J} = \mathcal{J}_0 \exp \left[\sum_{n=2}^{\infty} (iz)^n \mathcal{D}'_n \right], \quad (\text{C13})$$

where the explicit expressions for $\mathcal{D}'_n$ for the first few values of n are given by,

$$\mathcal{D}'_2 = \frac{1}{1!0!} \mathcal{D}_0^{(1)}, \quad (\text{C14})$$

$$\mathcal{D}'_3 = \frac{1}{1!1!} \mathcal{D}_1^{(1)}, \quad (\text{C15})$$

$$\mathcal{D}'_4 = \frac{1}{1!2!} \mathcal{D}_2^{(1)} + \frac{1}{2!0!} \mathcal{D}_0^{(2)}, \quad (\text{C16})$$

$$\mathcal{D}'_5 = \frac{1}{1!3!} \mathcal{D}_3^{(1)} + \frac{1}{2!1!} \mathcal{D}_1^{(2)}, \quad (\text{C17})$$

$$(\text{C18})$$

or diagrammatically,

$$\mathcal{D}'_2 = \text{diagram} \quad (\text{C19})$$

$$\mathcal{D}'_3 = \text{diagram} \quad (\text{C20})$$

and so on. Now, all that remains is to compute $\mathcal{J}_0$. Recall that $\mathcal{D}_{n'}^{(A')}$ is the set of all possible kinds of contracted diagrams one can build out of n' number of A' blobs. Any such diagram can be obtained by putting together p_1 number of one-blob contractions, p_2 number of two-blob contractions, ..., $p_{n'}$ number of n' blob contractions with the constraint: $\sum_{i=1}^{n'} i p_i = n'$. Here all the i -blob contractions are assumed to be fully connected. We will denote them with $\mathcal{D}_{i,0}$. Any given configuration $\{p_i\}$ can be obtained in $\frac{n'!}{\prod_{i=1}^{n'} (i!)^{p_i} p_i!}$ number of ways. If we sum over all possible configurations of $\{p_i\}$ subject to the constraint mentioned above, we get the entire set $\mathcal{D}_{n'}^{(A')}$,

$$\mathcal{D}_{n'}^{(A')} = \sum_{\{p_i\}} \frac{n'!}{(1!)^{p_1} (2!)^{p_2} \dots (n'!)^{p_{n'}}} \frac{1}{p_1! p_2! \dots p_{n'}!} \times \mathcal{D}_{1,0}^{p_1} \mathcal{D}_{2,0}^{p_2} \dots \mathcal{D}_{n',0}^{p_{n'}}. \quad (\text{C21})$$

Here we draw the first few connected diagrams $\mathcal{D}_{i,0}$.

$$\mathcal{D}_{1,0} = \text{diagram} \quad (\text{C22})$$

$$\mathcal{D}_{2,0} = \text{diagram} + \text{diagram} + \text{diagram} \quad (\text{C23})$$

$$\begin{aligned}
\mathcal{D}_{3,0} = & \text{Diagram 1} + \text{Diagram 2} + \text{Diagram 3} \\
& + 3 \times \text{Diagram 4} + 6 \times \text{Diagram 5} + 3 \times \text{Diagram 6}
\end{aligned}
\tag{C24}$$

Now we notice again that in (C10), we are summing over all possible values of n' . Hence the restricted p sums in (C21) can be performed independently. And we get,

$$\mathcal{J}_0 = \sum_{p_1=0}^{\infty} \frac{1}{p_1!} \left(\frac{(iz)}{1!} \mathcal{D}_{1,0} \right)^{p_1} \sum_{p_2=0}^{\infty} \frac{1}{p_2!} \left(\frac{(iz)^2}{2!} \mathcal{D}_{2,0} \right)^{p_2} \cdots = \exp \left(\sum_{p=1}^{\infty} \frac{(iz)^p}{p!} \mathcal{D}_{p,0} \right). \tag{C25}$$

We have computed all the pieces we require to write down a closed form expression for $\overline{e^{iW_{U,B}}}$. From (C1) and using (C13), (C25) we finally get,

$$\mathcal{I}(q, p; z) = \exp \left(\sum_{n=1}^{\infty} (iz)^n \mathcal{D}_n \right), \tag{C26}$$

APPENDIX D: HIGHER ORDER TERMS

While performing the Gaussian approximation in Eq. (106), we argued that higher-order contributions—cubic, quartic, and beyond—are suppressed. In this Appendix, we explain why this happens. Specifically, for $q \neq 0$, we demonstrate that the ratio $\frac{\mathcal{D}_3 \mathcal{D}_1}{\mathcal{D}_2^2}$ appearing in Eq. (96) is indeed suppressed at early as well as late times. Building on this analysis, we will then identify a general pattern indicating that higher-order terms, such as the quartic contribution, are similarly suppressed. Next, we will consider the $q = 0$ case and show how in this setting, the Gaussian approximation can fail.

$q \neq 0$ case: We begin by evaluating $\mathcal{D}_3$. From Eq. (94), it follows that $\mathcal{D}_3 = \frac{\mathcal{D}_{3,0}}{3!} + \mathcal{D}'_3$. Using (C24) and (C20), we find that the diagrams contributing for the case $q \neq 0$ are given by,

$$\begin{aligned}
\mathcal{D}_3 = & \frac{1}{3!} \times \text{Diagram 1} + \frac{1}{3!} \times \text{Diagram 2} + \frac{1}{3!} \times \text{Diagram 3} \\
& + \frac{1}{2} \times \text{Diagram 4} + \text{Diagram 5} + \frac{1}{2} \times \text{Diagram 6} \\
& + \text{Diagram 7}
\end{aligned}
\tag{D1}$$

We now proceed to evaluate these diagrams. For this purpose, it is convenient to introduce two projectors: $P_0 = |\psi_0\rangle\langle\psi_0|$ and $P_{\perp} = \mathbb{1} - P_0$. Expressed in terms of these projectors, and retaining only the leading contributions of the relevant Weingarten functions, we obtain

$$\mathcal{D}_3 \simeq \frac{1}{D^6} \left[-\frac{1}{3} \text{Tr}((\rho_R P_\perp)^3) + \frac{1}{3D^2} (\text{Tr}(\rho_R P_\perp))^3 + \frac{3}{2D} \text{Tr}((\rho_R P_\perp)^2) \text{Tr}(\rho_R P_\perp) - \text{Tr}((\rho_R P_\perp)^2 \rho_R P_0) \right]. \quad (\text{D2})$$

Note that $\rho_R(t)$ is time dependent. We have already evaluated $\mathcal{D}_2$ for $q \neq 0$ in Eq. (105). For convenience, we reproduce it here,

$$2\mathcal{D}_2 \simeq \frac{1}{D^4} (D \text{Tr}((\rho_R P_\perp)^2) - (\text{Tr}(\rho_R P_\perp))^2 + 2D \text{Tr}(\rho_R P_\perp \rho_R P_0)). \quad (\text{D3})$$

And for $\mathcal{D}_1$ we have,

$$\mathcal{D}_1 = \frac{1}{D^2} (1 - \langle \psi_0 | \rho_R | \psi_0 \rangle). \quad (\text{D4})$$

Now, we can write,

$$\begin{aligned} \text{Tr}((\rho_R P_\perp)^3) &= \text{Tr}((\rho_R)^3) - 3\langle \psi_0 | \rho_R^3 | \psi_0 \rangle + 3\langle \psi_0 | \rho_R | \psi_0 \rangle \langle \psi_0 | \rho_R^2 | \psi_0 \rangle - \langle \psi_0 | \rho_R | \psi_0 \rangle^3, \\ \text{Tr}((\rho_R P_\perp)^2) &= \text{Tr}((\rho_R)^2) - 2\langle \psi_0 | \rho_R^2 | \psi_0 \rangle + \langle \psi_0 | \rho_R | \psi_0 \rangle^2, \\ \text{Tr}(\rho_R P_\perp) &= 1 - \langle \psi_0 | \rho_R | \psi_0 \rangle, \\ \text{Tr}((\rho_R P_\perp)^2 \rho_R P_0) &= \langle \psi_0 | \rho_R^3 | \psi_0 \rangle - 2\langle \psi_0 | \rho_R^2 | \psi_0 \rangle \langle \psi_0 | \rho_R | \psi_0 \rangle + \langle \psi_0 | \rho_R | \psi_0 \rangle^3, \\ \text{Tr}(\rho_R P_\perp \rho_R P_0) &= \langle \psi_0 | \rho_R^2 | \psi_0 \rangle - \langle \psi_0 | \rho_R | \psi_0 \rangle^2. \end{aligned} \quad (\text{D5})$$

We now redefine all these invariants by scaling out a suitable factor of $e^{-S_2^R}$,

$$\text{Tr}((\rho_R)^3) = e^{-2S_2^R} \hat{S}_3, \quad (\text{D6})$$

$$\langle \psi_0 | \rho_R^3 | \psi_0 \rangle = e^{-3S_2^R} \hat{h}, \quad (\text{D7})$$

$$\langle \psi_0 | \rho_R^2 | \psi_0 \rangle = e^{-2S_2^R} \hat{g}, \quad (\text{D8})$$

$$\langle \psi_0 | \rho_R | \psi_0 \rangle = e^{-S_2^R} \hat{S}_0. \quad (\text{D9})$$

All these hatted quantities remain $\mathcal{O}(1)$ at all times. With these redefinitions, the quantities $\mathcal{D}_3$, $\mathcal{D}_2$, and $\mathcal{D}_1$ can be expressed as follows:

$$\mathcal{D}_3 \simeq \frac{1}{D^6} e^{-3S_2^R} \left[-\frac{1}{3} (e^{S_2^R} \hat{S}_3 - 3\hat{h} + 3\hat{g}\hat{S}_0 - \hat{S}_0^3) + \frac{1}{3D^2} (e^{S_2^R} - \hat{S}_0)^3 + \frac{3}{2D} (e^{S_2^R} - 2\hat{g} + \hat{S}_0^2) (e^{S_2^R} - \hat{S}_0) - (\hat{h} - 2\hat{g}\hat{S}_0 + \hat{S}_0^3) \right], \quad (\text{D10})$$

$$\mathcal{D}_2 \simeq \frac{1}{D^4} \frac{1}{2!} e^{-2S_2^R} [D(e^{S_2^R} - 2\hat{g} + \hat{S}_0^2) + 2D(\hat{g} - \hat{S}_0) - (e^{S_2^R} - \hat{S}_0)^2], \quad (\text{D11})$$

$$\mathcal{D}_1 = \frac{1}{D^2} e^{-S_2^R} [e^{S_2^R} - \hat{S}_0]. \quad (\text{D12})$$

At small times, $e^{S_2^R}$ is of order unity. Consequently, at early times, the quantities inside the square brackets in Eqs. (D10) and (D12) are also of order one, while the quantity inside the square bracket in equation (D11) is of order D . As a result, we obtain

$$\mathcal{D}_3 \simeq \frac{1}{D^6} e^{-3S_2^R} \times \mathcal{O}(1), \quad \mathcal{D}_2 \simeq \frac{1}{D^4} e^{-2S_2^R} \times D \times \mathcal{O}(1), \quad \mathcal{D}_1 \simeq \frac{1}{D^2} e^{-S_2^R} \times \mathcal{O}(1).$$

Thus, we find that

$$\frac{\mathcal{D}_3 \mathcal{D}_1}{\mathcal{D}_2^2} \simeq \mathcal{O}\left(\frac{1}{D^2}\right), \quad \dots (\text{small times}). \quad (\text{D13})$$

As time progresses, $e^{S_2^R}$ continues to grow. At late times, $e^{S_2^R}$ becomes large and we consider the limit where $D \rightarrow \infty$, $e^{S_2^R} \rightarrow \infty$, with the ratio $\frac{D}{e^{S_2^R}} = \hat{r}$ held fixed. In this regime, we find from (D10), (D11), and (D12),

$$\mathcal{D}_3 \simeq \frac{1}{D^6} \times e^{-2S_2^R} \times \left(-\frac{1}{3}\hat{S}_3 + \frac{1}{3\hat{r}^2} + \frac{3}{2\hat{r}}\right), \quad (\text{D14})$$

$$\mathcal{D}_2 \simeq \frac{1}{D^4} \times (\hat{r} - 1), \quad (\text{D15})$$

$$\mathcal{D}_1 \simeq \frac{1}{D^2}. \quad (\text{D16})$$

Hence at late times, we get,

$$\frac{\mathcal{D}_3 \mathcal{D}_1}{\mathcal{D}_2^2} \simeq e^{-2S_2^R} \times f_1(\hat{r}), \quad \dots (\text{late times}), \quad (\text{D17})$$

where $f_1(\hat{r})$ is $\mathcal{O}(1)$. Thus, we show that the cubic term is suppressed for $q \neq 0$, either by powers of D at early times or by powers of $e^{S_2^R}$ at late times. At this point, the trend becomes evident. The quartic term in (96) appears with the prefactor $\frac{\mathcal{D}_4 \mathcal{D}_1^2}{\mathcal{D}_2^3}$. A similar evaluation of the relevant diagrams contributing to $\mathcal{D}_4$ shows that, for small times, $\mathcal{D}_4 \simeq \frac{1}{D^8} e^{-4S_2^R} \times D \times \mathcal{O}(1)$, while at late times $\mathcal{D}_4 \simeq \frac{1}{D^8} e^{-2S_2^R} \times f_2(\hat{r})$. As a consequence, one can readily verify that at early times the ratio $\frac{\mathcal{D}_4 \mathcal{D}_1^2}{\mathcal{D}_2^3}$ is of order $\frac{1}{D^2}$. On the other hand, at late times this ratio is of order $\frac{1}{e^{2S_2^R}}$. Thus, one can systematically check that the higher-order terms appearing in the exponent in (96) are indeed suppressed.

$q = 0$ case: We now turn to the case $q = 0$. The structure of the calculation is, to a large extent, similar to the $q \neq 0$ case, so we will only sketch the steps here while highlighting the key differences that arise. For $\mathcal{D}_3$, we have the following diagrams:

$$\mathcal{D}_3 = \frac{1}{3!} \left(\text{Diagram 1} + \text{Diagram 2} + \text{Diagram 3} \right) \quad (\text{D18})$$

$$= \frac{1}{D^6} \frac{1}{3!} \left(\text{Tr}((\rho_R P_\perp)^3) + \frac{2}{D^2} (\text{Tr}(\rho_R P_\perp))^3 - \frac{3}{D} \text{Tr}((\rho_R P_\perp)^2) \text{Tr}(\rho_R P_\perp) \right)$$

Using the same redefinitions in terms of the hatted quantities introduced above for $q \neq 0$ case, we can then express the result as,

$$\mathcal{D}_3 = \frac{1}{D^6} \frac{1}{3!} e^{-3S_2^R} \left[(e^{S_2^R} \hat{S}_3 - 3\hat{h} + 3\hat{g}\hat{S}_0 - \hat{S}_0^3) + \frac{2}{D^2} (e^{S_2^R} - \hat{S}_0)^3 - \frac{3}{D} (e^{S_2^R} - 2\hat{g} + \hat{S}_0^2)(e^{S_2^R} - \hat{S}_0) \right]. \quad (\text{D19})$$

Similarly, for $\mathcal{D}_2$, we obtain,

$$2\mathcal{D}_2 = \frac{1}{D^4} (D \text{Tr}((\rho_R P_\perp)^2) - (\text{Tr}(\rho_R P_\perp))^2) = \frac{1}{D^4} e^{-2S_2^R} [D(e^{S_2^R} - 2\hat{g} + \hat{S}_0^2) - (e^{S_2^R} - \hat{S}_0)^2]. \quad (\text{D20})$$

The key point is that, in this case, $\mathcal{D}_1$ is enhanced; that is,

$$\mathcal{D}_1 = \frac{1}{D} \langle \psi_0 | \rho_R | \psi_0 \rangle = \frac{1}{D} e^{-S_2^R} \hat{S}_0. \quad (\text{D21})$$

At small times, one readily finds

$$\mathcal{D}_3 \simeq \frac{1}{D^6} e^{-3S_2^R} \times \mathcal{O}(1), \quad \mathcal{D}_2 \simeq \frac{1}{D^4} e^{-2S_2^R} \times D \times \mathcal{O}(1). \quad (\text{D22})$$

Consequently, the ratio $\frac{\mathcal{D}_3 \mathcal{D}_1}{\mathcal{D}_2^2}$ is of order $\frac{1}{D}$ at early times. At late times, however, $e^{S_2^R}$ becomes large, and one finds

$$\mathcal{D}_3 \simeq \frac{1}{D^6} e^{-2S_2^R} \times f_3(\hat{r}), \quad \mathcal{D}_2 \simeq \frac{1}{D^4} (\hat{r} - 1), \quad \mathcal{D}_1 \simeq \frac{1}{D^2} \hat{r} \hat{S}_0, \quad (\text{D23})$$

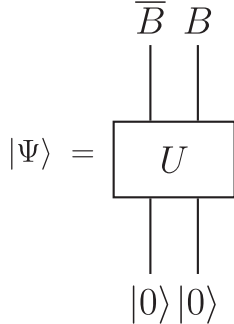
where $f_3(\hat{r})$ is again of order one. Thus, at late times the ratio $\frac{\mathcal{D}_3 \mathcal{D}_1}{\mathcal{D}_2^2}$ is suppressed as $\mathcal{O}(e^{-2S_2^R})$. So far everything seems fine. But the next nontrivial (quartic) term in (96) with the coefficient $\frac{\mathcal{D}_4 \mathcal{D}_1^2}{\mathcal{D}_2^3}$ turns out to be problematic. The discussion above already indicates its form of $\mathcal{D}_4$ at early and late times. At early times one finds $\mathcal{D}_4 \simeq \frac{1}{D^8} e^{-4S_2^R} \times D \times \mathcal{O}(1)$. At late times $\mathcal{D}_4$ instead scales as $\mathcal{D}_4 \simeq \frac{1}{D^8} e^{-2S_2^R} \times f_4(\hat{r})$, with $f_4(\hat{r}) = \mathcal{O}(1)$. It is then straightforward to verify that, at early times, the ratio $\frac{\mathcal{D}_4 \mathcal{D}_1^2}{\mathcal{D}_2^3}$ is of order 1, while at late times the same ratio is suppressed as $e^{-2S_2^R}$. Furthermore, the pattern is as follows: at early times, the coefficients of s^n in (96) with $n = 3, 5, 7, \dots$ are suppressed as $1/D$, whereas the coefficients with $n = 4, 6, 8, \dots$ are of order 1. At late times, all terms beyond s^2 are suppressed by powers of $e^{S_2^R}$. Thus, the Gaussian approximation appears to break down for $q = 0$ at early times, but remains valid at late times where the scaling limit is valid.

APPENDIX E: WIGNER NEGATIVITY FOR RANDOM TENSOR NETWORKS

In this Section, we will compute Wigner negativity in the simplest instance of a random tensor network—a bipartite state with a single random tensor—quite similar to Page’s setup. Consider a bipartite system $\mathcal{H}_B \otimes \mathcal{H}_{\bar{B}}$ with Hilbert space dimensions d_B and $d_{\bar{B}}$, respectively. We prepare a random pure state,

$$|\Psi\rangle = U|0\rangle_{\bar{B}} \otimes |0\rangle_B, \quad (\text{E1})$$

where U is a Haar-random unitary acting on the full Hilbert space. The seed state $|0\rangle_{\bar{B}} \otimes |0\rangle_B$ is chosen to be the one where both subsystems are in first basis state in their respective computational bases. The corresponding circuit diagram is shown below.



For the state under consideration, we are interested in evaluating the Wigner negativity of subsystem B . To establish a connection with fixed-area states in quantum gravity, we can identify, $d_B = \exp(A_1/4G_N)$ and $d_{\bar{B}} = \exp(A_2/4G_N)$, where A_1 and A_2 denote the areas of the two

competing extremal surfaces as we have discussed earlier. In the present setting, the negativity will be expressed as a function of d_B and $d_{\bar{B}}$. In what follows, we focus on the regime in which both d_B and $d_{\bar{B}}$ are large, while their ratio $d_B/d_{\bar{B}}$ is held fixed.

We expand the state $|\Psi\rangle$ in the computational basis $\{|i\rangle_B \otimes |j\rangle_{\bar{B}}\}$ where i runs from 0 to $d_B - 1$ and j runs from 0 to $d_{\bar{B}} - 1$. The reduced density matrix of subsystem B is then obtained by tracing over $\bar{B}$, yielding,

$$\langle l|\rho_B|k\rangle = \sum_{i,i'} U_{il;00} U_{i'k;00}^* \delta_{ii'}. \quad (\text{E2})$$

Subsequently, the Wigner function corresponding to this density matrix is given by,

$$W(q, p) = \frac{1}{d_B} \sum_{k,l=0}^{d_B-1} A_{kl} \langle l|\rho_B|k\rangle$$

where, A_{kl} is the phase point operator: $A_{kl} = e^{\frac{2i\pi p}{d_B}(k-l)} \hat{\delta}_{2q,k+l}$.

Our goal is to compute the Wigner negativity for the state ρ_B , or more precisely the Haar-averaged Wigner negativity (where this time the Haar average is over the random tensor). As in previous Sections, in order to handle the absolute value of the Wigner function, we employ the following integral representation:

$$\begin{aligned} \bar{\mathcal{N}} &= \sum_{q,p} |\overline{W(q,p)}|, \\ \overline{W} &= \lim_{\epsilon \rightarrow 0} \frac{1}{2\pi i} \int_{-\infty}^{\infty} dz \frac{2z}{z^2 + \epsilon^2} \left(-i \frac{\partial}{\partial z} \right) e^{izW}. \end{aligned} \quad (\text{E3})$$

Here, the overline denotes Haar averaging over the random tensor. The calculation proceeds in a manner similar to that outlined in the main text, so we will only sketch the essential steps here. We expand out e^{izW} and evaluate $\overline{W^n}$ by computing the relevant Haar integrals.

First, using (E2), we express W^n in terms of the Haar-random unitaries,

$$\begin{aligned} W^n &= \frac{1}{d_B^n} \sum_{\substack{k_1, l_1, \dots, k_n, l_n \\ i_1, i'_1, \dots, i_n, i'_n}} A_{k_1 l_1} A_{k_2 l_2} \dots A_{k_n l_n} \delta_{i_1 i'_1} \delta_{i_2 i'_2} \dots \delta_{i_n i'_n} \\ &\quad \times U_{i_1 l_1; 00} U_{i_2 l_2; 00} \dots U_{i_n l_n; 00} U_{i'_1 k_1; 00}^* \\ &\quad \times U_{i'_2 k_2; 00}^* \dots U_{i'_n k_n; 00}^*. \end{aligned} \quad (\text{E4})$$

The above expression can be conveniently recast in a compact diagrammatic form. We show this below.

$$W^n = \begin{array}{c} \text{Diagram 1} \end{array} \times \begin{array}{c} \text{Diagram 2} \end{array} \times \dots \times \begin{array}{c} \text{Diagram n} \end{array} \quad (\text{E5})$$

where each piece denotes,

$$\begin{array}{c} \text{Diagram} \end{array} := \frac{1}{d_B} \sum_{k, \ell; i, i'} A_{k\ell} U_{i\ell;00} U_{i'k;00}^* \delta_{ii'} \quad (\text{E6})$$

In the diagram, the diagonal legs represent the row indices of the unitaries U and U^* . The column indices are suppressed, as they are always fixed to zero. The indices k, ℓ correspond to the subsystem B and are depicted by blue lines, while the indices i, i' correspond to the subsystem $\bar{B}$ and are represented by black lines. The symbol A denotes the phase-point operator, and $\mathbb{1}$ corresponds to $\delta_{ii'}$. This diagrammatic representation is particularly convenient for visualizing the effect of averaging over the Haar random unitaries. We now perform the Haar integrals using the following formula:

$$\begin{aligned} \int dU U_{i_1 l_1;00} U_{i_2 l_2;00} \dots U_{i_n l_n;00} U_{i'_1 k_1;00}^* U_{i'_2 k_2;00}^* \dots U_{i'_n k_n;00}^* &= \sum_{\sigma \in S_n} \delta_{i_1 i'_{\sigma(1)}} \delta_{k_1 l_{\sigma(1)}} \delta_{i_2 i'_{\sigma(2)}} \delta_{k_2 l_{\sigma(2)}} \dots \delta_{i_n i'_{\sigma(n)}} \delta_{k_n l_{\sigma(n)}} \left(\sum_{\tau} \text{Wg}(\tau, d) \right) \\ &\approx \frac{1}{d_B^n d_{\bar{B}}^n} \sum_{\sigma \in S_n} \delta_{i_1 i'_{\sigma(1)}} \delta_{k_1 l_{\sigma(1)}} \delta_{i_2 i'_{\sigma(2)}} \delta_{k_2 l_{\sigma(2)}} \dots \delta_{i_n i'_{\sigma(n)}} \delta_{k_n l_{\sigma(n)}}. \end{aligned} \quad (\text{E7})$$

Here, $d = d_B d_{\bar{B}}$, and the second line follows in the limit where d_B and $d_{\bar{B}}$ are both large, i.e., d is large. Using this result, we can compute $\overline{W^n}$ by contracting the free legs in (E5) according to all elements of the permutation group S_n . Importantly, (E7) dictates that the upper and lower legs are contracted identically. As an illustration, consider the case $n = 5$. The quantity $\overline{W^5}$ admits $5!$ possible contractions, one of which is shown below. This contraction corresponds to the permutation $\sigma = (123)(45)$. The blue lines, associated with subsystem B , form loops that produce the factor $\text{Tr}(A^3)\text{Tr}(A^2)$, while the black lines, associated with subsystem $\bar{B}$, yield $\text{Tr}(\mathbb{1})^2$. Now for any n ,

$$\text{Tr}(A^n) = \begin{cases} 1 & \text{if } n \text{ is odd,} \\ d_B & \text{if } n \text{ is even,} \end{cases} \quad (\text{E8})$$

and $\text{Tr}(\mathbb{1}) = d_{\bar{B}}$.

$$\begin{aligned} \overline{W^5} &\supset \begin{array}{c} \text{Diagram 1} \end{array} \quad \begin{array}{c} \text{Diagram 2} \end{array} \quad (\text{E9}) \\ &= \frac{1}{d_B^5} \times \frac{1}{d_{\bar{B}}^5} \times \text{Tr}(A^3)\text{Tr}(A^2)\text{Tr}(\mathbb{1})^2 \quad \dots \quad \sigma = (123)(45) \end{aligned}$$

This example highlights a key point: the value of any term in $\overline{W}^n$ depends solely on the cycle structure of the corresponding permutation. For S_n , any generic element can be decomposed into c_1 one-cycles, c_2 two-cycles, ..., c_n n cycles, subject to the constraint $\sum_{i=1}^n i c_i = n$. The number of permutations with such a cycle structure is $\frac{n!}{c_1! c_2! \dots c_n!} \times \frac{1}{1^{c_1} 2^{c_2} \dots n^{c_n}}$. Each of these permutations contributes the same value, namely $(\text{Tr}(A)\text{Tr}(1))^{c_1} (\text{Tr}(A^2)\text{Tr}(1))^{c_2} \dots (\text{Tr}(A^n)\text{Tr}(1))^{c_n}$. Now we define

$$D_n := \text{Tr}(A^n)\text{Tr}(1). \quad (\text{E10})$$

These objects D_n may be regarded as the analogs of the fully connected irreducible diagrams introduced in the main text. With this definition, the sum over all permutations in S_n can be reorganized as a sum over distinct cycle structures, subject to the constraint $\sum_i i c_i = n$. Thus, we can write

$$\overline{W}^n = \frac{1}{(d_B^2 d_{\bar{B}})^n} \sum_{\{c_i\}} \frac{n!}{c_1! c_2! \dots c_n!} \prod_{i=1}^n \left(\frac{D_i}{i}\right)^{c_i}. \quad (\text{E11})$$

At this stage, one can easily recognize the pattern. Having encountered similar structures in our earlier discussions, we can immediately write down the corresponding expression for $e^{iz\overline{W}}$.

$$e^{iz\overline{W}} = \sum_{n=0}^{\infty} \frac{(iz)^n}{n!} \overline{W}^n = \exp \left[\sum_{n=1}^{\infty} \left(\frac{iz}{d_B^2 d_{\bar{B}}} \right)^n \frac{D_n}{n} \right]. \quad (\text{E12})$$

Following the same line of reasoning laid out in the PSSY analysis, we can effectively truncate the exponent in the above expression at $n=2$. Substituting this truncated expression into (E3), we obtain the ensemble-averaged absolute value of the Wigner function, $|W|$,

$$|\overline{W}(q, p)| \approx \frac{D_1}{d_B^2 d_{\bar{B}}} \text{erf} \left(\frac{D_1}{\sqrt{2D_2}} \right) + \frac{1}{d_B^2 d_{\bar{B}}} \sqrt{\frac{2D_2}{\pi}} e^{-\frac{D_1^2}{2D_2}}. \quad (\text{E13})$$

From (E10) and (E8), it follows that $D_1 = d_{\bar{B}}$, $D_2 = d_B d_{\bar{B}}$. Consequently, the averaged negativity is given by,

$$\bar{\mathcal{N}} = \text{erf} \left(\sqrt{\frac{d_{\bar{B}}}{2d_B}} \right) + \sqrt{\frac{2d_B}{\pi d_{\bar{B}}}} e^{-\frac{d_{\bar{B}}}{2d_B}}. \quad (\text{E14})$$

Hence, we obtain an explicit expression for the Wigner negativity as a function of the Hilbert space dimensions of the subsystems B and $\bar{B}$ in the simplest tensor network model. More precisely, the negativity depends on the ratio $d_B/d_{\bar{B}}$. Drawing an analogy with fixed-area states in quantum gravity, this corresponds to a dependence on $\exp((A_1 - A_2)/4G_N)$, i.e., the difference between the areas of the competing extremal surfaces γ_1 and γ_2 . Equation (E14) is our final expression for the Wigner negativity in the simplest random tensor network state considered here.

-
- [1] S. W. Hawking, Breakdown of predictability in gravitational collapse, *Phys. Rev. D* **14**, 2460 (1976).
 - [2] D. N. Page, Information in black hole radiation, *Phys. Rev. Lett.* **71**, 3743 (1993).
 - [3] G. Penington, Entanglement wedge reconstruction and the information paradox, *J. High Energy Phys.* **09** (2020) 002.
 - [4] A. Almheiri, N. Engelhardt, D. Marolf, and H. Maxfield, The entropy of bulk quantum fields and the entanglement wedge of an evaporating black hole, *J. High Energy Phys.* **12** (2019) 063.
 - [5] A. Almheiri, R. Mahajan, J. Maldacena, and Y. Zhao, The page curve of Hawking radiation from semiclassical geometry, *J. High Energy Phys.* **03** (2020) 149.
 - [6] A. Almheiri, R. Mahajan, and J. Maldacena, Islands outside the horizon, *arXiv:1910.11077*.
 - [7] G. Penington, S. H. Shenker, D. Stanford, and Z. Yang, Replica wormholes and the black hole interior, *J. High Energy Phys.* **03** (2022) 205.
 - [8] A. Almheiri, T. Hartman, J. Maldacena, E. Shaghoulian, and A. Tajdini, Replica wormholes and the entropy of Hawking radiation, *J. High Energy Phys.* **05** (2020) 013.
 - [9] A. Almheiri, T. Hartman, J. Maldacena, E. Shaghoulian, and A. Tajdini, The entropy of Hawking radiation, *Rev. Mod. Phys.* **93**, 035002 (2021).
 - [10] M. Headrick, V. E. Hubeny, A. Lawrence, and M. Rangamani, Causality & holographic entanglement entropy, *J. High Energy Phys.* **12** (2014) 162.
 - [11] B. Czech, J. L. Karczmarek, F. Nogueira, and M. Van Raamsdonk, The gravity dual of a density matrix, *Classical Quantum Gravity* **29**, 155009 (2012).
 - [12] A. Almheiri, X. Dong, and D. Harlow, Bulk locality and quantum error correction in AdS/CFT, *J. High Energy Phys.* **04** (2015) 163.
 - [13] X. Dong, D. Harlow, and A. C. Wall, Reconstruction of bulk operators within the entanglement wedge in gauge-gravity duality, *Phys. Rev. Lett.* **117**, 021601 (2016).
 - [14] D. Harlow, The Ryu-Takayanagi formula from quantum error correction, *Commun. Math. Phys.* **354**, 865 (2017).
 - [15] T. Faulkner and A. Lewkowycz, Bulk locality from modular flow, *J. High Energy Phys.* **07** (2017) 151.
 - [16] J. Cotler, P. Hayden, G. Penington, G. Salton, B. Swingle, and M. Walter, Entanglement wedge reconstruction

- via universal recovery channels, *Phys. Rev. X* **9**, 031011 (2019).
- [17] V. Balasubramanian, D. Marolf, and M. Rozali, Information recovery from black holes, *Gen. Relativ. Gravit.* **38**, 1529 (2006).
- [18] S. Raju, Is holography implicit in canonical gravity? *Int. J. Mod. Phys. D* **28**, 1944011 (2019).
- [19] C. Chowdhury, O. Papadoulaki, and S. Raju, A physical protocol for observers near the boundary to obtain bulk information in quantum gravity, *SciPost Phys.* **10**, 106 (2021).
- [20] D. Harlow and P. Hayden, Quantum computation vs. firewalls, *J. High Energy Phys.* **06** (2013) 085.
- [21] A. R. Brown, H. Gharibyan, G. Penington, and L. Susskind, The Python's lunch: Geometric obstructions to decoding Hawking radiation, *J. High Energy Phys.* **08** (2020) 121.
- [22] I. H. Kim, E. Tang, and J. Preskill, The ghost in the radiation: robust encodings of the black hole interior, *J. High Energy Phys.* **06** (2020) 031.
- [23] C. Akers, N. Engelhardt, D. Harlow, G. Penington, and S. Vardhan, The black hole interior from non-isometric codes and complexity, *J. High Energy Phys.* **06** (2024) 155.
- [24] L. Yang and N. Engelhardt, The complexity of learning (pseudo)random dynamics of black holes and other chaotic systems, *J. High Energy Phys.* **03** (2025) 153.
- [25] V. Balasubramanian, A. Kar, C. Li, and O. Parrikar, Quantum error correction in the black hole interior, *J. High Energy Phys.* **07** (2022) 189.
- [26] V. Balasubramanian, A. Kar, C. Li, O. Parrikar, and H. Rajgadia, Quantum error correction from complexity in Brownian SYK, *J. High Energy Phys.* **08** (2023) 071.
- [27] N. Engelhardt, G. Penington, and A. Shahbazi-Moghaddam, A world without pythons would be so simple, *Classical Quantum Gravity* **38**, 234001 (2021).
- [28] N. Engelhardt, G. Penington, and A. Shahbazi-Moghaddam, Finding pythons in unexpected places, *Classical Quantum Gravity* **39**, 094002 (2022).
- [29] A. May, S. Pasterski, C. Waddell, and M. Xu, Cryptographic tests of the Python's lunch conjecture, *SciPost Phys.* **19**, 084 (2025).
- [30] G. Arora, M. Headrick, A. Lawrence, M. Sasieta, and C. Wolfe, Geometric surprises in the Python's lunch conjecture, *SciPost Phys.* **16**, 152 (2024).
- [31] N. Engelhardt, G. Penington, and A. Shahbazi-Moghaddam, Twice upon a time: Timelike-separated quantum extremal surfaces, *J. High Energy Phys.* **01** (2024) 033.
- [32] D. Gottesman, The Heisenberg representation of quantum computers, *arXiv:quant-ph/9807006*.
- [33] S. Aaronson and D. Gottesman, Improved simulation of stabilizer circuits, *Phys. Rev. A* **70**, 052328 (2004).
- [34] A. Mari and J. Eisert, Positive Wigner functions render classical simulation of quantum computation efficient, *Phys. Rev. Lett.* **109**, 230503 (2012).
- [35] S. Bravyi and A. Kitaev, Universal quantum computation with ideal Clifford gates and noisy ancillas, *Phys. Rev. A* **71**, 022316 (2005).
- [36] V. Veitch, C. Ferrie, D. Gross, and J. Emerson, Negative quasi-probability as a resource for quantum computation, *New J. Phys.* **14**, 113011 (2012).
- [37] V. Veitch, S. A. H. Mousavian, D. Gottesman, and J. Emerson, The resource theory of stabilizer quantum computation, *New J. Phys.* **16**, 013009 (2014).
- [38] W. K. Wootters, A Wigner-function formulation of finite-state quantum mechanics, *Ann. Phys. (N.Y.)* **176**, 1 (1987).
- [39] U. Leonhardt, Quantum-state tomography and discrete Wigner function, *Phys. Rev. Lett.* **74**, 4101 (1995).
- [40] S. Heiss and S. Weigert, Discrete Moyal-type representations for a spin, *Phys. Rev. A* **63**, 012105 (2000).
- [41] C. Miquel, J. P. Paz, and M. Saraceno, Quantum computers in phase space, *Phys. Rev. A* **65**, 062309 (2002).
- [42] K. S. Gibbons, M. J. Hoffman, and W. K. Wootters, Discrete phase space based on finite fields, *Phys. Rev. A* **70**, 062101 (2004).
- [43] D. Gross, Hudson's theorem for finite-dimensional quantum systems, *J. Math. Phys. (N.Y.)* **47**, 122107 (2006).
- [44] C. Cormick, E. F. Galvão, D. Gottesman, J. P. Paz, and A. O. Pittenger, Classicality in discrete Wigner functions, *Phys. Rev. A* **73**, 012301 (2006).
- [45] H. Pashayan, J. J. Wallman, and S. D. Bartlett, Estimating outcome probabilities of quantum circuits using quasiprobabilities, *Phys. Rev. Lett.* **115**, 070501 (2015).
- [46] E. Wigner, On the quantum correction for thermodynamic equilibrium, *Phys. Rev.* **40**, 749 (1932).
- [47] X. Wang, M. M. Wilde, and Y. Su, Quantifying the magic of quantum channels, *New J. Phys.* **21**, 103002 (2019).
- [48] R. Basu, P. Chowdhury, A. Ganguly, S. Nath, O. Parrikar, and S. Paul, Wigner negativity, random matrices and gravity, *J. High Energy Phys.* **01** (2026) 106.
- [49] E. Hostens, J. Dehaene, and B. D. Moor, Stabilizer states and Clifford operations for systems of arbitrary dimensions and modular arithmetic, *Phys. Rev. A* **71**, 042315 (2005).
- [50] C. D. White, C. Cao, and B. Swingle, Conformal field theories are magical, *Phys. Rev. B* **103**, 075145 (2021).
- [51] Z.-W. Liu and A. Winter, Many-body quantum magic, *PRX Quantum* **3**, 020333 (2022).
- [52] X. Turkeshi, A. Dymarsky, and P. Sierant, Pauli spectrum and nonstabilizerness of typical quantum many-body states, *Phys. Rev. B* **111**, 054301 (2025).
- [53] E. Tirrito, X. Turkeshi, and P. Sierant, Anticoncentration and nonstabilizerness spreading under ergodic quantum dynamics, *Phys. Rev. Lett.* **135**, 220401 (2025).
- [54] J. R. Fliss, Knots, links, and long-range magic, *J. High Energy Phys.* **04** (2021) 090.
- [55] C. D. White and J. H. Wilson, Mana in Haar-random states, *arXiv:2011.13937*.
- [56] K. Goto, T. Nosaka, and M. Nozaki, Probing chaos by magic monotones, *Phys. Rev. D* **106**, 126009 (2022).
- [57] L. Leone, S. F. E. Oliviero, Y. Zhou, and A. Hamma, Quantum chaos is quantum, *Quantum* **5**, 453 (2021).
- [58] L. Chen, R. J. Garcia, K. Bu, and A. Jaffe, Magic of random matrix product states, *Phys. Rev. B* **109**, 174207 (2024).
- [59] P. Niroula, C. D. White, Q. Wang, S. Johri, D. Zhu, C. Monroe, C. Noel, and M. J. Gullans, Phase transition in magic with random quantum circuits, *Nat. Phys.* **20**, 1786 (2024).
- [60] X. Turkeshi and P. Sierant, Error-resilience phase transitions in encoding-decoding quantum circuits, *Phys. Rev. Lett.* **132**, 140401 (2024).

- [61] Y. Zhang and Y. Gu, Quantum magic dynamics in random circuits, *npj Quantum Inf.* **12**, 87 (2026).
- [62] S. F. E. Oliviero, L. Leone, S. Lloyd, and A. Hamma, Unscrambling quantum information with clifford decoders, *Phys. Rev. Lett.* **132**, 080402 (2024).
- [63] C. Cao, Non-trivial area operators require non-local magic, *J. High Energy Phys.* **11** (2024) 105.
- [64] C. Cao, G. Cheng, A. Hamma, L. Leone, W. Munizzi, and S. F. E. Oliviero, Gravitational back-reaction is magical, *PRX Quantum* **6**, 040375 (2025).
- [65] R. Basu, A. Ganguly, S. Nath, and O. Parrikar, Complexity growth and the Krylov-Wigner function, *J. High Energy Phys.* **05** (2024) 264.
- [66] G. Vidal, Efficient classical simulation of slightly entangled quantum computations, *Phys. Rev. Lett.* **91**, 147902 (2003).
- [67] M. A. Nielsen, A geometric approach to quantum circuit lower bounds, *Quantum Inf. Comput.* **6**, 213 (2006).
- [68] M. A. Nielsen, M. R. Dowling, M. Gu, and A. C. Doherty, Quantum computation as geometry, *Science* **311**, 1133 (2006).
- [69] L. Susskind, Entanglement is not enough, *Fortschr. Phys.* **64**, 49 (2016).
- [70] L. Susskind, Computational complexity and black hole horizons, *Fortschr. Phys.* **64**, 24 (2016).
- [71] R. A. Jefferson and R. C. Myers, Circuit complexity in quantum field theory, *J. High Energy Phys.* **10** (2017) 107.
- [72] S. Chapman, M. P. Heller, H. Marrochio, and F. Pastawski, Toward a definition of complexity for quantum field theory states, *Phys. Rev. Lett.* **120**, 121602 (2018).
- [73] A. R. Brown and L. Susskind, Second law of quantum complexity, *Phys. Rev. D* **97**, 086015 (2018).
- [74] V. Balasubramanian, M. Decross, A. Kar, and O. Parrikar, Quantum complexity of time evolution with chaotic hamiltonians, *J. High Energy Phys.* **01** (2020) 134.
- [75] F. G. S. L. Brandão, W. Chemsissany, N. Hunter-Jones, R. Kueng, and J. Preskill, Models of quantum complexity growth, *PRX Quantum* **2**, 030316 (2021).
- [76] J. Haferkamp, P. Faist, N. B. T. Kothakonda, J. Eisert, and N. Y. Halpern, Linear growth of quantum circuit complexity, *Nat. Phys.* **18**, 528 (2022).
- [77] V. Balasubramanian, M. DeCross, A. Kar, Y. C. Li, and O. Parrikar, Complexity growth in integrable and chaotic models, *J. High Energy Phys.* **07** (2021) 011.
- [78] A. Streltsov, G. Adesso, and M. B. Plenio, Colloquium: Quantum coherence as a resource, *Rev. Mod. Phys.* **89**, 041003 (2017).
- [79] V. Balasubramanian, P. Caputa, J. M. Magan, and Q. Wu, Quantum chaos and the complexity of spread of states, *Phys. Rev. D* **106**, 046007 (2022).
- [80] P. Nandy, A. S. Matsoukas-Roubeas, P. Martínez-Azcona, A. Dymarsky, and A. del Campo, Quantum dynamics in Krylov space: Methods and applications, *Phys. Rep.* **1125–1128**, 1 (2025).
- [81] S. Baiguera, V. Balasubramanian, P. Caputa, S. Chapman, J. Haferkamp, M. P. Heller, and N. Y. Halpern, Quantum complexity in gravity, quantum field theory, and quantum information science, *Phys. Rep.* **1159**, 1 (2026).
- [82] P. Saad, S. H. Shenker, and D. Stanford, JT gravity as a matrix integral, *arXiv:1903.11115*.
- [83] A. Hamilton, D. N. Kabat, G. Lifschytz, and D. A. Lowe, Holographic representation of local bulk operators, *Phys. Rev. D* **74**, 066009 (2006).
- [84] N. Engelhardt and A. C. Wall, Coarse graining holographic black holes, *J. High Energy Phys.* **05** (2019) 160.
- [85] J. Chandra and T. Hartman, Coarse graining pure states in AdS/CFT, *J. High Energy Phys.* **10** (2023) 030.
- [86] P. Hayden, S. Nezami, X.-L. Qi, N. Thomas, M. Walter, and Z. Yang, Holographic duality from random tensor networks, *J. High Energy Phys.* **11** (2016) 009.
- [87] J. Chandra and T. Hartman, Toward random tensor networks and holographic codes in CFT, *J. High Energy Phys.* **05** (2023) 109.
- [88] H. Geng, L.-Y. Hung, and Y. Jiang, It from ETH: Multi-interval entanglement and replica wormholes from Large- c BCFT ensemble, *arXiv:2505.20385*.
- [89] N. Bao, H. Geng, and Y. Jiang, Ryu-Takayanagi formula for multi-boundary black holes from 2D large- c CFT ensemble, *J. High Energy Phys.* **10** (2025) 042.
- [90] X. Dong, D. Harlow, and D. Marolf, Flat entanglement spectra in fixed-area states of quantum gravity, *J. High Energy Phys.* **10** (2019) 240.
- [91] C. Akers and P. Rath, Holographic Renyi entropy from quantum error correction, *J. High Energy Phys.* **05** (2019) 052.
- [92] P. Caputa, J. Kruthoff, and O. Parrikar, Building tensor networks for holographic states, *J. High Energy Phys.* **05** (2021) 009; **09** (2022) 112(E).
- [93] O. Parrikar, H. Rajgadia, V. Singh, and J. Sorce, Relational bulk reconstruction from modular flow, *J. High Energy Phys.* **07** (2024) 138.
- [94] R. Basu, On the stabilizer complexity of Hawking radiation, GitHub repository (2026), <https://github.com/ritambasu61/Onthe-stabilizer-complexity-of-Hawking-radiation>.