Machine learning statistical gravity from multi-region entanglement entropy

Jonathan Lam◉ and Yi-Zhuang You◉

Department of Physics, University of California, San Diego, California 92093, USA

The Ryu-Takayanagi formula directly connects quantum entanglement and geometry. Yet the assumption of static geometry lead to an exponentially small mutual information between far-separated disjoint regions, which does not hold in many systems such as free fermion conformal field theories. In this paper, we proposed a microscopic model by superimposing entanglement features of an ensemble of random tensor networks of different bond dimensions, which can be mapped to a statistical gravity model consisting of a massive scalar field on a fluctuating background geometry. We propose a machine-learning algorithm that recovers the underlying geometry fluctuation from multi-region entanglement entropy data by modeling the bulk geometry distribution via a generative neural network. To demonstrate its effectiveness, we tested the model on a free fermion system and showed mutual information can be mediated effectively by geometric fluctuation. Remarkably, locality emerged from the learned distribution of bulk geometries, pointing to a local statistical gravity theory in the holographic bulk.

I. INTRODUCTION

The holographic duality [1–5] is a duality between boundary d-dimensional quantum field theories and bulk $(d + 1)$-dimensional gravitational theories in asymptotically anti-de Sitter (AdS) space. It provides an appealing explanation for the emergence of spacetime geometry from quantum entanglement [6–16]. The connection is manifested in the Ryu-Takayanagi (RT) formula [17,18] $S(A) = \frac{1}{4G_N} \min_{\gamma_A} |\gamma_A|$ that relates the entanglement entropy $S(A)$ of a boundary region A to the area of the extremal surface γ_A in the bulk that is homologous to the same region A. Progress has been made to reconstruct the bulk geometry from the boundary data in terms of geodesic lengths [19–22], extremal areas [23–25] or entanglement entropies [26,27]. A majority of the effort has been focused on reconstructing a classical geometry from *single-region* entanglement entropies (or independent extremal surfaces). However, *multi-region* entanglement entropies further encode the correlation among multiple extremal surfaces, which could reveal how the bulk geometry fluctuations around its classical background (assuming a semiclassical description of the bulk gravity). In this paper, we will explore the possibility to extract information about fluctuating holographic bulk geometries from multi-region entanglement entropies of a quantum system using generative models in machine learning.

A feature of the holography entanglement entropy based on the RT formula is that the mutual information $I_{A:B} = S_A + S_B - S_{AB}$ vanishes between two disjoint boundary regions A and B that are far separated from each other [28,29], because the minimum surface enclosing the combined region AB will be a disjoint union of γ_A and γ_B such that the entropies simply add up as $S_{AB} = S_A + S_B$, leaving no room for mutual information. While the vanishing mutual information is a correct feature of holographic conformal field theories (CFT), it is not generally the case for many other quantum systems (e.g., free-fermion CFT). One idea to remedy the problem is to introduce bulk matter fields to mediate the mutual information [30–32]. Another possibility is to consider statistical fluctuations of bulk geometries such that γ_A and γ_B are correlated to produce the finite mutual information. The statistical gravitational fluctuation may be viewed as an effective description arising from tracing out bulk matter fields. We will further explore the second possibility of fluctuating geometry using a concrete model of random tensor network (RTN) [33,34] with fluctuating bond dimensions. The bond dimension fluctuation translates to the bulk geometry fluctuation in the context of tensor network holography [35–37], which is presumably governed by some statistical gravity model.

However, it is unclear what should be the appropriate bulk statistical gravity model that best reproduces the entanglement feature of a given quantum system on the boundary. To address this challenge, we propose to apply data-driven and machine-learning approaches to uncover the statistical gravity model behind the observational data of quantum many-body entanglement. What needs to be learned is the joint probability distribution of bond dimensions (or bulk geometries). Generative models [38–44] in machine learning provide us precisely the tool to learn unknown distributions from data. In particular, we apply a deep generative model [39,41] to describe the bulk geometry fluctuation. We train the model by matching the model predictions of multi-region entanglement entropies with their actual values evaluated in the given
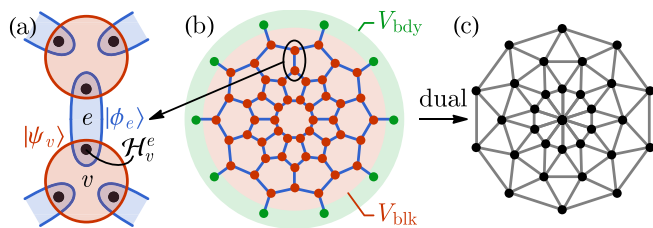
FIG. 1. (a) Details of the RTN near an edge. (b) The planar graph G on which the RTN is defined. The vertices are classified into the bulk vertices (in red) and the boundary vertices (in green). (c) The dual graph $\tilde{G}$ of the RTN graph G.

quantum many-body state. After training, the generative model should tell us the statistical gravity model that emerges from learning.

The approach developed in this work extends the general idea of entanglement feature learning [26], which aims to reconstruct the bulk geometry by learning from the entanglement data on the boundary. Compare to the previous work, this paper makes significant progress in including the gravitational fluctuation in the model, which will enable us to learn an emergent gravity theory rather than a static classical geometry. We will focus on (1+1)D quantum systems, and assume that the system admits an approximate semiclassical geometry description in the holographic bulk. Based on a random tensor network model with fluctuating bond dimensions, we first establish a holographic model for quantum entanglement involving a scalar matter field on a statistically fluctuating spatial geometry. Applying our approach to a free-fermion CFT state with a large central charge, we uncover a statistical gravity model governed by Weyl field fluctuations propagating on the hyperbolic background geometry. We show that the Weyl field fluctuation has the emergent bulk locality by studying its bulk correlation. We further analyze the spectrum and the leading collective modes of the emergent gravity theory. We also show that the matter field mass gets renormalized by the gravitational fluctuation as expected.

II. HOLOGRAPHIC MODELS OF ENTANGLEMENT

A. Random tensor network model

The random tensor network (RTN) model is an intuitive toy model for holographic duality, which directly connects quantum states and emergent geometries. The original proposal [33] of RTN assumes a fixed bond dimension on every link of the tensor network. It can be generalized to include bond dimension fluctuations (or more precisely, bond entanglement fluctuations) [34,45]. The generalized RTN model in consideration is defined as follows: (i) A planar graph $G = (V, E)$ is given to describe the background network geometry, where V denotes the vertex set and E denotes the edge set. $V = V_{\mathrm{blk}} \cup V_{\mathrm{bdy}}$ is divided into two subsets: the bulk V_{blk} and the boundary V_{bdy} sets, see Fig. 1(b). (ii) A local Hilbert space $\mathcal{H}_v^e$ is associated with each pair (v, e) of vertex $v \in V$ and its adjacent edge $e \in E$ (for e not adjacent to v, the associated Hilbert space is considered trivial $\mathcal{H}_v^e \cong \mathbb{C}$), see Fig. 1(a). (iii) A random pure state $|\psi_v\rangle \in \mathcal{H}_v \equiv \bigotimes_{e \in \mathrm{d}v} \mathcal{H}_v^e$ is defined on every bulk vertex $v \in V_{\mathrm{blk}}$. (iv) A random

entangled state $|\phi_e\rangle \in \mathcal{H}^e \equiv \bigotimes_{v \in \partial e} \mathcal{H}_v^e$ is defined across every edge $e \in E$. (v) RTN defines an ensemble $\mathcal{E}_{\mathrm{RTN}} = \{|\Psi\rangle\}$ of pure states in the boundary Hilbert space $\mathcal{H}_{\mathrm{bdy}} \equiv \bigotimes_{v \in V_{\mathrm{bdy}}} \mathcal{H}_v$ by taking a (partial) projection in the bulk Hilbert space $\mathcal{H}_{\mathrm{blk}} \equiv \bigotimes_{v \in V_{\mathrm{blk}}} \mathcal{H}_v$ as

$$|\Psi\rangle = \langle\psi|\phi\rangle : |\psi\rangle = \bigotimes_{v \in V_{\mathrm{blk}}} |\psi_v\rangle, \ |\phi\rangle = \bigotimes_{e \in E} |\phi_e\rangle. \quad (1)$$

The probability measure of $|\Psi\rangle$ in the RTN ensemble $\mathcal{E}_{\mathrm{RTN}}$ is given by $P(|\Psi\rangle) = P(|\psi\rangle)P(|\phi\rangle)$. The vertex state distribution $P(|\psi\rangle) = \prod_{v \in V_{\mathrm{blk}}} P(|\psi_v\rangle)$ is assumed to be factorized, and on each vertex, the distribution $P(|\psi_v\rangle)$ is taken to be the Haar measure (i.e., uniform random states in $\mathcal{H}_v$). The edge (link) state distribution $P(|\phi\rangle)$ is generally a nontrivial joint distribution depending on all $|\phi_e\rangle$ on all edges, which allows the quantum entanglement across different edges to fluctuate collectively.

For any operator $O^{(k)}$ defined in k copies of the boundary Hilbert space $\mathcal{H}_{\mathrm{bdy}}^{\otimes k}$, its expectation value in the product state $|\Psi\rangle^{\otimes k}$ is defined to be

$$\langle O^{(k)}\rangle = \mathop{\mathbb{E}}_{|\Psi\rangle \in \mathcal{E}_{\mathrm{RTN}}} \frac{\mathrm{Tr}((|\Psi\rangle\langle\Psi|)^{\otimes k} O^{(k)})}{\langle\Psi|\Psi\rangle^k}. \quad (2)$$

We assume that the correlation between denominator and numerator is not important (which is generally valid in the semiclassical regime when fluctuations are weak), so that we can approximate the ensemble average of the ratio by the ratio of separate averages,

$$\langle O^{(k)}\rangle \simeq \frac{1}{\mathcal{N}_k} \mathop{\mathbb{E}}_{|\Psi\rangle \in \mathcal{E}_{\mathrm{RTN}}} \mathrm{Tr}((|\Psi\rangle\langle\Psi|)^{\otimes k} O^{(k)}), \quad (3)$$

where $\mathcal{N}_k = \mathbb{E}_{|\Psi\rangle \in \mathcal{E}_{\mathrm{RTN}}} \langle\Psi|\Psi\rangle^k$ is the kth moment of the state norm squared. For example, the 2nd Rényi entanglement entropy S_A (or more precisely, the purity e^{-S_A}) of RTN states in a boundary region A can be calculated by taking $k = 2$ and $O^{(k)} = X_A$ (the swap operator supported in region A),

$$e^{-S_A} \propto \mathop{\mathbb{E}}_{|\Psi\rangle \in \mathcal{E}_{\mathrm{RTN}}} \mathrm{Tr}((|\Psi\rangle\langle\Psi|)^{\otimes 2} X_A). \quad (4)$$

We will suppress the Rényi index throughout this paper, and use S_A to denote the 2nd Rényi entropy. The RTN model provides an effective description of entanglement entropies of typical quantum states on the holographic boundary, given the background geometry G together with fluctuations of states $|\psi_v\rangle, |\phi_e\rangle$ in the holographic bulk.

It worth mention that in modeling the 2nd Rényi entanglement entropy by Eq. (4), the average over the RTN ensemble $\mathcal{E}_{\mathrm{RTN}}$ is taken neither on the state vector level (i.e., not a pure state superposition $\mathbb{E}_{|\Psi\rangle \in \mathcal{E}_{\mathrm{RTN}}} |\Psi\rangle$), nor on the density matrix level (i.e., not a mixed state superposition $\mathbb{E}_{|\Psi\rangle \in \mathcal{E}_{\mathrm{RTN}}} |\Psi\rangle\langle\Psi|$), but on the double density matrix level (as $\mathbb{E}_{|\Psi\rangle \in \mathcal{E}_{\mathrm{RTN}}}(|\Psi\rangle\langle\Psi|)^{\otimes 2}$). The same average strategy commonly appeared in random tensor network/quantum circuit literatures [26,33,34,46–48]. Such average may not have direct physical realization, nevertheless it defines a RTN model for entanglement entropy, which can produce (i) positive mutual information $I_{A:B}$ that does not vanish between distant regions and (ii) possibly negative tripartite information $I_{A:B:C}$ (see Appendix A for a perturbative proof). These features indicate that

the generalized RTN model is expressive enough to describe quantum chaotic states with information scrambling [49,50] and to capture mutual information between distant entanglement regions, which goes beyond holographic CFT states.

B. Ising and dual Ising models

Evaluating the ensemble average in Eq. (4) following the approach developed in Ref. [33], the RTN purity e^{-S_A} can be map to the partition function of an Ising model on the graph G with fluctuating coupling constants

$$e^{-S_A} = \sum_{[\sigma, J]} P[\sigma|J]P[J]\,\delta[\sigma_{\text{bdy}} \Leftrightarrow A], \qquad (5)$$

with $P[\sigma|J]$ given by

$$P[\sigma|J] = \frac{e^{-E[\sigma|J]}}{Z[J]},$$

$$E[\sigma|J] = -\sum_{e \in E}\left(\frac{J_e}{2}\prod_{v \in \partial e}\sigma_v\right),$$

$$Z[J] = \sum_{[\sigma]} e^{-E[\sigma|J]}\delta[\sigma_{\text{bdy}} \Leftrightarrow \emptyset], \qquad (6)$$

and $P[J]$ given by

$$P[J] = \int_{|\phi\rangle}\prod_{e \in E}\delta(J_e - S(|\phi_e\rangle))P(|\phi\rangle). \qquad (7)$$

Here $\sigma_v = \pm 1$ is the Ising variable defined on every vertex $v \in V$, $J_e \geqslant 0$ is the ferromagnetic coupling strength on every edge $e \in E$. J_e is determined by $S(|\phi_e\rangle)$, the 2nd Rényi entropy of the state $|\phi_e\rangle$ (entangled between the Hilbert spaces $\mathcal{H}_{v_{+e}}$ and $\mathcal{H}_{v_{-e}}$ where $v_\pm$ are the two vertices on the boundary of e). J_e characterizes how much the tensors are entangled with each other across the edge e in the tensor network. It corresponds to the notion of bond dimension when $|\phi_e\rangle$ is maximally entangled. The distribution $P[J]$ describes the how the effective bond dimension (bond entanglement) fluctuates in the RTN ensemble. Finally, the partition function is subject to the boundary condition that is set by the boundary region A of S_A,

$$\forall v \in V_{\text{bdy}} : \sigma_v = \begin{cases} +1 & v \notin A, \\ -1 & v \in A, \end{cases} \qquad (8)$$

which is denoted as $\delta[\sigma_{\text{bdy}} \Leftrightarrow A]$ in Eq. (5). The partition function $Z[J]$ properly normalizes the Boltzmann weight of the Ising model, such that $S_A = 0$ when the entanglement region $A = \emptyset$ is empty.

Given that G is a planar graph [51], we can use the Kramers-Wannier duality to rewrite the Ising model Eq. (5) on the dual lattice $\tilde{G} = (\tilde{V}, \tilde{E})$, as shown in Fig. 1(c), where $\tilde{V}$ corresponds to the set of faces in G and $\tilde{E} \cong E$. The dual Ising model takes the similar form

$$e^{-S_A} = \sum_{[\tilde{\sigma}, \tilde{J}]}\left(\prod_{\tilde{v} \in \partial A}\tilde{\sigma}_{\tilde{v}}\right)P[\tilde{\sigma}|\tilde{J}]P[\tilde{J}], \qquad (9)$$
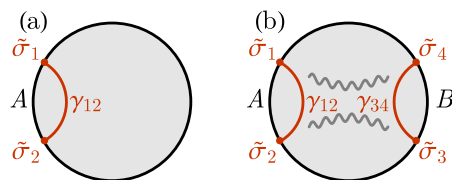

FIG. 2. (a) Two-point correlation and (b) four-point correlation of dual Ising spins.

with $P[\tilde{\sigma}|\tilde{J}]$ given by

$$P[\tilde{\sigma}|\tilde{J}] = \frac{e^{-E[\tilde{\sigma}|\tilde{J}]}}{Z[\tilde{J}]},$$

$$E[\tilde{\sigma}|\tilde{J}] = -\sum_{\tilde{e} \in \tilde{E}}\left(\frac{\tilde{J}_{\tilde{e}}}{2}\prod_{\tilde{v} \in \partial \tilde{e}}\tilde{\sigma}_{\tilde{v}}\right),$$

$$Z[\tilde{J}] = \sum_{[\tilde{\sigma}]} e^{-E[\tilde{\sigma}|\tilde{J}]}, \qquad (10)$$

and $P[\tilde{J}]$ related to $P[J]$ by

$$P[\tilde{J}] = \left(\prod_e \frac{\partial J_e}{\partial \tilde{J}_{\tilde{e}}}\right)P[J]. \qquad (11)$$

Here $\tilde{\sigma}_{\tilde{v}} = \pm 1$ is the dual Ising variable and $\tilde{J}_{\tilde{e}} = -\ln\tanh(J_e/2)$ is the dual coupling. The boundary condition in the original Ising model translates to the insertion of the dual Ising variable at every boundary point of entanglement region A (i.e., at every entanglement cut). The partition function $Z[\tilde{J}]$ on the denominator ensures $S_A = 0$ when the entanglement region $A = \emptyset$ is empty, i.e., when there is no insertion of dual Ising variables. Both $P[\tilde{\sigma}|\tilde{J}]$ and $P[\tilde{J}]$ are normalized probability distributions, which defines the joint distribution $P[\tilde{\sigma}, \tilde{J}] = P[\tilde{\sigma}|\tilde{J}]P[\tilde{J}]$ for dual Ising variables and their couplings. Therefore the purity of the RTN state can be interpreted as the boundary correlation of dual Ising variables $e^{-S_A} = \langle\prod_{\tilde{v} \in \partial A}\tilde{\sigma}_{\tilde{v}}\rangle$ in an Ising model with fluctuating couplings. The RT formula can be recovered in the classical limit when the RTN bond dimensions are large and fixed, which corresponds to the deep ferromagnetic phase of the original Ising model ($J_e \gg 1$) or equivalently the deep paramagnetic phase of the dual Ising model ($\tilde{J}_{\tilde{e}} \ll 1$). In such limit, the dual Ising correlation decays exponentially with the geodesic distance $\langle\tilde{\sigma}_1\tilde{\sigma}_2\rangle \propto e^{-|\gamma_{12}|/\xi}$, as illustrated in Fig. 2(a), which reproduces the RT formula $S_A = |\gamma_{12}|/\xi$ with some appropriate choice of the correlation length ξ. Multi-region entanglement entropies will correspond to higher-point correlations functions, such as $e^{-S_{AB}} \sim \langle\tilde{\sigma}_1\tilde{\sigma}_2\tilde{\sigma}_3\tilde{\sigma}_4\rangle \sim e^{-|\gamma_{12}|/\xi}e^{-|\gamma_{34}|/\xi}$ in Fig. 2(b). Allowing the dual Ising coupling $\tilde{J}$ to fluctuate collectively will introduce perturbations to the geodesic distance $|\gamma_{12}| \to |\gamma_{12}| + \delta|\gamma_{12}|$ in a correlated manner, such that

$$e^{-S_{AB}} \sim \mathbb{E}(e^{-(|\gamma_{12}| + \delta|\gamma_{12}|)/\xi}e^{-(|\gamma_{34}| + \delta|\gamma_{34}|)/\xi})$$

$$\sim e^{-|\gamma_{12}|/\xi}e^{-|\gamma_{34}|/\xi}\,e^{\frac{1}{2\xi^2}\,\mathbb{E}\,\delta|\gamma_{12}|\delta|\gamma_{34}|}$$

$$\sim e^{-S_A}e^{-S_B}e^{I_{A:B}}. \qquad (12)$$

Thus the correlated geometric fluctuation provides an effective mechanism to generate the mutual information between far-separated regions A and B (beyond the classical RT formula). Therefore we anticipate the fluctuating RTN model to be a more expressive holographic model for entanglement entropies. However, it is not clear how the dual Ising coupling $\tilde{J}$ (or the effective bond dimension J) should fluctuate precisely in order to quantitatively reproduce all multi-region entanglement entropies of a given quantum many-body state. The remaining task is learn the distribution $P[\tilde{J}]$ (or other equivalent distributions) from data.

C. Effective statistical gravity Model

Suppose the fluctuation of $\tilde{J}$ is small around its static background configuration, such that there is a meaningful notion of background geometry in the bulk. The dual Ising model can be described by an effective field theory in the continuum limit

$$S[\phi|g] = \frac{1}{2} \int d^2x \sqrt{g}(g^{ij}\partial_i\phi\partial_j\phi + m^2\phi^2), \quad (13)$$

where the dual Ising variable $\tilde{\sigma}_{\tilde{v}}$ is coarse-grained to a massive real scalar field $\phi(x)$, as the Ising model universally flows to this massive Gaussian fixed point in the paramagnetic phase. The theory is defined in the holographic space (without time dimension). The fluctuating Ising coupling $\tilde{J}$ can be translated to a fluctuating bulk metric tensor g around a reference background geometry $\bar{g}$ [52], since a stronger local coupling creates a larger local correlation, which effectively reduces the local distance measure $ds^2 = g_{ij}dx^i dx^j$ between the correlated Ising variables. Therefore the purity of RTN state Eq. (9) can be effectively described by a statistical gravity model

$$e^{-S_A} = \int_{[\phi,g]} \left(\prod_{x\in\partial A} \phi(x)\right) P[\phi|g]P[g],$$

$$P[\phi|g] = \frac{e^{-S[\phi|g]}}{Z[g]}, \quad Z[g] = \int_{[\phi]} e^{-S[\phi|g]}, \quad (14)$$

where the gravity is "quenched" in the sense that the metric configuration is generated with a probability distribution $P[g]$ independent of the scalar field ϕ configuration.

In two-dimensional space, the metric tensor has three independent components. However two of them can be removed by gauge transformation $g_{ij} \to g_{ij} + \nabla_i\xi_j + \nabla_j\xi_i$. We can choose the conformal gauge where the metric tensor $g_{ij}(x)$ is parametrized by a Weyl field $\omega(x)$ that rescales a fixed background $\bar{g}_{ij}(x)$

$$g_{ij}(x) = e^{2\omega(x)}\bar{g}_{ij}(x), \quad (15)$$

such that each Weyl field configuration represents a physically distinct geometry. As a result, the integration $\int_{[g]} P[g]$ can be replaced by $\int_{[\omega]} P[\omega]$ in Eq. (14). The unknown joint distribution $P[\omega]$ will be what we aim to learn from the entanglement entropy data.

To numerically evaluate the multi-point scalar field correlation, we can place the bulk field theory back on a lattice, say on the dual graph $\tilde{G} = (\tilde{V}, \tilde{E})$. Using Regge calculus [53] to

discretize the action,

$$S[\phi|\omega] = \sum_{\langle xy\rangle\in\tilde{E}} \frac{A_{xy}}{2}\left(\frac{\phi_x - \phi_y}{\ell_{xy}}\right)^2 + \sum_{x\in\tilde{V}} \frac{m^2 A_x}{2}e^{2\omega_x}\phi_x^2, \quad (16)$$

where ℓ_{xy} can be interpreted as the geodesic distance between two vertices x and y on the background geometry. A_x and A_{xy} are the areas associated to the vertex x and the edge $\langle xy\rangle$, respectively. ℓ_{xy}, A_x, A_{xy} are all fixed according to the choice of background metric, which will be specified later. The statistical variables in the model are the scalar field ϕ_x and the Weyl field ω_x in the holographic bulk. The model predicts the entanglement entropy on the holographic boundary by

$$e^{-S_A} = \int_{[\phi,\omega]} \left(\prod_{x\in\partial A} \phi_x\right) P[\phi|\omega]P[\omega],$$

$$P[\phi|\omega] = \frac{e^{-S[\phi|\omega]}}{Z[\omega]}, \quad Z[\omega] = \int_{[\phi]} e^{-S[\phi|\omega]} \quad (17)$$

which is the underlying lattice model that will be used in the machine-learning algorithm. The unknown distribution $P[\omega]$ will be parameterized by a generative model. By matching the model prediction with the actual data of entanglement entropies calculated from a quantum state, the algorithm can reconstruct the distribution $P[\omega]$ and infer the statistical gravity model behind the entanglement structure.

III. MACHINE LEARNING METHOD

A. Generative modeling

Generative modeling is about learning probability distributions [54]. We will apply the simplest latent-variable generative model [55] in this paper. The basic idea is to start with a easy-to-sample prior distribution, such as a Gaussian distribution. Draw a random vector $z \in \mathbb{R}^n$ (as a collection of latent variables) from the prior distribution $P(z)$. Then transform the latent variables z by a deep neural network G_ϑ (parametrized by some variational parameter ϑ) to the designated random variable ω, i.e., $z \to \omega = G_\vartheta(z)$. The mapping G_ϑ is called the generator, which defines the distribution of generated samples

$$P_\vartheta[\omega] = \int dz\, \delta(\omega - G_\vartheta(z))P(z). \quad (18)$$

A large batch of ω can be sampled efficiently in parallel, when hardware accelerators (e.g GPU or TPU) are available. If the neural network G_ϑ is expressive enough, Eq. (18) will provide a sufficiently expressive probability model $P_\vartheta[\omega]$ for the Weyl field ω configuration.

The distribution $P_\vartheta[\omega]$ defines the model prediction of the purity based on Eq. (17)

$$e^{-S_{A|\vartheta}} = \mathop{\mathbb{E}}_{[\omega]\sim P_\vartheta} \left\langle\prod_{\partial A} \phi\right\rangle_\omega. \quad (19)$$

We will use $S_{A|\vartheta}$ to denote the Rényi entropy predicted by the machine-learning model as it depends on the model
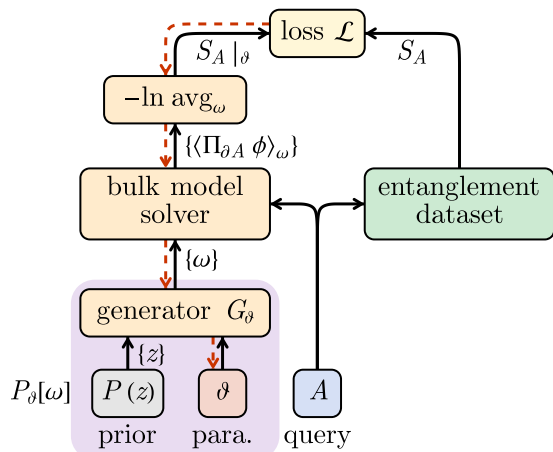
FIG. 3. Flow diagram of the machine-learning algorithm. Black arrows denotes the forward evaluation of the loss function. Red-dashed arrows denotes the gradient back propagation to train the parameter.

parameters ϑ. In Eq. (19), we introduced the short-hand notation

$$\left\langle \prod_{\partial A} \phi \right\rangle_\omega \equiv \int_{[\phi]} \left(\prod_{x \in \partial A} \phi_x \right) P[\phi|\omega] \tag{20}$$

to denote the scalar field correlation on a background Weyl field configuration. The conditional distribution $P[\phi|\omega]$ is defined in Eq. (17) with $S[\phi|\omega]$ given in Eq. (16). The scalar field correlation $\langle \prod_{\partial A} \phi \rangle_\omega$ can be efficiently evaluated when $S[\phi|\omega]$ is a Gaussian action (which is the case here).

The task is to learn the optimal Weyl field distribution $P_\vartheta[\omega]$ that gives the best prediction of the purity data based on Eq. (19). The dataset will contain the purity $\{e^{-S_A}\}$ of a quantum state in different regions A. The distribution $P_\vartheta[\omega]$ can be learned by optimizing model parameters ϑ to minimize the following loss function (to be explained later)

$$\mathcal{L}_\vartheta = \underset{A}{\text{avg}}(1 - e^{S_A|_\vartheta - S_A})^2. \tag{21}$$

As illustrated in Fig. 3, the training initiates from randomly choosing a batch of entanglement regions A. On one hand, we query the dataset to get the ground truth of S_A. On the other hand, a collection of Weyl field configurations are sampled from the generative model, based on which the model prediction $S_A|_\vartheta$ is estimated. Then the loss function is calculated by comparing $S_A|_\vartheta$ with S_A, and the gradient signal propagates back to train the parameters via gradient descent $\vartheta \to \vartheta - r\partial_\vartheta \mathcal{L}_\vartheta$. After some iterations, the parameters are expected to converge. In the following, we will explain different modules in Fig. 3 in detail.

B. Entanglement dataset

While efficient experimental approaches [56,57] have been developed to estimate Rényi entropies from randomized measurements, which enables the acquisition of a large amount of entanglement data to drive the entanglement feature learning, preparing an entanglement dataset by numerically computing entanglement entropies from a given quantum many-body

state remains difficult in general. As a proof of concept, we choose to use the ground state of a free fermion system for demonstration, on which entanglement entropies can be efficiently calculated.

Consider N copies of the $(1 + 1)$D massless Majorana fermion chain, described by the Hamiltonian

$$H = \sum_{a=1}^{N} \sum_j \mathrm{i}\chi_{j,a}\chi_{j+1,a}, \tag{22}$$

where $\{\chi_{i,a}, \chi_{j,b}\} = \delta_{ij}\delta_{ab}$. Let $|\Psi\rangle$ be the ground state of H. The 2nd Rényi entropy can be efficiently computed from the fermion correlation function,

$$S_A = -\frac{1}{2} \operatorname{Tr} \ln \left(C_A^2 + (1 - C_A)^2 \right), \tag{23}$$

where $C_{A,ij} = \langle\Psi|\chi_{i,a}\chi_{j,a}|\Psi\rangle$ (for $i, j \in A$) is the two-point correlation function (matrix) of Majorana fermions restricted inside the entanglement region A. The quantum system is critical and is described by the free-fermion CFT at low energy. To construct the dataset, we will take the Majorana fermion chain of 32 sites, and randomly sample a large collection of single-region, two-region, and three-region subsets. We then compute the entanglement entropy using Eq. (23) for every region and record the results in the entanglement dataset.

C. Bulk Model Solver

The bulk model solver is expected to calculate the scalar field correlation given the Weyl field background ω and the entanglement region A that specifies the scalar field inserting position on the boundary. We will use the lattice model specified by the action in Eq. (16), which describe a free scalar field ϕ. The action can be written as the bilinear form

$$S[\phi|\omega] = \frac{1}{2} \sum_{x,y \in \tilde{V}} \phi_x K_{xy}^{(\phi)}[\omega]\phi_y \tag{24}$$

where x, y label the vertices on the dual graph $\tilde{G} = (\tilde{V}, \tilde{E})$ on which the holographic model is defined. The kernel matrix takes the form of $K^{(\phi)}[\omega] = \nabla^2 + M[\omega]$, with

$$M_{xy}[\omega] = m^2 A_x e^{2\omega_x}\delta_{xy} \tag{25}$$

being the mass term, and

$$\nabla^2_{xy} = \sum_{\langle x'y'\rangle \in \tilde{E}} \frac{A_{x'y'}}{\ell^2_{x'y'}}(\delta_{xx'} - \delta_{xy'})(\delta_{yx'} - \delta_{yy'}) \tag{26}$$

being the discrete Laplace operator on the dual graph. The length ℓ_{xy} and area A_x, A_{xy} constants are fixed and are set by the background geometry, as to be specified soon. The two-point correlation is given by the inverse of the kernel matrix,

$$\langle\phi_x\phi_y\rangle_\omega = \frac{1}{\beta}((K^{(\phi)}[\omega])^{-1})_{xy}, \tag{27}$$

where a trainable constant β is introduced to take care of the field renormalization. Higher-point correlations follow from
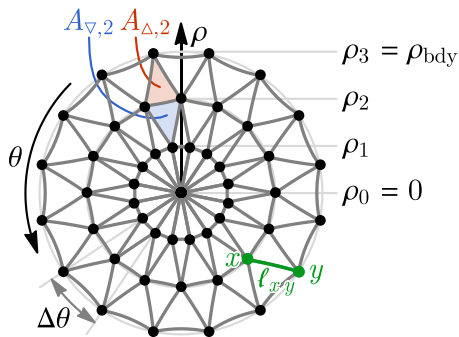
FIG. 4. Triangular lattice discretization of hyperbolic space in (ρ, θ) coordinate. The lattice is divided into different layers along the the radius direction. The ith layer corresponds to the radius ρ_i.

Wick's theorem. For example,

$$
\langle \phi_x \phi_y \phi_z \phi_w \rangle_\omega = \langle \phi_x \phi_y \rangle_\omega \langle \phi_z \phi_w \rangle_\omega + \langle \phi_x \phi_z \rangle_\omega \langle \phi_y \phi_w \rangle_\omega
$$
$$
+ \langle \phi_x \phi_w \rangle_\omega \langle \phi_y \phi_z \rangle_\omega. \tag{28}
$$

We will treat β and m^2 as trainable parameters, which will be optimized (together with other model parameters for $P[\omega]$) to fit the entanglement data.

Since we intend to apply our approach to entanglement data collected from CFTs, following the idea of AdS/CFT correspondence, it is natural to choose the two-dimension hyperbolic geometry (the spatial slice of AdS$_3$) as the background geometry. We use the following background metric

$$
ds^2 = d\rho^2 + \sinh^2 \rho \, d\theta^2, \tag{29}
$$

where $0 \leqslant \theta < 2\pi$ and $\rho \leqslant \rho_{\text{bdy}}$ (the UV cutoff scale is set by ρ_{bdy}, which is another parameter to learn). The geodesic distance any two points on the boundary $\rho = \rho_{\text{bdy}}$ separated by θ is given by

$$
|\gamma|(\theta) = \text{arccosh}(1 + 2 \sinh^2 \rho_{\text{bdy}} \sin^2(\theta/2))
$$
$$
\xrightarrow{e^{\rho_{\text{bdy}}} \gg 1} 2 \ln(\sin(\theta/2) + \rho_{\text{bdy}}).
$$

Without loss of generality, we chose to discretize space using a triangular lattice with periodic boundary condition along the θ direction. All vertices in the same layer are of the same ρ coordinate and their θ coordinates are uniformly spaced, see Fig. 4. The geodesic distance ℓ_{xy} between two vertices x and y is given by

$$
\cosh \ell_{xy} = \cosh \rho_x \cosh \rho_y
$$
$$
- \sinh \rho_x \sinh \rho_y \cos(\theta_x - \theta_y). \tag{30}
$$

The area of an elementary triangle in the ith layer reads

$$
\tan \frac{A_{\triangle,i}}{4} = \tanh \left(\frac{\rho_{i+1} - \rho_i}{2} \right) \tanh \left(\frac{b_i}{4} \right),
$$
$$
\tan \frac{A_{\triangledown,i}}{4} = \tanh \left(\frac{\rho_i - \rho_{i-1}}{2} \right) \tanh \left(\frac{b_i}{4} \right),
$$
$$
\cosh b_i = \cosh^2 \rho_i - \sinh^2 \rho_i \cos \Delta\theta, \tag{31}
$$

which defines the vertex and edge areas in a barycentric scheme. Specifically, the vertex area A_x is given by

$$
A_x = \frac{1}{3}(2A_{\triangle,i} + 2A_{\triangledown,i} + A_{\triangle,i-1} + A_{\triangledown,i+1}), \tag{32}
$$

for $\rho_x = \rho_i$. The edge area A_{xy} is given by

$$
A_{xy} = \begin{cases} \frac{1}{3}(A_{\triangle,i} + A_{\triangledown,i}) & \rho_x = \rho_y = \rho_i; \\ \frac{1}{3}(A_{\triangle,i} + A_{\triangledown,i+1}) & \rho_x = \rho_i, \rho_y = \rho_{i+1}; \\ \frac{1}{3}(A_{\triangle,i-1} + A_{\triangledown,i}) & \rho_x = \rho_i, \rho_y = \rho_{i-1}. \end{cases} \tag{33}
$$

These equations defines ℓ_{xy}, A_x, and A_{xy} used in the lattice model Eq. (16), which all rely on the values of ρ_i for different layers. The discretization scheme in the radial dimension is specified by how ρ_i is spaced from 0 to ρ_{bdy}. A bad choice of the discretization scheme may cause some triangle elements to have high aspect ratios, reducing the quality of the triangulation in approximating the continuous background geometry. We will take a data-driven approach to learn the optimal discretization scheme by treating $\{\rho_i\}$ as trainable parameters.

To summarize, the bulk model solver contains the following parameters: the normalization β and the squared mass m^2 associated with the scalar field dynamics, and the radial coordinates $\{\rho_i\}$ associated with the discretization of background geometry. These parameters will be trained together with other neural network parameters (see Sec. III D) to optimize the model prediction of the entanglement entropy data.

D. Neural network design

The central goal is to learn the Weyl field distribution $P[\omega]$ using a latent-variable generative model $P_\vartheta[\omega]$, recall Eq. (18). The key component of the generative model is a generator G_ϑ that maps the latent variable z to a Weyl field configuration $\omega = G_\vartheta(z)$. The generator is realized as a deep neural network consists of consecutive layers of simpler maps

$$
G_\vartheta(z) = g_N \circ \cdots g_2 \circ g_1(z), \tag{34}
$$

where each layer $g_n(z)$ is an affine transformation followed by some nonlinearity such as ReLU [58]. The weight and bias parameters are introduced to parametrize the affine transformations, which constitute part of the training parameters ϑ.

It is both practical and theoretically motivated to enforce the neural network's architecture such that the learned distribution $P_\vartheta[\omega]$ will respects certain symmetries, i.e., to construct an equivariant neural network [59]. Let Q be a symmetry transformation that we wish to impose. The sufficient condition for the generated distribution to be symmetric (i.e., $P_\vartheta[Q\omega] = P_\vartheta[\omega]$) is to require (i) $Qg_n(z) = g_n(Qz)$ and (ii) $P(Qz) = P(z)$. The symmetries in consideration are

(1) $\omega_{(\rho,\theta)} \to \omega_{(\rho,\theta+a)}$ [translation],

(2) $\omega_{(\rho,\theta)} \to \omega_{(\rho,-\theta)}$ [reflection].

The translation symmetry can be imposed by parameter sharing between relation-related weights and biases, making the affine transformation in each layer effectively a convolution along the translation direction. The reflection symmetry can be imposed by using a reflection symmetric convolution kernel. The prior distribution $P(z) = \prod_x P(z_x)$ automatically satisfies the symmetry condition as it factorizes to identical independent Gaussian distributions on every site.

We would like to emphasize that although each layer looks like a convolutional layer under the symmetry constraint, we do not restrict the convolution kernel to be local (the kernel size extends to the whole lattice), because we do not want to impose locality by hand. As we will see, a sense of locality could emerge in the neural network as the holographic model gets trained, which corresponds to the emergent locality in the bulk gravity theory.

E. Loss function design

The loss function is designed to evaluate the average difference between the purity $e^{-S_{A|\vartheta}}$ predicted by the holographic model and the purity e^{-S_A} given by the entanglement data. A straightforward option would be the mean squared error (MSE) loss

$$\mathcal{L}_{\vartheta,\mathrm{MSE}} = \underset{A}{\mathrm{avg}}(e^{-S_{A|\vartheta}} - e^{-S_A})^2. \qquad (35)$$

The model prediction $e^{-S_{A|\vartheta}}$ should be evaluated according to Eq. (19), which involves the ensemble expectation $\mathbb{E}_{[\omega]\sim P_\vartheta}$. In practice, the expectation value can only be estimated by sampling a finite number of Weyl field configurations from the generative model P_ϑ and take the average

$$e^{-S_{A|\vartheta}} = \frac{1}{N_\omega} \sum_{[\omega]\sim P_\vartheta} \left\langle \prod_{\partial A} \phi \right\rangle_\omega, \qquad (36)$$

where N_ω denotes the number of Weyl field samples. With the help of modern GPU, $\langle \prod_{\partial A} \phi \rangle_\omega$ can be computed in parallel efficiently. The sample size N_ω is thus ultimately limited by the GPU memory. In our case, N_ω ranges from 512 to 2048.

For any finite sample size N_ω, the finite average $e^{-S_{A|\vartheta}}$ will have a finite variance, which bias the MSE loss

$$\mathcal{L}_{\vartheta,\mathrm{MSE}} = \underset{A}{\mathrm{avg}} \left((\mathbb{E}\, e^{-S_{A|\vartheta}} - e^{-S_A})^2 + \frac{\mathrm{var}\, e^{-S_{A|\vartheta}}}{N_\omega} \right), \qquad (37)$$

causing the parameter to converge to a wrong saddle point. The bias can be corrected by assigning a larger weight to the prediction with a higher precision, i.e.,

$$\underset{A}{\mathrm{avg}} \frac{(e^{-S_{A|\vartheta}} - e^{-S_A})^2}{\mathrm{var}\, e^{-S_{A|\vartheta}}}, \qquad (38)$$

which can also be argued from the maximum-likelihood estimation. The variance is generally proportional to the square of the purity $\mathrm{var}\, e^{-S_{A|\vartheta}} \propto (e^{-S_{A|\vartheta}})^2$, which leads to the mean squared relative error (MSRE) loss

$$\begin{aligned} \mathcal{L}_{\vartheta,\mathrm{MSRE}} &= \underset{A}{\mathrm{avg}} \left(\frac{e^{-S_{A|\vartheta}} - e^{-S_A}}{e^{-S_{A|\vartheta}}} \right)^2 \\ &= \underset{A}{\mathrm{avg}}(1 - e^{S_{A|\vartheta} - S_A})^2. \end{aligned} \qquad (39)$$

We numerically test the loss function by generating some data using a model with known parameters, and train new models with different loss function on the generated data to see if the parameter converges to the known result. Our test shows that Eq. (39) indeed converges better compare to Eq. (35). Therefore, we will use the MSRE loss function to train the model, as mentioned in Eq. (21).

TABLE I. MSRE loss on test sets for different models and training sets. The model can be static geometry or fluctuating geometry. The training set can include on single-regions or both single- and double-regions. The test set can be either single-, double-, or triple-regions separately.

	Model	static	static	fluctuating
	Training set	single	single+double	single+double
Test set	single	8.7×10^{-6}	2.1×10^{-2}	1.5×10^{-3}
	double	1.1×10^{-1}	3.9×10^{-2}	5.7×10^{-3}
	triple	7.5×10^{-1}	6.0×10^{-1}	3.1×10^{-1}

IV. NUMERICAL RESULTS

A. Fitting entanglement data with static and fluctuating geometry

We apply the proposed machine-learning approach to learn the entanglement feature of a Majorana fermion chain of 32 sites (16 unit cells) with a relatively large central charge $c = 8$. The entanglement data is partitioned into the training set and the test set that does not overlap. Within the training/test set, the data can be further classified by the number of subregions of the entanglement region, including the single-region, double-region, and triple-region entanglement. To demonstrate the effect of introducing gravitational fluctuations, we will compare two holographic models: (i) the *fluctuating* model, i.e., the model $e^{-S_{A|\vartheta}} = \mathbb{E}_{[\omega]\sim P_\vartheta} \langle \prod_{\partial A} \phi \rangle_\omega$ proposed in Eq. (19) with fluctuating geometries, and (ii) the *static* model, i.e., the model $e^{-S_{A|\vartheta}} = \langle \prod_{\partial A} \phi \rangle_{\omega\equiv0}$ with a fixed static geometry. We train both models using the MSRE loss in Eq. (21). The algorithm is implemented in the TensorFlow [60] framework using the ADAM [61] optimizer. Upon convergence, the MSRE loss is evaluated on the test set to characterize the performance of the model. The result is summarized in Table I.

If we train the static geometry model with single-region data only, the model can easily achieve high accuracy (MSRE $\sim 10^{-5}$) in predicting single-region entanglement, as also shown in Fig. 5(a). But the prediction of multi-region entanglement is rather inaccurate (MSRE $\sim 10^{-1}$), meaning that the static geometry model overfits the single-region data and can not be generalized to multi-region data. If we include the double-region data in the training set, and train the
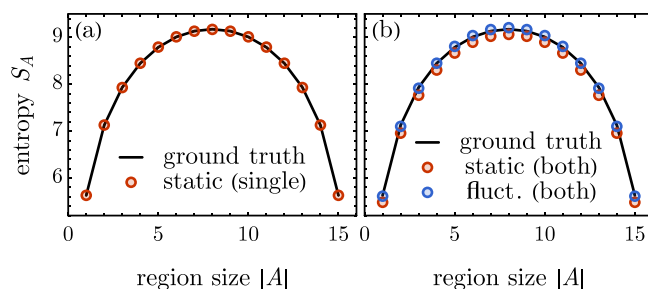

FIG. 5. Model predicted entanglement entropy vs ground truth for (a) static model trained on single-region entanglement data, (b) static and fluctuating models trained on both single-region and double-region entanglement data.
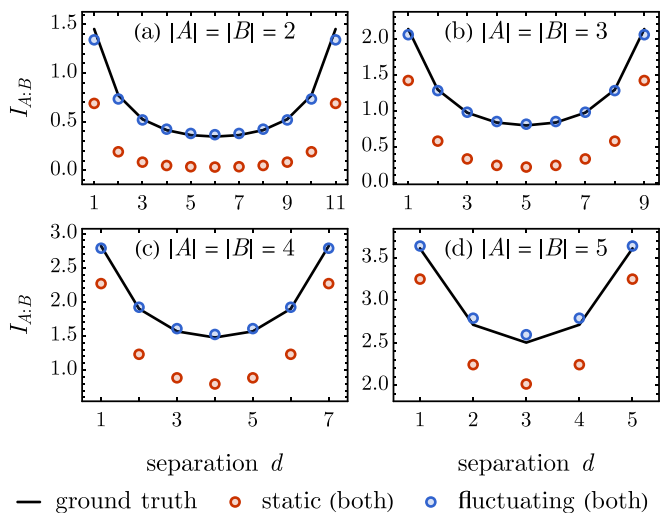
FIG. 6. Mutual information $I_{A:B}$ between two equal-sized regions A and B for different inter-region separations. Data points show predictions by the static model (red) and the fluctuating model (blue). Both models are trained with *both* single- and double-region entanglement data.

FIG. 7. (a) Weyl field local covariance $\Sigma_{xx}^{(\omega)}$ vs the radius coordinate ρ_x. (b) Intralayer correlation $C_{xy}^{(\omega)}$ of the Weyl field (in logarithmic scale) vs the bulk geodesic distance d_{xy}, where x, y points are taken from the same layer. The distance between points on a larger radius (or a higher layer) appears farther due to the hyperbolic background geometry, but the inverse correlation length Δ (the slope) remains roughly on the same order of magnitude across layers.

static geometry model with both single- and double-region entanglement, the model will learn to predict double-region entanglement better at the price of losing the accuracy in predicting single-region entanglement, with the MSRE saturates at the $\sim 10^{-2}$ level. This implies an intrinsic conflict for the static geometry model in modeling the single- and multi-region entanglement simultaneously.

However, by introducing gravitational fluctuations to the model, the fluctuating geometry model achieves one order of magnitude improvement in the prediction accuracy of both single- and double-region entanglement, as the MSRE drops to the $\sim 10^{-3}$ level, which is also manifest in Fig. 5(b). This indicates that the gravitational fluctuation indeed helps to reconcile the conflict between single- and multi-region entanglements in the classical gravity model (RT formula) (see Appendix B for an analytic analysis of how the conflict can be reconciled in principle). Moreover, the prediction accuracy in triple-regions is also improved significantly, even if the model is never trained on the triple-region data. This speaks for the better generalizability of the fluctuating geometry model.

As argued previously, the static geometry model suffers from the problem of vanishing mutual information between far separated regions. One motivation to introduce gravitational fluctuations is to mediate the mutual information between distant regions through the holographic bulk. Indeed, as shown in Fig. 6, by allowing the geometry to fluctuate, the model can better capture the behavior of mutual information. In particular, the static model fails to produce the nonvanishing mutual information between distant regions, which is most obviously seen in Fig. 6(a), where the regions are most far separated (compare to their sizes). However, the fluctuating model fixes this problem, demonstrating the importance of introducing the geometric fluctuation in modeling the multi-region entanglement.
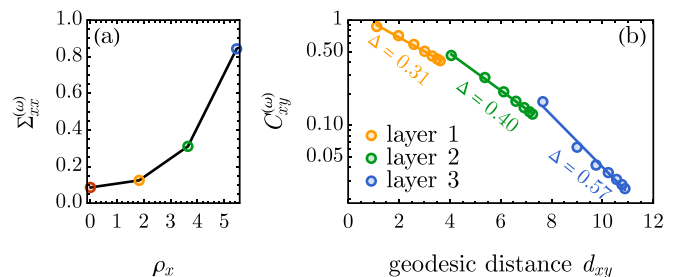
B. Weyl field correlation and effective bulk gravity theory

After training, we want to open up the model and see what bulk gravity theory has been learned. With the trained generative model $P_{\vartheta}[\omega]$ that describes the statistical fluctuation of the Weyl field ω, we can explore various statistical properties of the distribution $P_{\vartheta}[\omega]$ to gain a deeper understanding of the optimal bulk gravity theory that emerges from learning the boundary entanglement data.

We first study the covariance function $\Sigma_{xy}^{(\omega)}$ of the Weyl field ω, defined as

$$\Sigma_{xy}^{(\omega)} = \mathop{\mathbb{E}}_{[\omega] \sim P_{\vartheta}} \omega_x \omega_y = \int_{[\omega]} P_{\vartheta}[\omega] \omega_x \omega_y. \tag{40}$$

We observe that its diagonal elements $\Sigma_{xx}^{(\omega)}$ (i.e., the local covariance) grows with the radius ρ_x coordinate (as x approaches the boundary), see Fig. 7(a). This is because the discretization scale is changing along the radius direction. In our discretization scheme as shown in Fig. 4, the hyperbolic space is finer discretized towards the center of the bulk, therefore the field ω will appear to be stiffer near the bulk center, and hence its covariance is smaller. To eliminate this influence of the discretization scheme, we normalize (standardize) the covariance and define the correlation function

$$C_{xy}^{(\omega)} = \frac{\Sigma_{xy}^{(\omega)}}{\sqrt{\Sigma_{xx}^{(\omega)} \Sigma_{yy}^{(\omega)}}}. \tag{41}$$

We found that the Weyl field correlation $C_{xy}^{(\omega)}$ decays exponentially with respect to the geodesic distance d_{xy} in the holographic bulk

$$C_{xy}^{(\omega)} \sim \exp(-\Delta d_{xy}), \tag{42}$$

where the inverse correlation length Δ remains almost the same across different layers in the bulk, as shown in Fig. 7(b). The short-ranged nature of the Weyl field correlation is more obviously shown in Fig. 8, which is an unequivocal sign of locality. In other words, the machine has learned from the entanglement data that the Weyl field fluctuation can be described by a local model (as the correlation is short-ranged) in the bulk. This emergent locality is remarkable since locality
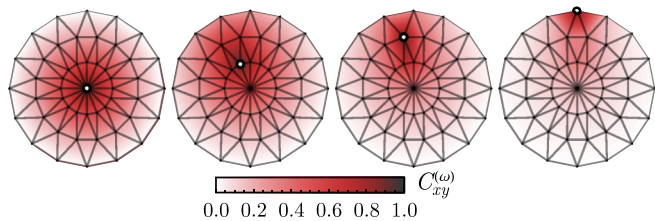
FIG. 8. Weyl field correlation in the bulk, between a marked reference point and the remaining points. Results are shown for the reference point placed in different layers.

was never explicitly given to the generative model at the architecture level: the neural network in the generator G_ϑ was fully connected, which in principle allows nonlocal/long-ranged correlation of ω across the bulk, yet a short-ranged correlation emerges from learning the entanglement data.

Further more, we can learn about the leading modes of gravitational fluctuations in the machine-learned distribution $P_\vartheta[\omega]$ by computing the spectral decomposition of the covariance function

$$\Sigma_{xy}^{(\omega)} = \sum_\alpha \lambda^{(\alpha)} \omega_x^{(\alpha)} \omega_y^{(\alpha)}, \tag{43}$$

where $\lambda^{(\alpha)}$ is the αth eigenvalue and $\omega^{(\alpha)}$ is the corresponding eigenmode. The result is shown in Fig. 9. The long wave-length collective fluctuations emerges as the leading (low-energy) modes of gravitational fluctuation automatically. Using the covariant function $\Sigma_{xy}^{(\omega)}$, one can reconstruct the effective gravitational action to the quadratic order (at Gaussian level)

$$S[\omega] = \frac{1}{2} \sum_{x,y} \omega_x \Sigma_{xy}^{(\omega)} \omega_y + \cdots, \tag{44}$$

such that $P[\omega] \propto e^{-S[\omega]}$ approximately. In this way, the machine-learning model helps us to extract a statistical gravity theory (in terms of the Weyl field theory) from the entanglement data, demonstrating a data-driven approach to establish the holographic duality.

C. Matter field mass renormalization effect

As we have seen, geometric fluctuation effectively introduces interactions between the bulk scalar field ϕ, which generates the desired behavior for mutual information. As a consequence, the bare mass m of the scalar field should also be renormalized by the gravitational interaction. Remarkably, we can observe such a renormalization effect in our holographic model, by comparing the static model (without geometric fluctuation) and the fluctuating model (with geometric fluctuation). The mass parameter m is trainable in both models, but their optimal values are different due to the renormalization effect. We train the static model on the single-region entanglement data, and the fluctuating model on both the single- and double-region entanglement data. For a range of total central charge $11/2 \leqslant c \leqslant 8$ studied, we observe that the trained value of the (bare) mass m in the fluctuating model is systematically larger than that m_0 in the static model, as shown in Fig. 10.
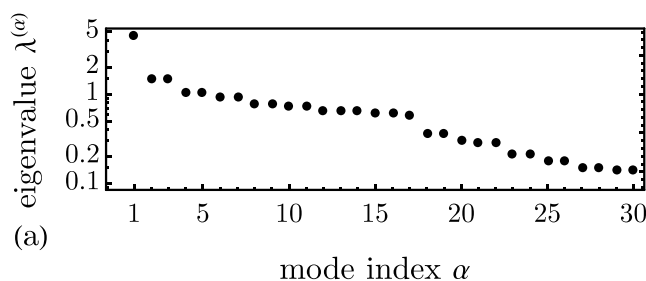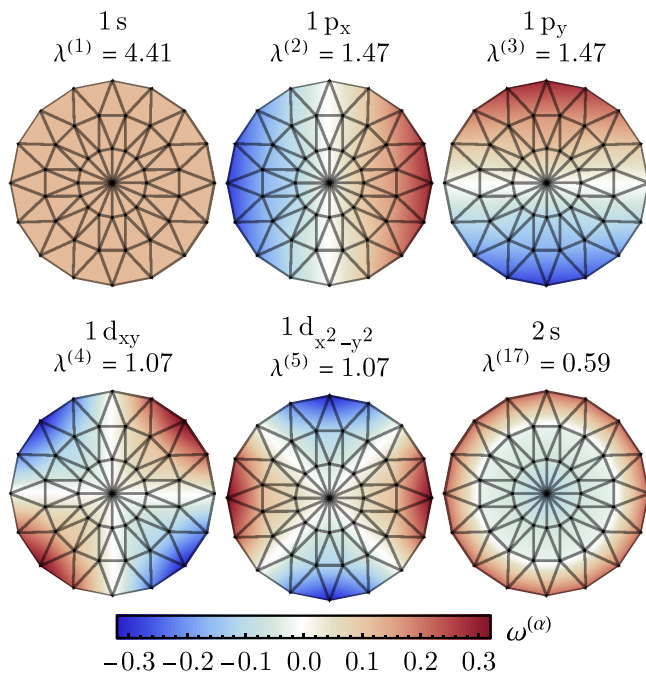

FIG. 9. Spectrum of the Weyl field fluctuation. (a) Leading eigenvalues $\lambda^{(\alpha)}$ of the covariance function (in logarithmic scale). (b) Selected eigenmodes $\omega^{(\alpha)}$ labeled by the principal and angular quantum numbers.

The mass renormalization effect can be understood heuristically by consider a single-region entanglement. In the static model, the entanglement entropy is modeled by $e^{-S_A} \sim$
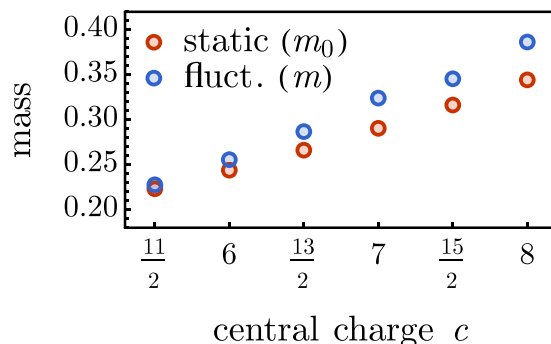

FIG. 10. Trained values of the scalar field mass m for different central charge c of the free fermion CFT, based on the static model (red) and the fluctuating model (blue).

$e^{-m_0|\gamma_A|}$ where γ_A is the geodesic connecting the entanglement cuts of A through the static bulk. With geometric fluctuation $|\gamma_A| \to |\gamma_A| + \delta|\gamma_A|$ (where $\delta|\gamma_A|$ is the additional geodesic length due to the Weyl field ω), the entanglement entropy will be modeled by

$$e^{-S_A} \sim \mathbb{E}_\omega \, e^{-m(|\gamma_A|+\delta|\gamma_A|)} \simeq e^{-m(|\gamma_A|-\frac{m}{2}\mathbb{E}_\omega(\delta|\gamma_A|)^2)}. \quad (45)$$

For these two models to match, we must have $m > m_0$, which qualitatively explains our observation.

V. SUMMARY AND DISCUSSION

We present a machine-learning approach to extract the holography statistical gravity theory from the data of multi-region entanglement entropy in a quantum many-body system. Our paper advances both the field of tensor network holography and the field of machine learning holography. (i) On the tensor network holography side, we generalize the random tensor network (RTN) model to incorporate the bond dimension fluctuation, which makes the model more expressive in capturing features of multi-region entanglement. We derive the holographic bulk theory for the RTN with bond dimension fluctuation and show that the dual gravity theory consists of a massive scalar field on a fluctuating background geometry. The idea of using Ising duality in the derivation is also quite original, which provides an alternative view of the bulk theory that has not been presented in literature, as we are aware of. (ii) On the machine learning holography side, our work goes beyond the previous approaches [26,62–67] of inferring only a static background geometry from the boundary quantum data. By modeling the bulk geometric fluctuation with a generative model, our approach can extract a statistical gravity theory from the quantum entanglement data. Remarkably, we found that the machine-constructed gravity theory exhibit an emergent locality, which reveals the hidden bulk locality behind the nonlocal quantum entanglement on the boundary.

Our paper provides a data-driven approach to explore the emergent gravity from quantum entanglement. Combining with the recent development of efficient numerical methods to simulate entanglement dynamics in quantum many-body systems [47,48,68], we can further explore the corresponding gravity dynamics in the holographic bulk, which will deepen our understanding of emergent gravity from quantum entanglement. On the practical side, our algorithm will boost the efficiency to model the entanglement structure of quantum many-body systems, which will find applications in quantum algorithm optimization and quantum circuit design.

ACKNOWLEDGMENTS

APPENDIX A: PERTURBATIVE ANALYSIS OF MUTUAL AND TRIPARTITE INFORMATION

The random tensor network model points to a bulk theory described by the following action:

$$S[\phi, \omega] = S[\phi|\omega] + S[\omega], \quad (A1)$$

where $S[\phi|\omega] = \frac{1}{2}\sum_{x,y}\phi_x K_{xy}^{(\phi)}[\omega]\phi_y$ follows from Eq. (24), and we take a quadratic action $S[\omega] = \frac{1}{2}\sum_{x,y}\omega_x K_{xy}^{(\omega)}\omega_y$ for simplicity. In the perturbative limit, we assume that the fluctuation of the field ω is small, such that we can expand $K_{xy}^{(\phi)}[\omega] = K_{xy}^{(\phi)} + g\omega_x\delta_{xy}$, where $K_{xy}^{(\phi)}$ denotes the bare kernel of ϕ on the $\omega = 0$ background. Therefore, the bulk theory becomes

$$S[\phi, \omega] = \frac{1}{2}\sum_{x,y}\left(\phi_x K_{xy}^{(\phi)}\phi_y + \omega_x K_{xy}^{(\omega)}\omega_y\right) + \frac{g}{2}\sum_x \omega_x\phi_x^2. \quad (A2)$$

Define the field theory average as

$$\langle\cdots\rangle = \frac{1}{Z}\int_{[\phi,\omega]} \cdots e^{-S[\phi,\omega]}, \quad (A3)$$

then the entanglement entropy S_A of a single-region A that ends at the dual sites (x_1, x_2) can be written as

$$e^{-S_A} = \langle\phi_{x_1}\phi_{x_2}\rangle. \quad (A4)$$

The entanglement entropy for multi-regions are modeled similarly as multi-point covariance of the ϕ field among all boundary points.

Now we consider three regions A, B and C boundaried by (x_1, x_2), (x_3, x_4) and (x_5, x_6), respectively. The mutual information $I_{A:B} = S_A + S_B - S_{AB}$ and the tripartite information $I_{A:B:C} = S_A + S_B + S_C - S_{AB} - S_{BC} - S_{AC} + S_{ABC}$ can be evaluated by the following ratios of covariance function

$$e^{-I_{A:B}} = \frac{\langle\phi_{x_1}\phi_{x_2}\rangle\langle\phi_{x_3}\phi_{x_4}\rangle}{\langle\phi_{x_1}\phi_{x_2}\phi_{x_3}\phi_{x_4}\rangle},$$

$$e^{-I_{A:B:C}} = \frac{\langle\phi_{x_1}\phi_{x_2}\rangle\langle\phi_{x_3}\phi_{x_4}\rangle\langle\phi_{x_5}\phi_{x_6}\rangle\langle\phi_{x_1}\phi_{x_2}\phi_{x_3}\phi_{x_4}\phi_{x_5}\phi_{x_6}\rangle}{\langle\phi_{x_1}\phi_{x_2}\phi_{x_3}\phi_{x_4}\rangle\langle\phi_{x_1}\phi_{x_2}\phi_{x_3}\phi_{x_4}\rangle\langle\phi_{x_1}\phi_{x_2}\phi_{x_3}\phi_{x_4}\rangle}. \quad (A5)$$

To simplify the notation in the following discussion, we introduce a few graphs in Fig. 11. Let $\mathcal{G}_4$ be the complete graph over x_1, x_2, x_3, x_4, and $\mathcal{G}_6$ be the complete graph over x_1, x_2, x_3, x_4, x_5. Further denote $\mathcal{G}'_{2n}$ graph (with a prime) to be the graph with edges $(x_{2k-1}x_{2k})$ (for $k = 1, \cdots, n$) removed from $\mathcal{G}_{2n}$. Define the set of perfect matchings on a graph $\mathcal{G}$ by $\mathcal{M}[\mathcal{G}]$ (where each perfect matching is a subset of edges such that every vertex is covered and only covered by one edge). We define the bare propagators (the covariance functions) $\Sigma^{(\phi)}$ and $\Sigma^{(\omega)}$ from the inverses of the bare kernels $K^{(\phi)}$ and $K^{(\omega)}$ for both ϕ and ω fields, respectively,

$$\Sigma_{xy}^{(\phi)} = ((K^{(\phi)})^{-1})_{xy}, \quad \Sigma_{xy}^{(\omega)} = ((K^{(\omega)})^{-1})_{xy}. \quad (A6)$$

Using perturbative field theory [treating g in Eq. (A1) as perturbation], to the 2nd order in g (and keeping only the tree
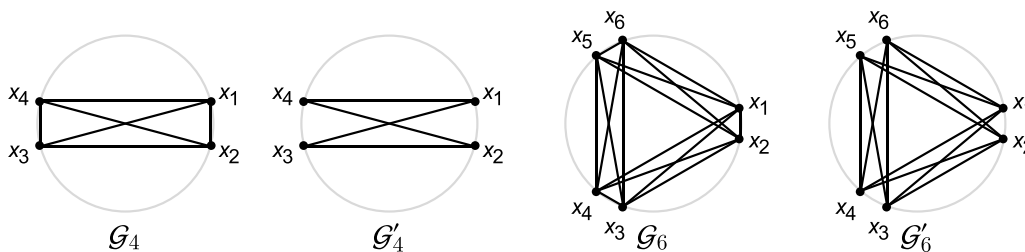
FIG. 11. A few useful graphs.

level diagrams), we can calculate the covariance functions

$$\langle \phi_{x_1} \phi_{x_2} \rangle = \Sigma^{(\phi)}_{x_1 x_2}, \tag{A7}$$

$$\langle \phi_{x_1} \phi_{x_2} \phi_{x_3} \phi_{x_4} \rangle = \sum_{(x_i x_j)(x_k x_l) \in \mathcal{M}[\mathcal{G}_4]} \left(\Sigma^{(\phi)}_{x_i x_j} \Sigma^{(\phi)}_{x_k x_l} + g^2 \sum_{y_1, y_2} \Sigma^{(\omega)}_{y_1 y_2} \Sigma^{(\phi)}_{x_i y_1} \Sigma^{(\phi)}_{y_1 x_j} \Sigma^{(\phi)}_{x_k y_2} \Sigma^{(\phi)}_{y_2 x_l} \right), \tag{A8}$$

$$\langle \phi_{x_1} \phi_{x_2} \phi_{x_3} \phi_{x_4} \phi_{x_5} \phi_{x_6} \rangle = \sum_{(x_i x_j)(x_k x_l)(x_m x_n) \in \mathcal{M}[\mathcal{G}_6]} \left(\Sigma^{(\phi)}_{x_i x_j} \Sigma^{(\phi)}_{x_k x_l} \Sigma^{(\phi)}_{x_m x_n} + g^2 \sum_{y_1, y_2} \Sigma^{(\omega)}_{y_1 y_2} \left(\Sigma^{(\phi)}_{x_i y_1} \Sigma^{(\phi)}_{y_1 x_j} \Sigma^{(\phi)}_{x_k y_2} \Sigma^{(\phi)}_{y_2 x_l} \Sigma^{(\phi)}_{x_m x_n} \right. \right.$$

$$\left. \left. + \Sigma^{(\phi)}_{x_i y_1} \Sigma^{(\phi)}_{y_1 x_j} \Sigma^{(\phi)}_{x_k x_l} \Sigma^{(\phi)}_{x_m y_2} \Sigma^{(\phi)}_{y_2 x_n} + \Sigma^{(\phi)}_{x_i x_j} \Sigma^{(\phi)}_{x_k y_1} \Sigma^{(\phi)}_{y_1 x_l} \Sigma^{(\phi)}_{x_m y_2} \Sigma^{(\phi)}_{y_2 x_n} \right) \right). \tag{A9}$$

Substitute the correlation functions Eq. (A7)–Eq. (A9) to Eq. (A5), we find (to the g^2 order)

$$I_{A:B} = \sum_{(x_i x_j)(x_k x_l) \in \mathcal{M}[\mathcal{G}'_4]} \frac{\Sigma^{(\phi)}_{x_i x_j} \Sigma^{(\phi)}_{x_k x_l}}{\Sigma^{(\phi)}_{x_1 x_2} \Sigma^{(\phi)}_{x_3 x_4}} + g^2 \sum_{(x_i x_j)(x_k x_l) \in \mathcal{M}[\mathcal{G}_4]} \sum_{y_1, y_2} \Sigma^{(\omega)}_{y_1 y_2} \frac{\Sigma^{(\phi)}_{x_i y_1} \Sigma^{(\phi)}_{y_1 x_j} \Sigma^{(\phi)}_{x_k y_2} \Sigma^{(\phi)}_{y_2 x_l}}{\Sigma^{(\phi)}_{x_1 x_2} \Sigma^{(\phi)}_{x_3 x_4}}, \tag{A10}$$

$$I_{A:B:C} = - \sum_{(x_i x_j)(x_k x_l)(x_m x_n) \in \mathcal{M}[\mathcal{G}'_6]} \frac{\Sigma^{(\phi)}_{x_i x_j} \Sigma^{(\phi)}_{x_k x_l} \Sigma^{(\phi)}_{x_m x_n}}{\Sigma^{(\phi)}_{x_1 x_2} \Sigma^{(\phi)}_{x_3 x_4} \Sigma^{(\phi)}_{x_5 x_6}}$$

$$- g^2 \sum_{(x_i x_j)(x_k x_l)(x_m x_n) \in \mathcal{M}[\mathcal{G}'_6]} \sum_{y_1, y_2} \Sigma^{(\omega)}_{y_1 y_2} \left(\frac{\Sigma^{(\phi)}_{x_i y_1} \Sigma^{(\phi)}_{y_1 x_j} \Sigma^{(\phi)}_{x_k y_2} \Sigma^{(\phi)}_{y_2 x_l} \Sigma^{(\phi)}_{x_m x_n}}{\Sigma^{(\phi)}_{x_1 x_2} \Sigma^{(\phi)}_{x_3 x_4} \Sigma^{(\phi)}_{x_5 x_6}} + \frac{\Sigma^{(\phi)}_{x_i y_1} \Sigma^{(\phi)}_{y_1 x_k} \Sigma^{(\phi)}_{x_j y_2} \Sigma^{(\phi)}_{y_2 x_l} \Sigma^{(\phi)}_{x_m x_n}}{2\Sigma^{(\phi)}_{x_1 x_2} \Sigma^{(\phi)}_{x_3 x_4} \Sigma^{(\phi)}_{x_5 x_6}} + \frac{\Sigma^{(\phi)}_{x_i y_1} \Sigma^{(\phi)}_{y_1 x_j} \Sigma^{(\phi)}_{x_k x_l} \Sigma^{(\phi)}_{x_m y_2} \Sigma^{(\phi)}_{y_2 x_n}}{\Sigma^{(\phi)}_{x_1 x_2} \Sigma^{(\phi)}_{x_3 x_4} \Sigma^{(\phi)}_{x_5 x_6}} \right.$$

$$\left. + \frac{\Sigma^{(\phi)}_{x_i y_1} \Sigma^{(\phi)}_{y_1 x_m} \Sigma^{(\phi)}_{x_k x_l} \Sigma^{(\phi)}_{x_j y_2} \Sigma^{(\phi)}_{y_2 x_n}}{2\Sigma^{(\phi)}_{x_1 x_2} \Sigma^{(\phi)}_{x_3 x_4} \Sigma^{(\phi)}_{x_5 x_6}} + \frac{\Sigma^{(\phi)}_{x_i x_j} \Sigma^{(\phi)}_{x_k y_1} \Sigma^{(\phi)}_{y_1 x_l} \Sigma^{(\phi)}_{x_m y_2} \Sigma^{(\phi)}_{y_2 x_n}}{\Sigma^{(\phi)}_{x_1 x_2} \Sigma^{(\phi)}_{x_3 x_4} \Sigma^{(\phi)}_{x_5 x_6}} + \frac{\Sigma^{(\phi)}_{x_i x_j} \Sigma^{(\phi)}_{x_k y_1} \Sigma^{(\phi)}_{y_1 x_m} \Sigma^{(\phi)}_{x_l y_2} \Sigma^{(\phi)}_{y_2 x_n}}{2\Sigma^{(\phi)}_{x_1 x_2} \Sigma^{(\phi)}_{x_3 x_4} \Sigma^{(\phi)}_{x_5 x_6}} \right). \tag{A11}$$

In the case that $\Sigma^{(\phi)}_{xy} \geqslant 0$ and $\Sigma^{(\omega)}_{xy} \geqslant 0$ (which is typically the case), we can ensure $I_{A:B} \geqslant 0$ and $I_{A:B:C} \leqslant 0$. The result proves that random tensor network model can produce a negative tripartite information $I_{A:B:C}$, which is a unique feature of quantum many-body entanglement that can not be achieved in classical systems. A negative tripartite information is an indication of quantum information scrambling and chaotic quantum dynamics in the quantum system. Although the bulk theory is a classical statistical gravity model, it can still model the quantum chaotic entanglement features on the holographic boundary, which speak for the strong expression power of the random tensor network model.

APPENDIX B: NECESSITY AND EXPECTED BEHAVIOR OF WEYL FIELD FLUCTUATION

We would like to take a closer look at the mutual information. It would more intuitive to present the diagrammatic representation of Eq. (A10)

$$\tag{B1}$$

where points on the boundary correspond to x_1, x_2, x_3, x_4 (following the arrangement of vertices in the $\mathcal{G}_4$ graph shown in Fig. 11) and the small circles in the bulk correspond to y_1, y_2 that should be summed over. The black lines represent $\Sigma^{(\phi)}_{xy}$ and the gray lines represent $\Sigma^{(\omega)}_{xy}$. The perturbation g parameterizes the coupling strength of the bulk scalar field ϕ to the background gravitational fluctuation (the Weyl field ω). Setting $g = 0$ will decouple the gravitational fluctuation,

which effectively corresponds to a static bulk model (because gravitational fluctuation will have no effect in the decoupled limit). Let us consider the case when regions A and B are far separated, meaning that the spacings $|x_1 - x_2|$ and $|x_3 - x_4|$ are small. In this case, we expect the mutual information to decay with the inter-region spacing in a power-law manner with the power set by the smallest scaling dimension of the critical field in the quantum system on the holographic boundary, because the mutual information upper bounds all correlation function between regions A and B, which can not decay faster than the lightest critical field. As we will see, this behavior can only be reproduce via the bulk model gravitational fluctuation is included.

To argue the necessity of including gravitational fluctuation, we first consider the decoupled limit (i.e., $g = 0$) to demonstrate why it fails to capture the correct behavior of mutual information. In the $g = 0$ limit, the first two terms can still contribute to a finite mutual information that decays with the inter-region separation in a power-law manner, but the power will be set by the central charge of the quantum system on the holographic boundary. Because the power-law comes from the ϕ-field correlation $\Sigma_{xy}^{(\phi)}$, whose scaling is determined by the single-region entanglement entropy, as $-\ln \Sigma_{x_1 x_2}^{(\phi)} \sim S_A \sim c \ln |A|$ (such that x_1, x_2 are boundary points of the region A). However, the total central charge $c = N/2$ can be as large as we wish in the large N limit. Therefore, although the first two terms (the g^0 terms) in Eq. (B1) can produce a power-law decay mutual information, but the power will typically be too large (i.e., the mutual information will decay too fast). This

reflects the internal inconsistency in describing both single- and double-region entanglements using a holographic bulk model without gravitational fluctuation.

An obvious solution is to introduce a different field from ϕ to mediate the mutual information across the holographic bulk. Then we will have an independent freedom, such that we can tune its scaling dimension to match that of the lightest critical field. This is one major motivation to introduce the gravitational fluctuation (or to couple the scaler field ϕ to a fluctuating Weyl field ω). As we turn on the coupling g, the mutual information will be dominated by

$$I_{A:B} = g^2 \sum_{y_1, y_2} \Sigma_{y_1 y_2}^{(\omega)} \frac{\Sigma_{x_1 y_1}^{(\phi)} \Sigma_{y_1 x_2}^{(\phi)} \Sigma_{x_3 y_2}^{(\phi)} \Sigma_{y_2 x_4}^{(\phi)}}{\Sigma_{x_1 x_2}^{(\phi)} \Sigma_{x_3 x_4}^{(\phi)}}, \qquad (B2)$$

whose long-range behavior scales with $\sim g^2 \Sigma_{y_1 y_2}^{(\omega)}$. The Weyl field ω has a different propagator $\Sigma_{xy}^{(\omega)}$, which can be independently tuned to make $-\ln \Sigma_{xy}^{(\omega)} \sim 2\Delta_{\min} \ln |x - y|$ with $\Delta_{\min}$ being the smallest scaling dimension in the quantum critical theory. Here $|x - y|$ denotes the distance between two boundary points x and y measured using the bound metric. Translate $|x - y|$ into the bulk distance d_{xy} assuming the bulk has a hyperbolic background geometry, we have $d_{xy} \sim \ln |x - y|$, which implies $\Sigma_{xy}^{(\omega)} \sim e^{-2\Delta_{\min} d_{xy}}$. This indicates that Weyl field must be heavy in the bulk to produce the exponential decay of its correlation function with the bulk distance. Indeed, such a massive Weyl field fluctuation does emerge in the machine-learnt bulk gravity theory.

[1] J. D. Brown and M. Henneaux, Central charges in the canonical realization of asymmetric symmetries: An example from three-dimensional gravity, Commun. Math. Phys. **104**, 207 (1986).

[2] E. Witten, Anti-de Sitter space and holography, Adv. Theor. Math. Phys. **2**, 253 (1998).

[3] E. Witten, Anti-de Sitter space, thermal phase transition, and confinement in gauge theories, Adv. Theor. Math. Phys. **2**, 505 (1998).

[4] S. S. Gubser, I. R. Klebanov, and A. M. Polyakov, Gauge theory correlators from non-critical string theory, Phys. Lett. B **428**, 105 (1998).

[5] J. Maldacena, The large-N Limit of superconformal field theories and supergravity, Int. J. Theor. Phys. **38**, 1113 (1999).

[6] M. van Raamsdonk, Comments on quantum gravity and entanglement, arXiv:0907.2939.

[7] M. van Raamsdonk, Building up spacetime with quantum entanglement, Gen. Relativ. Gravit. **42**, 2323 (2010).

[8] J. Maldacena and L. Susskind, Cool horizons for entangled black holes, Fortschr. Phys. **61**, 781 (2013).

[9] K. Jensen and A. Karch, Holographic Dual of an Einstein-Podolsky-Rosen Pair Has a Wormhole, Phys. Rev. Lett. **111**, 211602 (2013).

[10] V. Balasubramanian, B. D. Chowdhury, B. Czech, and J. de Boer, The entropy of a hole in spacetime, J. High Energy Phys. 10 (2013) 220.

[11] X.-L. Qi, Exact holographic mapping and emergent space-time geometry, arXiv:1309.6282.

[12] V. Balasubramanian, P. Hayden, A. Maloney, D. Marolf, and S. F. Ross, Multiboundary wormholes and holographic entanglement, Class. Quantum Gravity **31**, 185015 (2014).

[13] L. Susskind, ER=EPR, GHZ, and the consistency of quantum measurements, arXiv:1412.8483 .

[14] V. Balasubramanian, B. D. Chowdhury, B. Czech, J. de Boer, and M. P. Heller, Bulk curves from boundary data in holography, Phys. Rev. D **89**, 086004 (2014).

[15] B. Czech and L. Lamprou, Holographic definition of points and distances, Phys. Rev. D **90**, 106005 (2014).

[16] C. J. Cao, S. M. Carroll, and S. Michalakis, Space from Hilbert space: Recovering geometry from bulk entanglement, Phys. Rev. D **95**, 024031 (2017).

[17] S. Ryu and T. Takayanagi, Holographic Derivation of Entanglement Entropy from the anti de Sitter Space/Conformal Field Theory Correspondence, Phys. Rev. Lett. **96**, 181602 (2006).

[18] S. Ryu and T. Takayanagi, Aspects of holographic entanglement entropy, J. High Energy Phys. 0608 (2006) 045.

[19] M. Porrati and R. Rabadan, Boundary rigidity and holography, J. High Energy Phys. 0401 (2004) 034.

[20] J. Hammersley, Extracting the bulk metric from boundary information in asymptotically AdS spacetimes, J. High Energy Phys. 12 (2006) 047.

[21] S. Bilson, Extracting spacetimes using the AdS/CFT conjecture, J. High Energy Phys. 08 (2008) 073.

[22] C. Cao, X.-L. Qi, B. Swingle, and E. Tang, Building bulk geometry from the tensor Radon transform, J. High Energy Phys. 2020 (2020) 33.

[23] S. Bilson, Extracting spacetimes using the AdS/CFT conjecture. Part II, J. High Energy Phys. 02 (2011) 050.

[24] S. Alexakis, T. Balehowsky, and A. Nachman, Determining a Riemannian metric from minimal areas, Adv. Math. **366**, 107025 (2017).

[25] N. Bao, C. Cao, S. Fischetti, and C. Keeler, Towards bulk metric reconstruction from extremal area variations, Class. Quantum Grav. **36**, 185002 (2019).

[26] Y.-Z. You, Z. Yang, and X.-L. Qi, Machine learning spatial geometry from entanglement features, Phys. Rev. B **97**, 045153 (2018).

[27] S. R. Roy and D. Sarkar, Bulk metric reconstruction from boundary entanglement, Phys. Rev. D **98**, 066017 (2018).

[28] M. Headrick, Entanglement Rényi entropies in holographic theories, Phys. Rev. D **82**, 126010 (2010).

[29] M. Headrick, Lectures on entanglement entropy in field theory and holography, arXiv:1907.08126.

[30] T. Faulkner, A. Lewkowycz, and J. Maldacena, Quantum corrections to holographic entanglement entropy, J. High Energy Phys. 11 (2013) 074.

[31] N. Engelhardt and A. C. Wall, Quantum extremal surfaces: holographic entanglement entropy beyond the classical regime, J. High Energy Phys. 01 (2015) 073.

[32] X. Dong, X.-L. Qi, Z. Shangnan, and Z. Yang, Effective entropy of quantum fields coupled with gravity, J. High Energy Phys. 10 (2020) 052.

[33] 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.

[34] X.-L. Qi, Z. Yang, and Y.-Z. You, Holographic coherent states from random tensor networks, J. High Energy Phys. 08 (2017) 060.

[35] B. Swingle, Constructing holographic spacetimes using entanglement renormalization, arXiv:1209.3304.

[36] B. Swingle, Entanglement renormalization and holography, Phys. Rev. D **86**, 065007 (2012).

[37] F. Pastawski, B. Yoshida, D. Harlow, and J. Preskill, Holographic quantum error-correcting codes: Toy models for the bulk/boundary correspondence, J. High Energy Phys. 06 (2015) 149.

[38] R. Salakhutdinov, Learning deep generative models, Annu. Rev. Stat. Appl. **2**, 361 (2015).

[39] D. Jimenez Rezende and S. Mohamed, Variational inference with normalizing flows, arXiv:1505.05770.

[40] A. van den Oord, N. Kalchbrenner, and K. Kavukcuoglu, Pixel recurrent neural networks, arXiv:1601.06759.

[41] L. Dinh, J. Sohl-Dickstein, and S. Bengio, Density estimation using real NVP, arXiv:1605.08803.

[42] D. P. Kingma, T. Salimans, R. Jozefowicz, X. Chen, I. Sutskever, and M. Welling, Improving variational inference with inverse autoregressive flow, arXiv:1606.04934.

[43] A. van den Oord, S. Dieleman, H. Zen, K. Simonyan, O. Vinyals, A. Graves, N. Kalchbrenner, A. Senior, and K. Kavukcuoglu, WaveNet: A generative model for raw audio, arXiv:1609.03499.

[44] G. Papamakarios, T. Pavlakou, and I. Murray, Masked autoregressive flow for density estimation, arXiv:1705.07057.

[45] R. Vasseur, A. C. Potter, Y.-Z. You, and A. W. W. Ludwig, Entanglement transitions from holographic random tensor networks, Phys. Rev. B **100**, 134203 (2019).

[46] N. Bao, G. Penington, J. Sorce, and A. C. Wall, Holographic tensor networks in full AdS/CFT, arXiv:1902.10157.

[47] W.-T. Kuo, A. A. Akhtar, D. P. Arovas, and Y.-Z. You, Markovian entanglement dynamics under locally scrambled quantum evolution, Phys. Rev. B **101**, 224202 (2020).

[48] R. Fan, S. Vijay, A. Vishwanath, and Y.-Z. You, Self-organized error correction in random unitary circuits with measurement, Phys. Rev. B **103**, 174309 (2021).

[49] P. Hosur, X.-L. Qi, D. A. Roberts, and B. Yoshida, Chaos in quantum channels, J. High Energy Phys. 02 (2016) 004.

[50] A. Seshadri, V. Madhok, and A. Lakshminarayan, Tripartite mutual information, entanglement, and scrambling in permutation symmetric systems with an application to quantum chaos, Phys. Rev. E **98**, 052205 (2018).

[51] The model can be more expressive if the planar graph assumption is lifted, however it will be challenging to make connection to the RTN model on nonplanar graphs.

[52] Another interpretation is to translate the fluctuating Ising coupling to the fluctuating mass term, as the scalar field mass is the relevant perturbation that drives the order-disorder transition, which plays the same role as the Ising coupling. This alternative view turns out to be equivalent to the fluctuating metric interpretation in two-dimension, as we will see soon.

[53] T. Regge, General relativity without coordinates, Nuovo Cim. **19**, 558 (1961).

[54] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning* (MIT Press, Cambridge, 2016).

[55] I. J. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, Generative adversarial networks, arXiv:1406.2661.

[56] T. Brydges, A. Elben, P. Jurcevic, B. Vermersch, C. Maier, B. P. Lanyon, P. Zoller, R. Blatt, and C. F. Roos, Probing Rényi entanglement entropy via randomized measurements, Science **364**, 260 (2019).

[57] H.-Y. Huang, R. Kueng, and J. Preskill, Predicting many properties of a quantum system from very few measurements, Nat. Phys. **16**, 1050 (2020).

[58] A. F. Agarap, Deep learning using rectified linear units (ReLU), arXiv:1803.08375.

[59] T. Cohen and M. Welling, Group equivariant convolutional networks, in *Proceedings of The 33rd International Conference on Machine Learning*, Proceedings of Machine Learning Research, Vol. 48, edited by M. F. Balcan and K. Q. Weinberger (PMLR, New York, 2016), pp. 2990–2999.

[60] M. Abadi, A. Agarwal, P. Barham, E. Brevdo, Z. Chen, C. Citro, G. S. Corrado, A. Davis, J. Dean, M. Devin *et al.*, TensorFlow: Large-scale machine learning on heterogeneous systems (2015), software available from tensorflow.org.

[61] D. P. Kingma and J. Ba, Adam: A method for stochastic optimization, arXiv:1412.6980.

[62] W.-C. Gan and F.-W. Shu, Holography as deep learning, Int. J. Mod. Phys. D **26**, 1743020 (2017).

[63] X. Dong and L. Zhou, Spacetime as the optimal generative network of quantum states: A roadmap to QM=GR? arXiv:1804.07908.

[64] K. Hashimoto, S. Sugishita, A. Tanaka, and A. Tomiya, Deep learning and the ads/cft correspondence, Phys. Rev. D **98**, 046019 (2018).

[65] K. Hashimoto, S. Sugishita, A. Tanaka, and A. Tomiya, Deep learning and holographic QCD, Phys. Rev. D **98**, 106014 (2018).

[66] K. Hashimoto, AdS/CFT as a deep Boltzmann machine, Phys. Rev. D **99**, 106017 (2019).

[67] K. Hashimoto, H.-Y. Hu, and Y.-Z. You, Neural ODE and molographic QCD, arXiv:2006.00712.

[68] A. A. Akhtar and Y.-Z. You, Multiregion entanglement in locally scrambled quantum dynamics, Phys. Rev. B **102**, 134203 (2020).