Holevo information and ensemble theory of gravity

Xiao-Liang Qi, Zhou Shangnan and Zhenbin Yang

Stanford Institute for Theoretical Physics, Stanford University,
450 Serra Mall, Stanford, CA 94305, U.S.A.

E-mail: xlqi@stanford.edu, snzhou@stanford.edu, zhenbiny@stanford.edu

ABSTRACT: Holevo information is an upper bound for the accessible classical information of an ensemble of quantum states. In this work, we use Holevo information to investigate the ensemble theory interpretation of quantum gravity. We study the Holevo information in random tensor network states, where the random parameters are the random tensors at each vertex. Based on the results in random tensor network models, we propose a conjecture on the holographic bulk formula of the Holevo information in the gravity case. As concrete examples of holographic systems, we compute the Holevo information in the ensemble of thermal states and thermo-field double states in the Sachdev-Ye-Kitaev model. The results are consistent with our conjecture.

Contents

1 Introduction

In recent years, various studies of low dimensional gravitational models have suggested that a simple bulk gravitational theory may be described by an ensemble of boundary theories [1–15]. More precisely, the gravitational path integral including topology changes (Euclidean wormholes) is proposed to be the holographic dual of an ensemble average of a family of quantum many-body systems, rather than a single one.[1] From the quantum information perspective, this suggests two possible different interpretations of the bulk geometry as a quantum state. The traditional one, as in the standard holographic dictionary [18], is that a given bulk geometry and a bulk matter state is dual to a definite quantum state ρ on the boundary when a boundary Cauchy surface is given. The picture of ensemble duality suggests another possibility: that a given bulk geometry and a boundary Cauchy surface corresponds to an ensemble average of quantum states $\rho(J)$, where the ensemble parameter J can have a probability distribution p_J. The parameter J may be continuous or discrete. For concreteness, we will assume it is continuous, and normalize the probability distribution as $\int dJ p_J = 1$. These two descriptions correspond to two different sets of rules of gravitational quantizations, and various information paradox can arise if we are not careful enough about distinguishing these rules. One example that illustrates these two different descriptions is the appearance of bra-ket wormhole in a gravitational

---

[1]Historically, the connection between wormholes and ensemble average has been speculated by Coleman [16] and been sharpened in AdS/CFT by Maldacena-Maoz [17].

prepared state [19, 20]: these are global CFT states described by summation of two Euclidean semiclassical gravitational evolutions including the Hartle-Hawking geometry and a bra-ket wormhole geometry. On the one hand, with the standard holographic dictionary, the boundary state is a pure state with vanishing entropy, which can be correctly calculated by the island rules [21]. On the other hand, canonical quantization of the bra-ket wormhole geometry predicts a mixed state of the CFT. Such a state can be thought as an averaged description of an ensemble of different pure states depending on some unknown parameters J: $\rho_a = \int dJ p_J \rho(J)$. While these two interpretations are clear in this case, generically, the meaning of ensemble averaged states in AdS/CFT is less explored.

Assuming that simple gravity is an ensemble theory, a natural question is the "size" of the ensemble. There could be infinite possible values of parameter J. For example in the Sachdev-Ye-Kitaev (SYK) model [22–25], the natural random parameter labeling the states is the random coupling which is a set of independent Gaussian random variables. (We will discuss this model in more detail later.) However, the size of J parameter space is not a physically meaningful measure of the size of the ensemble. For example, one can consider an ensemble in which $\rho(J)$ only depends on certain linear superposition of different J parameters. An intrinsic measure of the size of the ensemble is the information contained in the ensemble, which can be measured by the Holevo information [26]. (We would like to note that Holevo information in quantum gravity has been studied for a different kind of ensemble [27, 28].)

Consider an observer who carries quantum measurements on $\rho(J)$, and tries to learn about the random parameter J. In general, we can consider a projected operator valued measurement (POVM), which is defined by a set of positive Hermitian operators M_a satisfying $\sum_a M_a = \mathbb{I}$. The measurement leads to an output a with probability $P(a|J) = \mathrm{tr}\,(M_a \rho(J))$, which corresponds to the joint probability $P(a, J) = p(J)P(a|J)$. The mutual information $I(a : J)$ between the two random variables a, J measures how much information one can learn about J from the measurement result a. For example, if J is a function of a, the mutual information is maximal. We can define a quantum mutual information that bounds $I(a : J)$ from above. By introducing an ancilla state $|J\rangle$ which are orthogonal for different J, we can define the auxiliary state

$$\rho_{SW} = \int dJ p_J \rho(J) \otimes |J\rangle \langle J| \tag{1.1}$$

where we denote S as the system and W as the ancilla. The mutual information between S and W is the Holevo information, which provides an upper bound to the classical mutual information between J and the measurement result a (since measurements are quantum channels applied to S, which can only reduce the mutual information). The explicit formula of the Holevo information is

$$H = I(S : W) = S\left(\int dJ p_J \rho(J)\right) - \int dJ p_J S\left(\rho(J)\right) \tag{1.2}$$

In this paper, we will study the Holevo information for the ensemble of states in quantum gravity. In section 2, we study the Holevo information for the ensemble of random tensor network states (RTN), which are toy models of quantum gravity. For RTN with

large bond dimension, we show that the Holevo information is determined by the difference between the generalized entropy with trivial entanglement wedge and that with the actual entanglement wedge (i.e. the Ryu-Takayanagi (RT) formula [29]). In section 3, we make a conjecture on the holographic formula of Holevo information for quantum gravity (more precisely, for asymptotically anti de-Sitter (AdS) space and AdS coupled with non-gravitational bath). In particular, in an evaporating black hole after Page time, the Holevo information is given by the difference between the "Hawking entropy" computed for the original geometry without replica wormhole, and the Page entropy which takes into account of the replica wormhole and entanglement island [19, 30–32]. We also discussed a possible generalization of the proposed Holevo information in systems with multiple quantum extremal surfaces. In section 4 we study the SYK model to gain further understanding on the Holevo information. The random parameters are the random coupling of the SYK model. We study two different ensembles of states, including the thermal states and the thermofield double (TFD) states. The Holevo information for the thermal states ensemble is given by the difference between maximal entropy and the thermal entropy. For the TFD states, the Holevo information is determined by the fermion equal-time correlation between the two sides. We discuss the holographic interpretation of these results and show that they are consistent with our conjectured bulk formula. Finally, we provide further discussion in section 5. In particular, we emphasize that the difference between RTN and gravity theory comes from the nonlocality of the random parameters in the latter case, and discuss how to possibly modify the RTN model to make it closer to the gravity theory.

2 Tensor network models

To understand the Holevo information H in quantum gravity, we first consider the toy model of random tensor networks (RTN). An RTN is defined in the following way [33]. Define two quantum systems B (bulk) and S (boundary, in the case of AdS/CFT, or boundary and bath in the case of evaporating black hole [34]). Their Hilbert spaces are denoted as $\mathbb{H}_B$ and $\mathbb{H}_S$, with the total Hilbert space $\mathbb{H}_{BS} \equiv \mathbb{H}_B \otimes \mathbb{H}_S$. In addition, $\mathbb{H}_B = \otimes_x \mathbb{H}_x$ has a tensor factorization into different qudits, which correspond to vertices of the tensor network. An RTN state ρ_s in $\mathbb{H}_S$ is defined by a random product state $|V(J)\rangle \equiv \otimes_x |V_x(J_x)\rangle \in \mathbb{H}_B$, and a state $\rho_P \in \mathbb{H}_{BS}$:

$$\rho_S(J) = p_J^{-1} \text{tr}_B \left(|V(J)\rangle \langle V(J)| \rho_P \right) \tag{2.1}$$

$$\text{with} \quad p_J = \text{tr}_{BS} \left(|V(J)\rangle \langle V(J)| \rho_P \right) \tag{2.2}$$

Here $|V_x(J_x)\rangle$ is a random state in the site Hilbert space $\mathbb{H}_x$. More precisely, $|V_x(J_x)\rangle = U_x |0\rangle$ is obtained by applying a Haar random unitary U_x on an arbitrary reference state $|0\rangle$. J_x parameterizes this symmetric manifold with a uniform probability distribution. If we choose the normalization

$$\int dJ |V(J)\rangle \langle V(J)| = \mathbb{I}_B \tag{2.3}$$

the ensemble of random states $|V(J)\rangle$ defines an ensemble of RTN $\rho_S(J)$ with probability p_J. Physically, we can consider the projection to $|V(J)\rangle$ as the consequence of a POVM, and

the state $\rho_S(J)$ is the state after the measurement for the qubits that are not measured [33]. In this construction, ρ_P can be a generic state, although conventionally a tensor network state refers to the situation when ρ_P consists of a product of EPR pairs defined on links of a graph.

Now we investigate the Holevo information for this ensemble, for a subsystem $A \subset S$. According to eq. (2.3), the averaged state $\overline{\rho_A}$ is easily determined:

$$\overline{\rho_A} = \int dJ p_J \rho_A(J) = \text{tr}_{BC}(\rho_P) \tag{2.4}$$

where $\rho_A(J) = \text{tr}_C \rho_S(J)$ is the density matrix of A for fixed J parameter. Here C is the complement of A in S. We will consider the situation that ρ_P consists of EPR pairs $\rho_P = \prod_{\langle xy \rangle} |xy\rangle \langle xy| \otimes \rho_r$, and the bond dimension of each EPR pair is large, with ρ_r entropy remains finite. In such large bond dimension limit, the Renyi entropy $S^{(n)}(\rho(J))$ is independent of the random parameter J [33], and is given by the RTN version of Ryu-Takayanagi formula [29]:[2]

$$S^{(n)}(\rho_A(J)) \simeq \min_{\Sigma \subset B} S^{(n)}_{\Sigma A}(\rho_P) \tag{2.5}$$

If we consider the large bond dimension limit such that the equation above holds for all integer n, the same formula applies to the von Neumann entropy by analytic continuation:

$$S(\rho_A(J)) \simeq \min_{\Sigma \subset B} S_{\Sigma A}(\rho_P) \tag{2.6}$$

The region Σ that minimizes the righthand side is the entanglement wedge of A, which we denote as Σ_A. Therefore we get

$$H_A = S(\overline{\rho_A}) - \int dJ p_J S(\rho_A(J)) \simeq S_A(\rho_P) - S_{A\Sigma_A}(\rho_P) \tag{2.7}$$

If for all subregion Σ of the bulk, $S_{\Sigma|A}(\rho_P) \geq 0$, the minimization in eq. (2.6) will be given by an empty Σ, which corresponds to zero Holevo information.

We would like to note that the right-hand side of eq. (2.7) is minus the conditional entropy of Σ_A and A:

$$H_A = -S_{\Sigma_A|A}(\rho_P) \tag{2.8}$$

This conditional entropy must be negative in order for an entanglement wedge to exist. Interestingly, this is also the necessary and sufficient condition for quantum teleportation from A to Σ_A [35, 36]. Our calculation shows that for random measurements, the Holevo information of the resulting ensemble (which measures the knowledge of the remaining quantum state about the classical measurement output) is equal to the coherent information. In the quantum setting, coherent information is positive if Σ_A is more strongly

---

[2]More precisely, for a given small deviation δ, there exists a critical bond dimension $D_c = \frac{\alpha}{\delta^2} e^{c_{2n} V}$ with V the number of vertices in the bulk, and α, c_{2n} order-one consants. For $D \gg D_c$, the deviation of the Renyi entropy to the RT value is smaller than δ with a high probability: $\left| S^{(n)}(\rho_A(J)) - \min_{\Sigma \subset B} S^{(n)}_{\Sigma A}(\rho_P) \right| < \delta$ with probability $P(\delta) = 1 - \frac{D_c}{D}$. For more discussions about this, see chapter 7 of ref. [33].
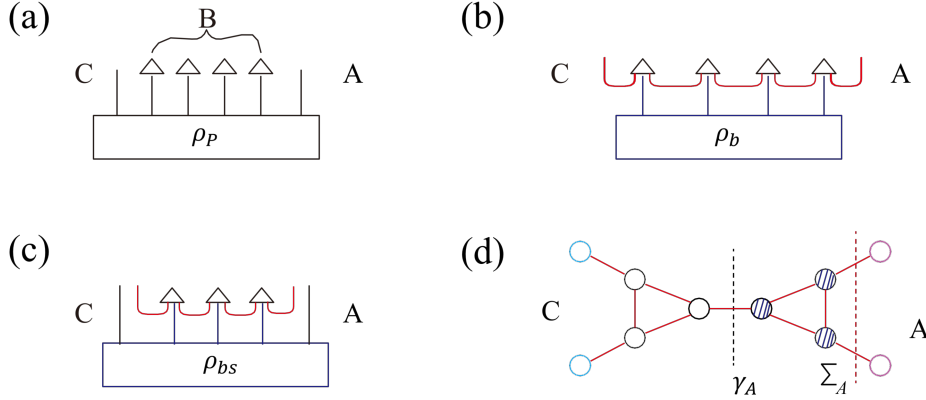
Figure 1. (a) A general random tensor network, with each triangle representing a vertex state. ρ_P is generally a mixed state. Each line here (and also in subfigures (b) and (c)) corresponds to a pair of indices, labeling a complete basis in $\mathbb{H}_x \otimes \overline{\mathbb{H}}$. (b) and (c) Two examples of RTN with ρ_P factorized into bulk QFT state ρ_b and EPR pairs (red curves). The difference between (b) and (c) is whether the boundary is maximally entangled with its complement in ρ_P. (d) A more specific example of RTN, where we have only drawn the EPR pairs. In this simple case, the entropy of boundary region A (purple circles on the right-hand side) is determined by the number of links at the minimal cut γ_A (black dashed line), and the Holevo information is determined by the difference between the area of A (number of links crossing the red dashed line) and that of γ_A.

correlated with A than with the environment, which agrees with our intuition about entanglement wedge.

Eq. (2.7) and (2.8) tells us that the information we can obtain from a subsystem A about the random parameter J is equal to the entropy difference between the contribution of trivial entanglement wedge and the actual entanglement wedge.

To illustrate this result, we can consider two different cases, shown in figure 1(b) and (c). The first example is the RTN corresponding to the AdS/CFT case, when the boundary is connected with the bulk through maximally entangled EPR pairs, and there is a bulk quantum field theory state $|\Psi_b\rangle$. (The bulk state does not have to be pure, but we write the formula for pure state case for simplicity.) The state $\rho_P = |\Psi_P\rangle\langle\Psi_P|$ has the structure

$$|\Psi_P\rangle = \otimes_{\langle xy\rangle \in B} |xy\rangle \otimes |\Psi_b\rangle \otimes \otimes_{x\in B, Y\in S} |xY\rangle \tag{2.9}$$

Here $|xY\rangle$ are maximally entangled EPR pairs between boundary and bulk sites. For simplicity we assume all EPR pairs have the same dimension D. In this case, the averaged state $\overline{\rho_A}$ is maximally mixed, and $S(\rho_A(J))$ is given by the RT formula with quantum corrections, which leads to the Holevo information

$$
\begin{aligned}
H_A &= |A|\log D - S_{\text{gen}}(A) \\
&= |A|\log D - (|\gamma_A|\log D + S_{\Sigma_A}(|\Psi_b\rangle\langle\Psi_b|))
\end{aligned}
\tag{2.10}
$$

$\gamma_A = \partial(\Sigma A)$ is the minimal cut separating A and the complement. This is illustrated in figure 1(b) (for the general setup) and (d) (for a simple example).

More generally, we can consider a state $|\Psi_P\rangle$ with A not maximally entangled with the complement. For example, we can slightly modify the state (2.9):

$$|\Psi_P\rangle = \otimes_{\langle xy\rangle \in B} |xy\rangle \otimes |\Psi_{bS}\rangle \otimes \otimes_{x\in B,Y\in S} |xY\rangle \qquad (2.11)$$

as is illustrated in figure 1(c). In this case, A contains degrees of freedom from $|\Psi_{bS}\rangle$ in addition to EPR pairs, leading to the entropy

$$H_A = |A|\log D + S_A((|\Psi_{bS}\rangle\langle\Psi_{bS}|)) - (|\gamma_A|\log D + S_{A\Sigma_A}(|\Psi_{bS}\rangle\langle\Psi_{bS}|)) \qquad (2.12)$$

One example of this case is the RTN model for an evaporating black hole coupled with a non-gravitational bath [34]. When A is a subsystem of the bath with no direct connection to the gravitational part, there is no area law term $|A|\log D$, and the only contribution to the averaged state entropy is that of the QFT $S_A(|\Psi_{bS}\rangle\langle\Psi_{bS}|)$. More discussions about the evaporating black hole will be carried in next section. Our discussion of the SYK model in section 4 will also be related to this case.

The RT formula (2.5) for RTN comes from a replica caculation of $\text{tr}(\rho_A^n)$ (see appendix A). The two terms in Holevo information correspond to taking n copies of ρ_P and inserting cyclic permutation operators only in A (for the first term) or in $\Sigma_A A$. This suggests that these two terms correspond to saddle points in gravitational calculation where replica wormhole is absent or present, correspondingly. We will discuss the gravity case in the next section.

3 Conjecture about the gravity case

Based on the tensor network results, we can make some conjecture on the Holevo information in AdS/CFT, or AdS/CFT coupled with bath. We start from the case of an AdS black hole coupled with a non-gravitational flat-space bath. As we learned from recent works [19, 30–32], for a large region of the radiation A, after Page time the entropy of A is given by a nontrivial quantum extremal surface which is a boundary of an Island region I:

$$S_A = \frac{|\partial I|}{4G_N} + S_{\text{qft}}(AI) \qquad (3.1)$$

If we use a replica trick and consider the calculation of Renyi entropy $S_A^{(n)}$, the bulk geometry contains a replica wormhole [19, 32] that connects different replicas in a region that becomes I in the $n \to 1$ limit. In this case, the gravitational path integral for integer $n > 1$ have multiple saddle points. Besides the leading saddle point of replica wormhole, there is always a trivial saddle point which is n disconnected copies of the original geometry. Considering the tensor network results, it is natural to conjecture that after the Page time, the Holevo information is given by the difference between the trivial and nontrivial saddle points, which leads to:

$$H_A = S_{\text{trivial}}(A) - S_{\text{actual}}(A) = S_{\text{qft}}(A) - S_{\text{qft}}(AI) - \frac{|\partial I|}{4G_N} \qquad (3.2)$$
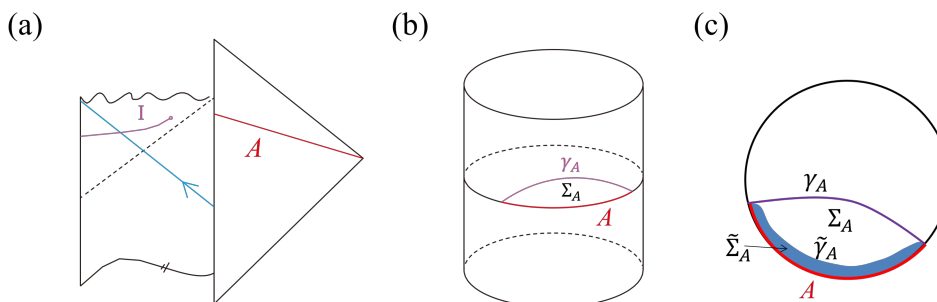
Figure 2. (a) Entanglement island in an AdS black hole coupled with flat space bath. (b) A boundary region A of the AdS space, when there is no bath. (c) A top view of the spatial slice in (b), with the regularized region $\tilde{\Sigma}_A$ and its boundary $\tilde{\gamma}_A$ (see text).

Before the Page time, we get $H_A = 0$. It should be noted that ref. [4] has already proposed that the trivial saddle point corresponds to the entropy of the ensemble-averaged state.

The situation is a bit more complicated in the AdS/CFT case without a bath. If we consider A to be part of the boundary, there is no trivial saddle point in the replica calculation of $S^{(n)}(A)$ since the boundary condition already connects the different replicas. The Holevo information seems to be divergent in this case. This situation is similar to the tensor network model result in eq. (2.10). Naively one expects

$$H_A = S_{\max}(A) - S_{\text{gen}}(A) = \frac{|A| - |\gamma_A|}{4G_N} + S_{\text{qft}}(A) - S_{\text{qft}}(A\Sigma_A) \tag{3.3}$$

Here $S_{\text{qft}}(A)$ is the entropy of A with trivial entanglement wedge. However, since A is co-dimension-2 in the bulk space-time, $S_{\text{qft}}(A)$ is a UV dependent quantity (just like the entropy of an infinitesimal interval in an $1+1$-dimensional CFT), which requires regularization. To regularize this quantity we introduce a region $\tilde{\Sigma}_A$ (see figure 2(c)), which is a small neighborhood of A. For example, we can cover A with balls (in the boundary space) with radius ϵ, and define $\tilde{\Sigma}_A$ as the union of the entanglement wedges of all the ϵ-balls. (see e.g. ref. [37] for a related construction.) Then we can define

$$H_A = S_{\text{gen}}(A\tilde{\Sigma}_A) - S_{\text{gen}}(A\Sigma_A) = \frac{|\tilde{\gamma}_A| - |\gamma_A|}{4G_N} + S_{\text{qft}}\left(A\tilde{\Sigma}_A\right) - S_{\text{qft}}(A\Sigma_A) \tag{3.4}$$

Here $\tilde{\gamma}_A = \partial\left(A\tilde{\Sigma}_A\right)$ is a regularized version of the area of A, and $S_{\text{qft}}\left(A\tilde{\Sigma}_A\right)$ is a regularized version of $S_{\text{qft}}(A)$.

It is interesting to discuss generalizations of our conjecture to systems where there is a subleading saddle point corresponding to a smaller entanglement wedge. For example, one can consider a two-sided black hole corresponding to a thermofield double state $|TFD\rangle$, and then apply a unitary U_R on the right boundary system. This will preserve the entanglement entropy between left and right, but if the unitary increases energy density of the R system, it can lead to a bigger coarse-grained entropy, corresponding to a subleading quantum extremal surface $\tilde{\gamma}_R$ that is closer to the R boundary. This is illustrated in figure 3(a). A simpler situation is two regions in the AdS vacuum (figure 3(b)), which has a connected
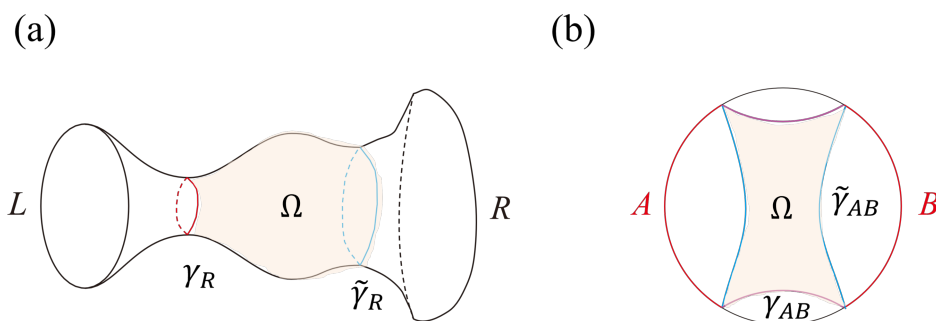
Figure 3. Two example cases with a subleading quantum extremal surface corresponding to a smaller entanglement wedge. (a) A two-sided black hole where the right boundary R has an coarse-grained entropy that is larger than entanglement entropy. (b) Two regions AB for which the actual entanglement wedge is connected, and there is a subleading quantum extremal surface $\tilde{\gamma}_{AB} = \gamma_A \gamma_B$ which is the disjoint HRT surface of A and B separately.

entanglement wedge (when the mutual information $I(A:B)$ is nonzero to the leading order of $\frac{1}{G_N}$). In this case, there is a subleading saddle point corresponding to the disconnected quantum extremal surface $\tilde{\gamma}_{AB} = \gamma_A \cup \gamma_B$. This corresponds to a smaller entanglement wedge $\tilde{\Sigma}_{AB} = \Sigma_A \cup \Sigma_B \subset \Sigma_{AB}$ which is a subset of the actual entanglement wedge.

In such situation, the difference of generalized entropy contribution between these two saddles do not directly corresponds to the Holevo information we have discussed so far, since the subleading saddle point still corresponds to nontrivial connection between different replicas. However, in random tensor networks there is a Holevo information interpretation of this situation. In RTN the random parameters are locally defined at each vertex. Thus we can define a restricted ensemble by denoting J as the random parameters in the tensors in a given bulk region, rather than everywhere. For example in figure 3(a), if we define $\Omega = \Sigma_R \backslash \tilde{\Sigma}_R$ to be the region between γ_R and $\tilde{\gamma}_R$, then averaging over the random tensor parameters in this region, denoted by J_Ω, corresponds to forbidding the entanglement wedge to overlap with this region when optimizing the location of the quantum extremal surface. Thus the optimization will lead to the subleading saddle point $\tilde{\gamma}_R$:

$$S\left(\int dJ_\Omega \rho\left(J_\Omega\right)\right) = \frac{|\tilde{\gamma}_R|}{4G_N} + S_{\text{qft}}\left(\tilde{\Sigma}_R R\right) \tag{3.5}$$

which implies that the restricted ensemble $\rho(J_\Omega)$ has the Holevo information

$$H_R = \frac{|\tilde{\gamma}_R| - |\gamma_R|}{4G_N} + S_{\text{qft}}\left(\tilde{\Sigma}_R R\right) - S_{\text{qft}}\left(\Sigma_R R\right) \tag{3.6}$$

A natural question is what the interpretation of H_R is in quantum gravity. The natural candidate is an ensemble of states that share the same semi-classical geometry and are only different in the bulk region Ω. For example, we can consider different bulk matter states in Ω with the same stress tensor. However, it is important to note that the ensemble we discuss is much bigger than just different bulk QFT states in Ω. For example in the two-interval case in figure 3(b), even if the system is in the AdS vacuum, such that there

are no other states sharing the same stress tensor, the Holevo information is still nonzero:

$$
\begin{aligned}
H_{AB}^{(\Omega)} &= \frac{|\gamma_A| + |\gamma_B| - |\gamma_{AB}|}{4G_N} + S_{\text{qft}}\left(\Sigma_A \cup \Sigma_B\right) - S_{\text{qft}}\left(\Sigma_{AB}\right) \\
&= I(A:B) - I_{\text{qft}}\left(\Sigma_A : \Sigma_B\right)
\end{aligned}
\tag{3.7}
$$

Our interpretation is that the random parameters are not associated with different states in the bulk, but with different mappings from bulk to boundary. Roughly speaking, for a given semi-classical geometry, small fluctuations around this geometry defines a "code subspace" of states which are mapped to the boundary with an isometry [38]. The random parameters in the tensors physically correspond to random parameters in the bulk-to-boundary isometry. In the case of subleading saddle point, the random tensor network model suggests that we can define a two-step isometry

$$
\Omega \longrightarrow \tilde{\gamma}_R \longrightarrow \text{boundary}
\tag{3.8}
$$

and J_Ω are random parameters in the first isometry. Different J_Ω's correspond to different mappings from the same bulk physics in region Ω to different boundary states. One should remember that the boundary Hamiltonian can correspondingly depend on J_Ω, which means the boundary dynamics can appear to be J_Ω independent (at least for simple correlation functions) even if the state and the Hamiltonian depends on J_Ω. We will see this more explicitly in the discussion of the SYK model in next section.

4   Sachdev-Ye-Kitaev model

The Majorana fermion version of SYK model describes N Majorana fermions with the q-body coupling

$$
H_J = i^{q/2} \sum_{j_1 < j_2 < \dots < j_q} J_{j_1 j_2 \dots j_q} \chi_{j_1} \chi_{j_2} \cdots \chi_{j_q}
\tag{4.1}
$$

where $J_{j_1 j_2 \dots j_q}$'s are antisymmetric in the indices, and are independent Gaussian variables with the probability distribution

$$
p_J = \Omega_q^{-1} \exp\left[-\frac{1}{2c_q \mathcal{J}^2} \sum_{j_1 < j_2 < \dots < j_q} J_{j_1 j_2 \dots j_q}^2 \right]
\tag{4.2}
$$

$$
c_q = \frac{2^{q-1}(q-1)!}{N^{q-1} q}, \ \Omega_q = \left(2\pi c_q \mathcal{J}^2\right)^{\frac{1}{2}\binom{N}{q}}
\tag{4.3}
$$

The SYK model has approximate conformal invariant low energy dynamics [23–25, 39] and is proposed to approximately dual to Jackiw-Teitelboim (JT) gravity [40–45]. The random coupling $J_{j_1 j_2 \dots j_q}$ naturally defines an ensemble of states. We will study the Holevo information in this ensemble in two different scenarios.

4.1 Thermal state

The first case we consider is the thermal state

$$\rho_\beta(J) = Z_J^{-1} e^{-\beta H_J} \tag{4.4}$$

In the large N limit, the SYK partition function is self-averaged. More precisely, for an order-1 integer n,

$$\overline{Z_J^n} \equiv \int dJ p_J Z_J^n \tag{4.5}$$

can be written in the path integral of collective variables $G^{ab}(\tau)$ and $\Sigma^{ab}(\tau)$, with $a, b = 1, 2, \ldots, n$. The path integral is dominated by a saddle point which is replica diagonal $G^{ab}(\tau) = G(\tau)\delta^{ab}$. The coupling between different replicas are suppressed by a factor $O(N^{2-q})$ [46]. This leads to

$$\overline{\log Z_J} = \log \overline{Z_J} + O\left(N^{2-q}\right) \tag{4.6}$$

Thus the second term in the Holevo information $\int dJ p_J S\left(\rho_\beta(J)\right)$ self-averages, and is equal to the thermal entropy $S_{\text{th}}(\beta)$ of SYK model.

To determine the first term of the Holevo information, we need to compute

$$\overline{\rho} = \int dJ p_J \rho_\beta(J) \tag{4.7}$$

$\overline{\rho}$ can be determined by symmetry. For each Majorana fermion operator χ_i, we have

$$\chi_i \rho_\beta(J) \chi_i = Z_J^{-1} e^{-\beta \chi_i H_J \chi_i} \tag{4.8}$$

χ_i commutes with terms in H_J that does not contain i, and anti-commutes with terms that contains an i. Denote $\tilde{J}_{j_1 j_2 \ldots j_q}$ by

$$\tilde{J}_{j_1 j_2 \ldots j_q} = \begin{cases} -J_{j_1 j_2 \ldots j_q}, & i \in \{j_1, j_2, \ldots, j_q\} \\ J_{j_1 j_2 \ldots j_q}, & i \notin \{j_1, j_2, \ldots, j_q\} \end{cases} \tag{4.9}$$

Apparently $Z_{\tilde{J}} = Z_J$ and $p_{\tilde{J}} = p_J$. Thus we have

$$\chi_i \rho_\beta(J) \chi_i = \rho_\beta(\tilde{J}) \tag{4.10}$$

$$\chi_i \overline{\rho} \chi_i = \overline{\rho} \tag{4.11}$$

In other words, $\overline{\rho}$ has to commute with all χ_i, which means it has to be proportional to the identity. Thus $\overline{\rho} = 2^{-N/2} \mathbb{I}$ is the maximally mixed state. Therefore we conclude that the Holevo information is given by

$$H_\beta = \frac{N}{2} \log 2 - S_{\text{th}}(\beta) \tag{4.12}$$

This formula is consistent with the random tensor network result in eq. (2.10). The first term is the maximal entropy and the second term is the actual entropy of the state.

In the gravity dual theory, S_β corresponds to the RT entropy given by the dilaton value at the black hole horizon, plus the contribution of the bulk fermion field (cf. [47]). As is illustrated in figure 4(a), the bulk interpretation of $S(\bar{\rho})$ is the generalized entropy of a small interval $\tilde{\Sigma}_R$, given by

$$S_{\text{gen}}\left(\tilde{\Sigma}_R R\right) = \frac{\phi_\epsilon}{4G} + S_{\text{qft}}\left(\tilde{\Sigma}_R\right) \tag{4.13}$$

Both terms on the right-hand side of this equation depends on the UV cutoff in the bulk, but the sum should be independent from the choice of cutoff. $S_{\text{qft}}\left(\tilde{\Sigma}_R\right)$ is the entropy of a single interval for the N bulk Majorana fermions. Interestingly, if we discretize the bulk AdS$_2$ fermion into a $1+1$-dimensional lattice fermion, and take the limit when $\tilde{\Sigma}_R$ is a single lattice site, the entropy $S_{\text{qft}}\left(\tilde{\Sigma}_R\right)$ will reach the maximal value $\frac{N}{2}\log 2$, which suggests that $\phi_\epsilon = 0$ at the lattice level. In other words, the dilaton term can be viewed as purely from integrating out bulk UV fermion modes.

4.2 Thermofield double state

Now we can consider a different scenario where for each J we are given the thermofield double state rather than the thermal state. The thermofield double state is a purification of the thermal state, defined as follows. First we choose a maximally entangled state $|I\rangle$ between two copies of the SYK model, and then define

$$|TFD_\beta(J)\rangle = 2^{N/4}\rho_{\beta L}^{1/2}|I\rangle \tag{4.14}$$

Here $\rho_{\beta L}^{1/2}$ is the square root of the thermal density operator acting on the left system. In the limit $\beta \to 0$, we get $|TFD_0(J)\rangle = |I\rangle$. It is convenient to choose $|I\rangle$ to be the state satisfying

$$(\chi_{iL} + i\chi_{iR})|I\rangle = 0 \tag{4.15}$$

For this choice, the two-point functions satisfy

$$\langle TFD_\beta(J)|\, i\chi_{iL}\chi_{iR}\, |TFD_\beta(J)\rangle = \text{tr}\left[\chi_i\left(\frac{\beta}{2}\right)\chi_i(0)\rho_\beta\right] \equiv G\left(\frac{\beta}{2}\right) \tag{4.16}$$

The average state is defined as

$$\bar{\rho} = \int dJ p_J |TFD_\beta(J)\rangle \langle TFD_\beta(J)| \tag{4.17}$$

The Holevo information is equal to the entropy of $\bar{\rho}$, since $|TFD_\beta(J)\rangle$ is a pure state and the second term vanishes.

To determine $\bar{\rho}$ we notice that in the large-N limit, correlation functions of $\bar{\rho}$ satisfies the Wick theorem (which is the consequence of the saddle point approximation of SYK model). Therefore $\bar{\rho}$ is approximately a Gaussian state, which is thus determined by the two-point function

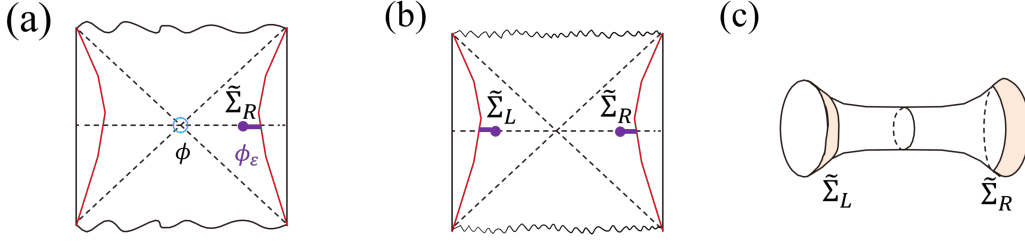$$\text{tr}\left(\bar{\rho}i\chi_{iL}\chi_{iR}\right) = G\left(\frac{\beta}{2}\right) \tag{4.18}$$

Figure 4. The holographic dual interpretation of the Holevo information, for two different ensembles of states. (a) For the ensemble of SYK thermal states, the Holevo information is the generalized entropy of a small interval $\tilde{\Sigma}_R$, minus the thermal entropy. (b) For the ensemble of thermofield double states, the Holevo information is given by the generalized entropy of two small intervals $\Sigma_L \Sigma_R$. (c) An illustration of Σ_L, Σ_R in higher dimensions.

This condition determines $\overline{\rho}$ to be

$$\overline{\rho} = Z_a^{-1} e^{-\lambda \sum_i i\chi_{iL}\chi_{iR}}, \text{ with } \tanh \lambda = G\left(\frac{\beta}{2}\right) \tag{4.19}$$

Thus the Holevo information is

$$H = S\left(\overline{\rho}\right) = -N \left(\frac{1 + G\left(\beta/2\right)}{2} \log \frac{1 + G\left(\beta/2\right)}{2} + \frac{1 - G\left(\beta/2\right)}{2} \log \frac{1 - G\left(\beta/2\right)}{2}\right) \tag{4.20}$$

There is also a more precise way to obtain this result, which we discuss in appendix B.

Now we discuss more about the behavior of Holevo information (4.20) and its gravity interpretation. In the limit $\beta \to 0$, H vanishes because $|TFD_0(J)\rangle = |I\rangle$ is independent from J. In the low temperature limit $\beta \to \infty$, H approaches the maximal value $N \log 2$:

$$H \simeq N \log 2 - NG\left(\frac{\beta}{2}\right)^2 \tag{4.21}$$

The correction term is proportional to $\frac{N}{(\beta\mathcal{J})^{4/q}}$ [39].

In the holographic dual interpretation, naively one would thought the expectation is $S\left(\overline{\rho}\right) = N \log 2$ reaching the maximum, just like in the thermal state case. However, one should remember that in SYK model the bulk theory contains N Majorana fermions (which is required to reproduce the N boundary Majorana fermions). Therefore the correction to area law (i.e. the dilaton value) is order N. We can view the entropy $S(\overline{\rho})$ as the generalized entropy of the two small intervals Σ_L, Σ_R in figure 4(b). (For clarity, a higher dimensional picture of $\tilde{\Sigma}_L, \tilde{\Sigma}_R$ is illustrated in figure 4(c).) The generalized entropy is given by

$$S_{\text{gen}}\left(\tilde{\Sigma}_L\tilde{\Sigma}_R\right) = S_{\text{bulk}}\left(\tilde{\Sigma}_L\tilde{\Sigma}_R\right) + 2\phi_\epsilon \tag{4.22}$$

Here ϕ_ϵ is the dilaton value at the boundary of $\tilde{\Sigma}_L$. We can write $S_{\text{gen}}\left(\tilde{\Sigma}_L\tilde{\Sigma}_R\right)$ in term of mutual information:

$$S_{\text{gen}}\left(\tilde{\Sigma}_L\tilde{\Sigma}_R\right) = S_{\text{gen}}\left(\tilde{\Sigma}_L\right) + S_{\text{gen}}\left(\tilde{\Sigma}_R\right) - I_{\text{bulk}}\left(\tilde{\Sigma}_L : \tilde{\Sigma}_R\right) \tag{4.23}$$

with $S_{\text{gen}}\left(\tilde{\Sigma}_L\right) = \phi_c + S_{\text{bulk}}\left(\tilde{\Sigma}_L\right)$ and similar for $S_{\text{gen}}\left(\tilde{\Sigma}_R\right)$. The mutual information term is only contributed by the bulk fermion. In the limit when the size of $\tilde{\Sigma}_L$ approaches zero, $S_{\text{gen}}(\tilde{\Sigma}_L)$ is the entropy of the averaged thermal state (figure 4(a)), which should reproduce the maximal entropy $\frac{N}{2}\log 2$. The mutual information $I_{\text{bulk}}\left(\tilde{\Sigma}_L : \tilde{\Sigma}_R\right)$ given by bulk fermion is expected to approach $N\log 2 - H$.

It is interesting to note that at large β, the TFD state is also the ground state of the global AdS$_2$ system, which is dual to two SYK models coupled by a bilinear term [48]. Therefore the Holevo information of the ensemble of TFD states can be viewed as an example of the fact that even the AdS vacuum corresponds to a nontrivial ensemble, as we discussed in section 3.

4.3 $G - \Sigma$ action approach to the thermal state

In this subsection, we will revisit the SYK thermal state discussion in section 4.1 with a different approach. We consider the large-N effective action, with collective fields G, Σ. This approach will provide us further understanding on the gravitational saddle point that contributes to the Renyi entropy calculation of the averaged state.

The n-th Renyi entropy of the nonaveraged state $\rho_\beta(J)$ is given by the replica calculation:[3]

$$\int dJ p_J \text{Tr} e^{-n\beta H_J} = \int D\Sigma DG\, \text{Pf}^N(\partial_t - \Sigma) e^{-\frac{N}{2}\int_0^{n\beta} dt_1 \int_0^{n\beta} dt_2 \Sigma(t_1,t_2)G(t_1,t_2) - \frac{J^2}{q}G^q(t_1,t_2)}. \tag{4.24}$$

Notice that since $t_{1,2}$ runs from 0 to $n\beta$, $\Sigma(t_1,t_2)$ and $G(t_1,t_2)$ represent correlations between all the n copies of the replicas. The saddle point of this path integral is just a Euclidean disk with period $n\beta$ which has nonvanishing correlation between different replica copies. Now let's look consider the case of the averaged state. We have:

$$\int dJ_1 \ldots dJ_n p_{J_1} \ldots p_{J_n} \text{Tr} e^{-\beta H(J_1)} \ldots e^{-\beta H(J_n)}$$
$$= \int \prod_{i=1}^n D\Sigma_i DG_i\, \text{Pf}^N(\partial_t - \Sigma) e^{-\frac{N}{2}\sum_i \int_0^\beta dt_1 \int_0^\beta dt_2 \Sigma_i(t_1,t_2)G_i(t_1,t_2) - \frac{J^2}{q}G_i^q(t_1,t_2)}. \tag{4.25}$$

where $\Sigma = \text{diag}(\Sigma_1, \Sigma_2, \ldots, \Sigma_n)$ is a block diagonal matrix of all the Σ_is. However, ∂_t term is not diagonal between replicas.:

$$\text{Pf}(\partial_t - \Sigma) = \int D\psi e^{-\frac{1}{2}\int_0^{n\beta} dt \psi(t)\partial_t\psi(t) + \frac{1}{2}\sum_i \int_0^\beta dt_1^i \int_0^\beta dt_2^i \Sigma_i(t_1,t_2)\psi(t_1^i)\psi(t_2^i)}; \quad t^i = t + (i-1)\beta \tag{4.26}$$

The action (4.25) can be viewed as a modification of the non-averaged one (4.24), with J^2 in $J^2 G^q(t_1,t_2)$ replaced by $J^2\theta(t_1,t_2)$. Here $\theta(t_1,t_2) = 1$ or 0 if t_1, t_2 belongs to the same or different replica, respectively. As a consequence, at the saddle point Σ is always replica diagonal even if G is not diagonal. The Schwinger-Dyson equation is:

$$\begin{aligned} G(t_1,t_2) &= (\partial_t - \Sigma)^{-1}; \\ \Sigma_i(t_1,t_2) &= J^2 G(t_1^i, t_2^i)^{q-1}. \end{aligned} \tag{4.27}$$

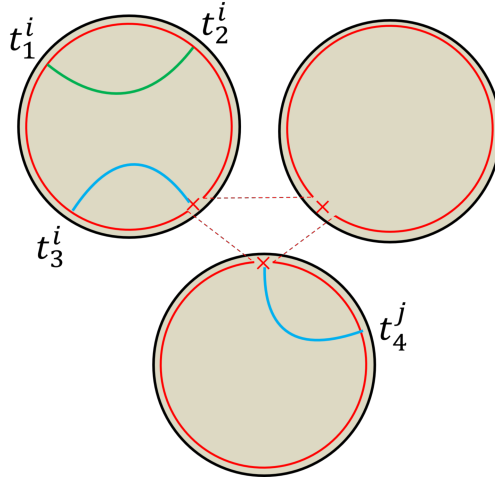[3]Here we do not include Z_J in the denominator since at large N it factorizes out.

Figure 5. Illustration of the Euclidean bulk geometry for n replica of the averaged thermal state. The red circles represent the boundary of each replica with perimeter β. n replica of hyperbolic disks (with $n = 3$ in the figure) are glued at a branching point (red cross) at the boundary. The points connected by the red dashed lines are identified. The green and blue curves are geodesics that contribute to the same-replica two-point function $G\left(t_1^i, t_2^i\right)$ and the across-replica two-point function $G\left(t_3^i, t_4^j\right)$, respectively.

Notice that $(\partial_t - \Sigma)^{-1}$ is a matrix with time range in $(0, n\beta) \times (0, n\beta)$ and the second equation only uses the diagonal blocks of G with t_1^i, t_2^i belonging to the same replica. It is easy to check that the solution of the SD equation is given by:

$$G(t_1^i, t_2^i) = G_\beta(t_1^i, t_2^i), \quad G(t_1^i, t_2^j) = -2G_\beta(t_1^i, 0)G_\beta(t_2^j, 0), \quad i < j, \qquad (4.28)$$

where $G_\beta(t_1, t_2)$ is the thermal correlator. One way to understand this result is that after average of J the operator $\int dJ p(J) e^{-(\beta-t)H(J)} \chi_i e^{-tH}$ needs to be proportional to χ_i due to O(N) symmetry. The proportional constant is uniquely fixed by the thermal two point function:

$$\int dJ p(J) \frac{1}{Z_J} e^{-(\beta-t)H(J)} \chi_i e^{-tH(J)} = 2G_\beta(t, 0)\chi_i. \qquad (4.29)$$

Combing this with the fact that $\int dJ p(J) \frac{1}{Z_J} e^{-\beta H(J)} = 2^{-N/2}\mathbb{I}$, one directly gets:

$$\langle \chi(t_1^i)\chi(t_2^i) \rangle = G_\beta(t_1^i, t_2^i); \quad \langle \chi(t_1^i)\chi(t_2^j) \rangle = -2G_\beta(t_1^i, 0)G_\beta(t_2^j, 0), \quad i < j. \qquad (4.30)$$

Taking the strong coupling limit, the above solution suggests a replica geometry in JT gravity shown in figure 5. On such a geometry, correlators with the same replica index (the green geodesic) is the same as the thermal correlator. Meanwhile, correlators between different replicas (the blue geodesic) will necessary factorize into product of two two-point functions since the bulk geodesic has to pass through the origin to enter a different replica system.[4] Clearly, this is like freezing the disk geometry in JT gravity and only allowing

---

[4]In pure JT such correlators will be exponential small due to the infinite distance of the origin. In SYK there is a natural IR cutoff of the boundary, which leads to a finite correlations between the off diagonal replicas.

the region near the boundary to glue together, which is consistent with the tensor network picture. In the $n \to 1$ limit, the quantum extremal surface is close to the boundary which leads to the maximum entropy in the system (see figure 4(a)).

To evaluate the Renyi entropy directly from $G - \Sigma$ action, we can use the following trick:

$$\partial_{J^2} \mathrm{Tr} \overline{\rho}^n = \partial_{J^2} \left(\frac{1}{Z^n} \int dJ_1 \ldots dJ_n p_{J_1} \ldots p_{J_n} \mathrm{Tr} e^{-\beta H(J_1)} \ldots e^{-\beta H(J_n)} \right) = 0 \qquad (4.31)$$

The variation of $\int dJ_1 \ldots dJ_n p_{J_1} \ldots p_{J_n} \mathrm{Tr} e^{-\beta H(J_1)} \ldots e^{-\beta H(J_n)})$ with respect to J is equal to $\frac{N}{2q} \sum_i \int \int G_i^q(t_1, t_2)$ by equation of motion. Using the on-shell solution that $G_i(t_1, t_2) = G_\beta(t_1, t_2)$, this is equal to $\frac{nN}{2q} \int \int G_\beta^q(t_1, t_2)$. This cancels exactly the variation of Z^n with respect of J which is also equal to $\frac{nN}{2q} \int \int G_\beta^q(t_1, t_2)$. Therefore the Renyi entropy is independent of J and we can simply evaluate the Renyi entropy in the limit of $J = 0$, which gives

$$\mathrm{Tr} \overline{\rho}^n = 2^{-(n-1)N/2}. \qquad (4.32)$$

This confirms our earlier conclusion that $\overline{\rho}$ is a maximally mixed state.

The $G - \Sigma$ action is a direct analog of the Einstein Hilbert action in gravity path integral. In the language of Euclidean gravitational path integral, allowing nontrivial correlation between different replica copies corresponds to allowing short Euclidean paths between replica systems. We see that in the Renyi entropy calculation of the nonaveraged state, correlations between different replica systems are allowed and gives a nontrivial action. In the Renyi entropy calculation of the ensemble averaged state, correlations between different replica systems do not contribute to the action.

The $G - \Sigma$ action approach to the averaged TFD state will be left for future work.

5 Discussion

In this work, we explore the meaning of ensemble average in holographic system from a quantum information perspective. Our basic assumption is that a simple bulk geometry can be represented as an ensemble of microstates $\rho(J)$ parameterized by ensemble parameter J. This assumption is true both in the random tensor network model, where the parameter J is related by the local random projections in the bulk, and in the SYK model and JT gravity where J is a random coupling in the Hamiltonian. Based on random tensor network model, we conjectured that the Holevo information of a boundary region is given by the difference between the generalized entropy of bulk regions with and without entanglement wedge (equation (3.3)):

$$H_A = S\left(\overline{\rho_A}\right) - \overline{S\left(\rho_A\right)} = \frac{|A| - |\gamma_A|}{4G_N} + S_{\mathrm{qft}}\left(A\right) - S_{\mathrm{qft}}(\Sigma_A A). \qquad (5.1)$$

Here A is the boundary region, and Σ_A is the usual entanglement wedge with quantum extremal sruface γ_A. From the perspective of entanglement wedge recontruction, this formula suggests that all the operators in the entanglement wedge are ensemble dependent boundary operators. Notice that even for the operators in the causal wedge, the HKLL reconstruction

depends on the ensemble parameter J through the dependence of the boundary Hamiltonian. We give an explicit check of our conjecture with two SYK examples including the thermal state and the thermofield double state.

In the RTN model, the random parameters are local, associated with each bulk vertex. However, in SYK and JT gravity the J parameter cannot be defined locally in the bulk. Here we would like to discuss a bit the meaning of locality of the J parameters in the bulk. In a random tensor network, the random projections $|V(J)\rangle$, after ensemble average, introduces permutations among the different bulk legs which share the same J parameter. Therefore if the J parameters are local, namely they are independently random at different bulk locations, then the averaging only introduces permutations among different replica systems, without introducing any nontrivial connections within a single copy. Such permutations among replica systems can be thought of as a direct analog of replica wormholes in quantum gravity. The replica wormholes do not change simple correlation functions in a single copy of the system, but do have nontrivial effects on the bulk Hilbert space by decreasing the number of independent states.

If we modify the RTN ansatz by taking the same random state $|V(J)\rangle$ on different vertices, then after ensemble average there will be nontrivial permutations within a single copy of the system. Such permutations change the geometry of the averaged state and correspond to more general types of spacetime wormholes. One type of such wormhole is the one that brings a particle inside the horizon to outside at late time [3]. Such wormholes are direct consequences of the random matrix behavior of the Hamiltonian, and they describe the system's memory of the initial condition. Even after ensemble averaging, such wormholes have observable consequences in a single copy system, such as the late time behavior of the two point functions.

A natural question is whether the nonlocality of J parameters lead to nontrivial modification of our conjecture. In the RTN model, we expect that the additional permutations due to the nonlocality of J parameters are non-perturbatively small corrections to the averaged density matrix because of the large bond dimension of the local EPR pairs in the state $|\Psi_p\rangle$. Therefore we can safely ignore their contributions in the calculation of the Holevo information, and the conjectured formula does not change. This is consistent with the bulk picture that even in the presence of non-local J parameters, there is an approximate notion of locality due to the non-perturbative nature of the wormholes. What we learn from the tensor network story is that the suppression of such non-perturbative contributions stems from the entanglement structure of the parent state ρ_P.

We would like to conclude the discussion with a few words about the situation when such non-local effects become significant. This occurs when the wormholes proliferate. Consider a late time thermofield double state, whose interior can be approximated by a very long cylinder. When the length of the cylinder is longer than exponential of the transverse area, the wormhole will have large fluctuations by splitting and joining different locations of the cylinder. This completely breaks the notion of locality in the interior. To say it differently, the interior geometry at such late time is no longer well defined and one cannot trust the classical geometric description. As a toy model, consider in the RTN model a geometry with L identical projections $|V(J)\rangle$ with bond dimension D. This can be

understood as a cylinder with length L and transverse area $\log D$. The average of tensor product $|V(J)\rangle\,\langle V(J)|^{\otimes L}$ results in a projection to the permutation-symmetric subspace of the L bulk vertices. When each site has Hilbert space dimension d, the permutation symmetric subspace has the dimension

$$D_L = \binom{L + d - 1}{L} \approx \begin{cases} d^L; & L \ll d; \\ L^d; & d \ll L. \end{cases} \tag{5.2}$$

In the simple model without bulk field, the vertex dimension is $d = D^2$. In the limit of $L \gg D^2$, the permutation symmetric subspace dimension D_L is polynomial in L, which is much smaller than the dimension before projection. This implies that the ensemble averaging has a strong effect to single copy physical properties, although the argument is not rigorous. Notice that this is a truncation of the naive bulk Hilbert space in the averaged state, which is different from the truncation of the black hole Hilbert space from replica wormhole. For instance, the effects discussed here will lead to an order-one change of simple correlation functions and the experience of an infalling observer. It will be interesting to study if there exists a new geometric description of the averaged state at late time, which is related to the typical state firewall paradox [49]. The breakdown of the interior geometry at an exponentially late time is also expected as a consequence of the complexity-equals-to-volume conjecture [50–52]. For a recent discussion about the behavior of the volume in JT gravity, see [53, 54].

Acknowledgments

A   Some more details about random tensor networks

In this appendix, we will discuss some more details about the replica calculation of the averaged entropy $\int dJ p_J S\left(\rho_A(J)\right)$ for the ensemble of RTN given in eq. (2.1) and (2.2). For convenience we write these equations again here in a slightly different way:

$$\pi_S(J) = \mathrm{tr}_B\left(|V(J)\rangle\,\langle V(J)|\,\rho_P\right) \tag{A.1}$$

$$p_J = \mathrm{tr}\left(\pi_S(J)\right) \tag{A.2}$$

$$\rho_S(J) = p_J^{-1}\pi_S(J) \tag{A.3}$$

Here we have defined the unnormalized state $\pi_S(J)$.

Using the standard replica trick

$$S(\rho) = -\left.\frac{\partial}{\partial n}\log\mathrm{tr}\left(\rho^n\right)\right|_{n\to 1} \tag{A.4}$$

we have

$$
\begin{aligned}
\int dJ p_J S\left(\rho_A(J)\right) &= -\int dJ p_J \frac{\partial}{\partial n}\log\mathrm{tr}\left(\rho_A(J)^n\right)\Big|_{n\to 1}\\
&= -\int dJ p_J \frac{\partial}{\partial n}\log\mathrm{tr}\left(p_J^{-1}\pi_A(J)^n\right)\Big|_{n\to 1}\\
&= -\int dJ p_J \frac{\partial}{\partial n}\left[\log\mathrm{tr}\left(\pi_A(J)^n\right) - n\log p_J\right]\Big|_{n\to 1}\\
&= -\int dJ p_J \left[\frac{1}{\mathrm{tr}\left(\pi_A(J)^n\right)}\frac{\partial\mathrm{tr}\left(\pi_A(J)^n\right)}{\partial n} - \frac{1}{p_J^n}\frac{\partial p_J^n}{\partial n}\right]\Big|_{n\to 1}
\end{aligned}
\tag{A.5}
$$

Taking $n\to 1$ the denominator in both terms of the last line is p_J, so that

$$\int dJ p_J S\left(\rho_A(J)\right) = -\frac{\partial}{\partial n}\left[\int dJ\mathrm{tr}\left(\pi_A(J)^n\right) - \int dJ\left(\mathrm{tr}\left(\pi_A(J)\right)\right)^n\right]\Big|_{n\to 1} \tag{A.6}$$

Define

$$Z_A^{(n)} = \int dJ\mathrm{tr}\left(\pi_A(J)^n\right) \tag{A.7}$$

$$Z_\emptyset^{(n)} = \int dJ\left(\mathrm{tr}\left(\pi_A(J)\right)\right)^n = \int dJ p_J^n \tag{A.8}$$

we notice that $Z_A^{(1)} = Z_\emptyset^{(1)} = 1$. Thus we can further transform eq. (A.6) into

$$
\begin{aligned}
\int dJ p_J S\left(\rho_A(J)\right) &= -\frac{\partial}{\partial n}\left[\log Z_A^{(n)} - \log Z_\emptyset^{(n)}\right]\Big|_{n\to 1}\\
&= -\frac{\partial}{\partial n}\log\frac{Z_A^{(n)}}{Z_\emptyset^{(n)}}\Big|_{n\to 1} = -\frac{\partial}{\partial n}\frac{Z_A^{(n)}}{Z_\emptyset^{(n)}}\Big|_{n\to 1}
\end{aligned}
\tag{A.9}
$$

We prefer the last expression because $Z_A^{(n)}$ and $Z_\emptyset^{(n)}$ have the interpretation of a classical statistical model partition function, and

$$-\log\frac{Z_A^{(n)}}{Z_\emptyset^{(n)}} \tag{A.10}$$

is the free energy cost of the spin model caused by changing the boundary condition in A region. The spin model comes from the fact that $\pi_S(J)$ is linear in $|V(J)\rangle\langle V(J)|$, and

$$\int dJ\,|V(J)\rangle\langle V(J)|^{\otimes n} = c_n^{-1}\sum_{g\in S^n} X_g \tag{A.11}$$

with g an element of the permutation group S^n, and X_g is the permutation g acting naturally by permuting different replica. c_n is a normalization constant. The explicit

expression of $Z_A^{(n)}$ and $Z_\emptyset^{(n)}$ is [33]

$$Z_A^{(n)} = c_n^{-V} \sum_{\{g_x\},\ g_x \in S^n} \text{tr}\left(\rho_{AB}^{\otimes n} \otimes_{x \in B} X_{g_x x} \otimes X_{An}\right) \tag{A.12}$$

$$Z_\emptyset^{(n)} = c_n^{-V} \sum_{\{g_x\},\ g_x \in S^n} \text{tr}\left(\rho_{AB}^{\otimes n} \otimes_{x \in B} X_{g_x x}\right) \tag{A.13}$$

It is interesting to note that the derivation so far has not used any specific property of ρ_P. The averaged entropy $\int dJ p_J S(\rho_A(J))$ is directly related to the analytic continuation of the free energy of the spin model. On comparison, the quantity discussed in ref. [33] is the average $\int dJ S^{(n)}(\rho_A(J))$ which only corresponds to $-\frac{1}{n-1} \log \frac{Z_A^{(n)}}{Z_\emptyset^{(n)}}$ in the large bond dimension limit when the numerator and denominator are self-averaging. In other words, going beyond the self-averaging limit, the averaged entropy $\int dJ p_J S(\rho_A(J))$ provides a physical interpretation of the free energy cost $-\log \frac{Z_A^{(n)}}{Z_\emptyset^{(n)}}$.

B Averaged state of the thermofield double state of SYK model

In this appendix, we carry a more detailed discussion on the averaged state of the SYK model TFD state. It can be seen that $\overline{\rho}$ is invariant under $SO(N)$ rotation that rotate χ_i as a vector $\chi_i \to O_{ij}\chi_j$, since such rotation is equivalent to $SO(N)$ transformation of the coefficients

$$J_{i_1 i_2 \ldots i_n} \to J_{j_1 j_2 \ldots j_n} O_{j_1 i_1} O_{j_2 i_2} \ldots O_{j_n i_n} \tag{B.1}$$

which preserves the probability p_J. The only $SO(N)$ scalar operators are

$$\hat{L} = \sum_i \frac{i\chi_{iL}\chi_{iR} + 1}{2}, \ F_L = i^{N/2}\chi_{1L}\chi_{2L}\ldots\chi_{NL}, \ F_R = i^{N/2}\chi_{1R}\chi_{2R}\ldots\chi_{NR} \tag{B.2}$$

In addition, eq. (4.10) translates to a property of the TFD state:

$$\chi_{iL}\chi_{iR} |TFD(J)\rangle \langle TFD(J)| \chi_{iR}\chi_{iL} = \left|TFD(\tilde{J})\right\rangle \left\langle TFD(\tilde{J})\right| \tag{B.3}$$

$$\Rightarrow [i\chi_{iL}\chi_{iR}, \overline{\rho}] = 0 \tag{B.4}$$

This symmetry condition further excludes F_L and F_R, so that $\overline{\rho}$ is a function of $\hat{L}$. $\overline{\rho}$ can be written in the eigenstate subspaces of $\hat{L}$ as

$$\overline{\rho} = \sum_{n=0}^N p_n \frac{\hat{\Pi}_n}{D_n} \tag{B.5}$$

with $\hat{\Pi}_n$ the projector to the subspace of $\hat{L} = n$, with the dimension $D_n = \begin{pmatrix} N \\ n \end{pmatrix}$. Thus the entropy is given by

$$S\left(\overline{\rho}\right) = \sum_{n=0}^N \left(-p_n \log p_n + p_n \log D_n\right) \tag{B.6}$$

p_n can be computed using a generating function:

$$F(\theta) \equiv \sum_n p_n e^{in\theta} = \text{tr}\left(\overline{\rho}\hat{\Pi}_n\right) e^{in\theta} = \text{tr}\left(\overline{\rho}e^{i\hat{L}\theta}\right) \tag{B.7}$$

If we analytically continue θ to $i\mu$, $F(i\mu)$ can be computed as a partition function of the coupled SYK model:

$$F(i\mu) = \langle TFD(J)|\, e^{-\mu\hat{L}}\, |TFD(J)\rangle = \frac{\mathcal{Z}_\mu}{\mathcal{Z}_0} \tag{B.8}$$

which is self-averaging. $F(i\mu)$ can be computed explicitly in the large-q limit of SYK model [58][5] Without going into detail of $F(i\mu)$, we know that in the large-N limit it has the form

$$F(i\mu) = e^{-N\mathcal{A}(\mu)} \tag{B.9}$$

with $\mathcal{A}(\mu)$ a finite function of μ. Thus according to eq. (B.7) we have

$$p_n = \int_0^{2\pi} \frac{d\theta}{2\pi} F(\theta) e^{-in\theta} = \int_0^{2\pi} \frac{d\theta}{2\pi} e^{-N\mathcal{A}(-i\theta)-in\theta} \tag{B.10}$$

In the large-N limit, we can expand $\mathcal{A}(\mu)$ into second order of μ:

$$\mathcal{A}(\mu) \simeq p\mu - \frac{1}{2}\alpha\mu^2 \tag{B.11}$$

with

$$p = \frac{\langle\hat{L}\rangle}{N} = \frac{1}{2}\left(G\left(\frac{\beta}{2}\right)+1\right) \tag{B.12}$$

$$\alpha = \frac{1}{N}\left(\langle\hat{L}^2\rangle - \langle\hat{L}\rangle^2\right) \tag{B.13}$$

This leads to

$$p_n \simeq C\exp\left[-\frac{N}{2\alpha}\left(p-\frac{n}{N}\right)^2\right] \tag{B.14}$$

In the large-N limit, p_n is strongly peaked, which has the entropy (B.6)

$$S\left(\overline{\rho}\right) \simeq \log D_{pN} + O(1) \simeq N\left[-p\log p - (1-p)\log(1-p)\right] + O(1) \tag{B.15}$$

This agrees with the Gaussian state result (4.20).

---

[5]See eq. (5.1) of ref. [58].

References

[1] P. Saad, S.H. Shenker and D. Stanford, *JT gravity as a matrix integral*, arXiv:1903.11115 [INSPIRE].

[2] D. Stanford and E. Witten, *JT gravity and the ensembles of random matrix theory*, *Adv. Theor. Math. Phys.* **24** (2020) 1475 [arXiv:1907.03363] [INSPIRE].

[3] P. Saad, *Late Time Correlation Functions, Baby Universes, and ETH in JT Gravity*, arXiv:1910.10311 [INSPIRE].

[4] R. Bousso and E. Wildenhain, *Gravity/ensemble duality*, *Phys. Rev. D* **102** (2020) 066005 [arXiv:2006.16289] [INSPIRE].

[5] D. Marolf and H. Maxfield, *Transcending the ensemble: baby universes, spacetime wormholes, and the order and disorder of black hole information*, *JHEP* **08** (2020) 044 [arXiv:2002.08950] [INSPIRE].

[6] D. Stanford, *More quantum noise from wormholes*, arXiv:2008.08570 [INSPIRE].

[7] J. Pollack, M. Rozali, J. Sully and D. Wakeham, *Eigenstate Thermalization and Disorder Averaging in Gravity*, *Phys. Rev. Lett.* **125** (2020) 021601 [arXiv:2002.02971] [INSPIRE].

[8] N. Afkhami-Jeddi, H. Cohn, T. Hartman and A. Tajdini, *Free partition functions and an averaged holographic duality*, *JHEP* **01** (2021) 130 [arXiv:2006.04839] [INSPIRE].

[9] A. Maloney and E. Witten, *Averaging over Narain moduli space*, *JHEP* **10** (2020) 187 [arXiv:2006.04855] [INSPIRE].

[10] A. Belin and J. de Boer, *Random statistics of OPE coefficients and Euclidean wormholes*, *Class. Quant. Grav.* **38** (2021) 164001 [arXiv:2006.05499] [INSPIRE].

[11] J. Cotler and K. Jensen, *AdS$_3$ gravity and random CFT*, *JHEP* **04** (2021) 033 [arXiv:2006.08648] [INSPIRE].

[12] Y. Chen and H.W. Lin, *Signatures of global symmetry violation in relative entropies and replica wormholes*, *JHEP* **03** (2021) 040 [arXiv:2011.06005] [INSPIRE].

[13] P.-S. Hsin, L.V. Iliesiu and Z. Yang, *A violation of global symmetries from replica wormholes and the fate of black hole remnants*, *Class. Quant. Grav.* **38** (2021) 194004 [arXiv:2011.09444] [INSPIRE].

[14] A. Belin, J. De Boer, P. Nayak and J. Sonner, *Charged Eigenstate Thermalization, Euclidean Wormholes and Global Symmetries in Quantum Gravity*, arXiv:2012.07875 [INSPIRE].

[15] A. Milekhin and A. Tajdini, *Charge fluctuation entropy of Hawking radiation: a replica-free way to find large entropy*, arXiv:2109.03841 [INSPIRE].

[16] S.R. Coleman, *Black Holes as Red Herrings: Topological Fluctuations and the Loss of Quantum Coherence*, *Nucl. Phys. B* **307** (1988) 867 [INSPIRE].

[17] J.M. Maldacena and L. Maoz, *Wormholes in AdS*, *JHEP* **02** (2004) 053 [hep-th/0401024] [INSPIRE].

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

[19] G. Penington, S.H. Shenker, D. Stanford and Z. Yang, *Replica wormholes and the black hole interior*, arXiv:1911.11977 [INSPIRE].

[20] Y. Chen, V. Gorbenko and J. Maldacena, *Bra-ket wormholes in gravitationally prepared states*, *JHEP* **02** (2021) 009 [arXiv:2007.16091] [INSPIRE].

[21] A. Almheiri, R. Mahajan, J. Maldacena and Y. Zhao, *The Page curve of Hawking radiation from semiclassical geometry*, *JHEP* **03** (2020) 149 [arXiv:1908.10996] [INSPIRE].

[22] S. Sachdev and J. Ye, *Gapless spin fluid ground state in a random, quantum Heisenberg magnet*, *Phys. Rev. Lett.* **70** (1993) 3339 [cond-mat/9212030] [INSPIRE].

[23] A. Kitaev, *Hidden correlations in the hawking radiation and thermal noise*, talk given at *Fundamental Physics Prize Symposium*, (2014) [www.youtube.com/watch?v=OQ9qN8j7EZI].

[24] A. Kitaev, *A simple model of quantum holography (part 1)*, talk at *KITP*, April 7, 2015, http://online.kitp.ucsb.edu/online/entangled15/kitaev/.

[25] A. Kitaev, *A simple model of quantum holography (part 2)*, talk at *KITP*, May 27, 2015, http://online.kitp.ucsb.edu/online/entangled15/kitaev2/.

[26] A.S. Holevo, *Bounds for the quantity of information transmitted by a quantum communication channel*, *Prob. Peredachi Info.* **9** (1973) 3.

[27] N. Bao and H. Ooguri, *Distinguishability of black hole microstates*, *Phys. Rev. D* **96** (2017) 066017 [arXiv:1705.07943] [INSPIRE].

[28] N. Bao, J. Harper and G.N. Remmen, *Holevo information of black hole mesostates*, *Phys. Rev. D* **105** (2022) 026010 [arXiv:2103.06888] [INSPIRE].

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

[30] G. Penington, *Entanglement Wedge Reconstruction and the Information Paradox*, *JHEP* **09** (2020) 002 [arXiv:1905.08255] [INSPIRE].

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

[32] A. Almheiri, T. Hartman, J. Maldacena, E. Shaghoulian and A. Tajdini, *Replica Wormholes and the Entropy of Hawking Radiation*, *JHEP* **05** (2020) 013 [arXiv:1911.12333] [INSPIRE].

[33] P. Hayden, S. Nezami, X.-L. Qi, N. Thomas, M. Walter and Z. Yang, *Holographic duality from random tensor networks*, *JHEP* **11** (2016) 009 [arXiv:1601.01694] [INSPIRE].

[34] X. Dong, X.-L. Qi, Z. Shangnan and Z. Yang, *Effective entropy of quantum fields coupled with gravity*, *JHEP* **10** (2020) 052 [arXiv:2007.02987] [INSPIRE].

[35] C.H. Bennett, G. Brassard, C. Crepeau, R. Jozsa, A. Peres and W.K. Wootters, *Teleporting an unknown quantum state via dual classical and Einstein-Podolsky-Rosen channels*, *Phys. Rev. Lett.* **70** (1993) 1895 [INSPIRE].

[36] E. Witten, *A Mini-Introduction To Information Theory*, *Riv. Nuovo Cim.* **43** (2020) 187 [arXiv:1805.11965] [INSPIRE].

[37] B. Czech, P. Hayden, N. Lashkari and B. Swingle, *The Information Theoretic Interpretation of the Length of a Curve*, *JHEP* **06** (2015) 157 [arXiv:1410.1540] [INSPIRE].

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

[39] J. Maldacena and D. Stanford, *Remarks on the Sachdev-Ye-Kitaev model*, *Phys. Rev. D* **94** (2016) 106002 [arXiv:1604.07818] [INSPIRE].

[40] R. Jackiw, *Lower Dimensional Gravity*, *Nucl. Phys. B* **252** (1985) 343 [INSPIRE].

[41] C. Teitelboim, *Gravitation and Hamiltonian Structure in Two Space-Time Dimensions*, *Phys. Lett. B* **126** (1983) 41 [INSPIRE].

[42] A. Almheiri and J. Polchinski, *Models of AdS₂ backreaction and holography*, *JHEP* **11** (2015) 014 [arXiv:1402.6334] [INSPIRE].

[43] K. Jensen, *Chaos in AdS₂ Holography*, *Phys. Rev. Lett.* **117** (2016) 111601 [arXiv:1605.06098] [INSPIRE].

[44] J. Maldacena, D. Stanford and Z. Yang, *Conformal symmetry and its breaking in two dimensional Nearly Anti-de-Sitter space*, *PTEP* **2016** (2016) 12C104 [arXiv:1606.01857] [INSPIRE].

[45] J. Engelsöy, T.G. Mertens and H. Verlinde, *An investigation of AdS₂ backreaction and holography*, *JHEP* **07** (2016) 139 [arXiv:1606.03438] [INSPIRE].

[46] A. Kitaev and S.J. Suh, *The soft mode in the Sachdev-Ye-Kitaev model and its gravity dual*, *JHEP* **05** (2018) 183 [arXiv:1711.08467] [INSPIRE].

[47] Y. Chen and P. Zhang, *Entanglement Entropy of Two Coupled SYK Models and Eternal Traversable Wormhole*, *JHEP* **07** (2019) 033 [arXiv:1903.10532] [INSPIRE].

[48] J. Maldacena and X.-L. Qi, *Eternal traversable wormhole*, arXiv:1804.00491 [INSPIRE].

[49] D. Marolf and J. Polchinski, *Gauge/Gravity Duality and the Black Hole Interior*, *Phys. Rev. Lett.* **111** (2013) 171301 [arXiv:1307.4706] [INSPIRE].

[50] L. Susskind, *The Typical-State Paradox: Diagnosing Horizons with Complexity*, *Fortsch. Phys.* **64** (2016) 84 [arXiv:1507.02287] [INSPIRE].

[51] A.R. Brown, D.A. Roberts, L. Susskind, B. Swingle and Y. Zhao, *Holographic Complexity Equals Bulk Action?*, *Phys. Rev. Lett.* **116** (2016) 191301 [arXiv:1509.07876] [INSPIRE].

[52] A.R. Brown, D.A. Roberts, L. Susskind, B. Swingle and Y. Zhao, *Complexity, action, and black holes*, *Phys. Rev. D* **93** (2016) 086006 [arXiv:1512.04993] [INSPIRE].

[53] Z. Yang, *The Quantum Gravity Dynamics of Near Extremal Black Holes*, *JHEP* **05** (2019) 205 [arXiv:1809.08647] [INSPIRE].

[54] L.V. Iliesiu, M. Mezei and G. Sárosi, *The volume of the black hole interior at late times*, arXiv:2107.06286 [INSPIRE].

[55] A. Almheiri and H.W. Lin, *The Entanglement Wedge of Unknown Couplings*, arXiv:2111.06298 [INSPIRE].

[56] R. Renner and J. Wang, *The black hole information puzzle and the quantum de Finetti theorem*, arXiv:2110.14653 [INSPIRE].

[57] X.-L. Qi, *Holevo information and ensemble gravity*, talk on *The Geoflow collaboration meeting*, September 29, 2021, https://youtu.be/IAnaS1Q4CSw.

[58] X.-L. Qi and A. Streicher, *Quantum Epidemiology: Operator Growth, Thermal Effects, and SYK*, *JHEP* **08** (2019) 012 [arXiv:1810.11958] [INSPIRE].