Title
Machine Learning Statistical Gravity from Multi-region Entanglement Entropy

Permalink
https://escholarship.org/uc/item/8h15k5mr

Author
Lam, Chun Cheong

Publication Date
2021

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA SAN DIEGO

Machine Learning Statistical Gravity from Multi-region Entanglement Entropy

A dissertation submitted in partial satisfaction of the
requirements for the degree
Doctor of Philosophy

in

Physics

by

Chun Cheong Lam

Committee in charge:

      Professor Yi-Zhuang You, Chair
      Professor Daniel Arovas
      Professor Michael Galperin
      Professor John McGreevy
      Professor Tatyanya Sharpee

2021

The dissertation of Chun Cheong Lam is approved, and it is acceptable in quality and form for publication on microfilm and electronically.

University of California San Diego

2021

DEDICATION

To those who fight for democracy in Hong Kong.

EPIGRAPH

What I cannot create, I do not understand.

—Richard Feynman

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGEMENTS

Throughout my journey in graduate school, I have received a lot of patient support and assistance.

I would first like to thank my advisor, Professor Yi-Zhuang You, whose expertise was invaluable in formulating the research questions and methodology. You have demonstrated the meaning of being a teacher with your passion in pedagogy, care to students and deep understanding to the subject matter.

I would like to acknowledge my friends and colleagues in UCSD. I have learned so much from the stimulating discussions and enjoyable coffee times with you all.

I would love to thank my partner Joyce Ip, who has supported and loved me wholeheartedly for more than a decade. In addition, I would like to thank my parents for nurturing my curiosity. Without that, life would be a boring one.

Material in Chapters 2-4 is currently being prepared for submission for publication of the material. Lam, Jonathan C. C.; Yi-Zhuang, You. The dissertation author was the primary investigator and author of this material.

VITA

2015	BSc. in Physics, The Chinese University of Hong Kong
2015-2021	Graduate Teaching Assistant, University of California San Diego
2021	Ph. D. in Physics, University of California San Diego

ABSTRACT OF THE DISSERTATION

Machine Learning Statistical Gravity from Multi-region Entanglement Entropy

by

Chun Cheong Lam

Doctor of Philosophy in Physics

University of California San Diego 2021

Professor Yi-Zhuang You, Chair

The holographic duality 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, in particular via the Ryu-Takayanagi (RT) formula which assumes the gravity theory is in the classical limit. Yet the assumption of classical geometry has lead to exponentially small mutual information between disjoint sub-regions, which is not true in many system such as free fermion. In this work, we study a generalized Random Tensor Network (RTN) model with fluctuating bond dimensions, which is mapped to a statistical gravity model consisting a massive scalar field on a fluctuating background geometry. A concrete algorithm is constructed

to recover the underlying geometry fluctuation from multi-region entanglement entropy data by modelling its distribution as a generative neural network. To demonstrate its effectiveness, we train the model using entanglement entropy of a free fermion system and showed mutual information can be mediated effectively by geometric fluctuation. Remarkably, locality emerges from the learned geometric distribution.

Chapter 1

Introduction

1.1 The Challenge Of Quantum Gravity

In 1687, Newton published Principia, in which he postulated his law of universal gravitation

$$F = G\frac{m_1 m_2}{r^2} \tag{1.1}$$

where F is the attractive force between the two objects of masses m_1 and m_2, separated by distance r, and G_N is the gravitational constant. Until a century ago, Newton's theory was superseded by Einstein's general relativity. As one of the most fundamental postulates in general relativity, the equivalence principle allowed Einstein to grant a completely new meaning to gravity. Instead of viewing gravity as a force, the effects of gravitation are ascribed to the geometry of spacetime. Around the same period of time, quantum mechanics was born. Buoyed by the success of the quantum theory, quantization of gravity became the next natural endeavor in theoretical physics. A century later, it remains difficult to describes gravity in a way that is consistent with quantum mechanics. There are many reasons to that. From a practical point of view, classical Einstein's gravity is such an excellent description of the universe across many scales. It is a blessing as well as a curse because quantum correction is expected to appear only at length scale

near the Planck scale, around 30^{-35}m which is many orders of magnitude beyond the reach of current accelerator technology. Besides practical considerations, the quantization of gravity also poses a unique technical challenge theoretically. In the quantum world, physical quantities are constantly fluctuating. In the path integral formulation of quantum mechanics one integrates over $e^{iS_{\text{matter}}[\phi|g]/\hbar}$ over all possible paths of the matter field ϕ living on a spacetime defined by the metric tensor g, where $S_{\text{matter}}[\phi|g]$ is the classical action. Therefore in a theory of quantum gravity, spacetime itself is also fluctuating . One has to evaluate the functional integral

$$\int DgD\phi e^{iS_{\text{gravity}}[\phi,g]/\hbar} \tag{1.2}$$

where $S_{\text{gravity}}[\phi,g] = S_{\text{matter}}[\phi|g] + S_{\text{geometry}}[g]$. The Einstein-Hilbert action is an excellent effective action for $S_{\text{geometry}}[g]$ at the low energy scale where human can experience. It gives rise to general relativity when evaluated at saddle point. However, it is known to be not renormalizable — meaning it will eventually break down at some very small length-scale.

1.2 Holographic Principle

Bekenstein[Bek20] and Hawking[Haw75] showed that, in order to comply with the laws of thermal dynamics in a gravitational system with black holes, a black hole should have an entropy

$$S_{\text{BH}} = \frac{A}{4G} \tag{1.3}$$

where A is the area of its event horizon and G is the Newton's gravitational constant. This 'area law' behavior has some far-reaching consequences. On one hand, the black hole entropy is an upper bound to that of all possible states of matter in a region of the same size. To see that, consider a region of space of volume V bounded by area A. Suppose the matter in this region has entropy $S > S_{BH}$, then by keep putting even more matter in, a black hole would form eventually,

with entropy S_{BH} less than what we started. This violation of the second law of thermodynamics implies $S < S_{BH}$ for all possible state of matter and a black hole to is most entropic configuration possible. On the other hand, as a measure of the number of independent degrees of freedom, the entropy of a physical system is normally extensive and proportional to its volume. This is known to be true for any theory with local interaction such as field theory. The sharp contradiction between the 'area law' of black hole and 'volume law' of general physical system suggests that a quantum theory of gravity, which describes a black hole, should have a number of degrees of freedom which scales like a system with local interaction in one lower dimension. This observation has ultimately lead to the holographic principle[BH86, Wit98b, Wit98a, GKP98, Mal99]. The notion of holography establishes an important duality between two seemingly completely different and unrelated system. On the one hand, one has a quantum mechanical system on a flat spacetime, which a priori has nothing to do with gravity. And on the other hand one has a theory of quantum gravity, one in which the geometry itself fluctuates quantum mechanically.

An important breakthrough was the discovery of AdS (anti-de Sitter)/CFT (conformal field theory) correspondence which is a precise implementation of the holographic principle. As a duality–physical equivalence between two theory, it relates the computation of the partition function of a CFT, a particular kind of quantum field theory (QFT) suitable for describing critical systems, to a quantum theory of gravity which has an asymptotic hyperbolic geometry. More precisely, the mathematical backbone is given by a relation known as the 'Gubser-Klebanov-Polyakov-Witten' (GKPW) rule. The relation identifies the generating functional of a CFT for an observable O, with the partition function of a bulk gravitation theory where the asymptotically anti-de Sitter boundary value of a field ϕ equated to the source h

$$\left\langle e^{i \int dx h(x) O(x)} \right\rangle_{\text{CFT}} = \int_{\phi \to h} Dg_{\mu\nu} D\phi \, e^{iS_{\text{gravity}}[\phi, g_{\mu\nu}]}. \tag{1.4}$$

Therefore, as a useful heuristic mnemonic, one can say that a CFT is living on the boundary of the

AdS spacetime bulk. As an example of weak-strong duality, it is a powerful computational device. For example, if a strongly interacting CFT has classical gravitational dual, one can compute quantities in such a CFT non-perturbatively, by evaluating the r.h.s of eq (1.4) at the saddle point of which the geometry is essential static. Perhaps even more interestingly, this duality allows one to use the l.h.s. of eq (1.4), of systems that are relatively well understood, to *define* the path integral of the r.h.s. In other words, a standard quantum field theory on flat spacetime could tell us about the nature of geometric fluctuation in a quantum theory of gravity in one higher dimension, via the holographic principle. At last but not least, it is a progress toward a UV-complete quantum gravity theory.

1.3   The Ryu-Takayanagi Formula

The holographic duality has provided an appealing explanation for the emergence of spacetime geometry from quantum entanglement[Van09, van10, MS13, JK13, BCCd13, Qi13, BHM$^+$14, Sus14, BCC$^+$14, CL14, CCM17]. Recently, using Ads/CFT correspondence, Ryu and Takayanagi (RT) gave a holographic interpretation[RT06b, RT06a] of the entanglement entropy in a quantum (conformal) field theory. The holographic entanglement entropy formula proposed by them provided a direct and quantitative relation between the quantum entanglement and spacetime geometry in the classical limit where fluctuation of geometry in suppressed.

Consider a equal-time slice Σ of an AdS spacetime on whose boundary $\partial\Sigma$ we define the dual CFT state $|\psi\rangle$. The RT formula states that the entanglement entropy $S_A = -\operatorname{Tr}\rho_A \ln\rho_A$ of a subregion A defined on the boundary CFT, where $\rho_A = \operatorname{Tr}_{\bar{A}}|\psi\rangle\langle\psi|$ can be also computed by

$$S_A = \frac{1}{4G_N}\min_{\gamma_A}|\gamma_A| \tag{1.5}$$

where $|\gamma_A|$ is the area of an extremal surface through the holographic bulk that encloses the boundary region A as illustrated in Figure 1.1. This has the same from as eq (1.3), the one
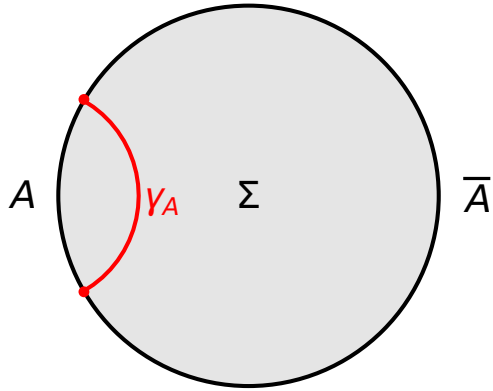
Figure 1.1: Holographic duality and Ryu-Takayanagi formula. Boundary entanglement region A and the corresponding extremal surface $|\gamma_A|$ through holographic bulk.

given by Bekenstein and Hawking, but with the black hole horizon area replaced by an minimal surface area $|\gamma_A|$. This formula has been extended to Rényi entanglement entropies[BMM13] and generalized to its covariant form[HRT07].

A distinct feature of the holography entanglement entropy based on the RT formula is that the mutual information

$$I_{A:B} = S_A + S_B - S_{AB} \qquad (1.6)$$

vanishes between two disjoint boundary regions A and B that are far separated from each other[Hea10, Hea19], 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 generally not 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 [FLM13, EW15, DQSY20]. 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.

1.4   Does Gravity Come From Quantum Information?

One on hand the emergent holographic space provides an intuitive geometric description for the sophisticated structure of quantum many-body entanglement [HT07, CHH12, LRSV15, BHRT15]. On the other hand it opens up the possibility to construct emergent spacetime geometry from entanglement. To make progress undertsanding emergent spacetime geometry from quantum entanglement, it is useful to construct and study toy models that capture aspects of the entanglement-geometry duality. A family of such toy models[EV14, Swi12a, Swi12b, BPSW19] using tensor networks have been devised. Originally, tensor networks have been constructed as variational wavefunctions for strongly correlated systems. On one hand, as a wavefunction, a tensor network is a representation of a quantum state. On the other hand, it is a graph whose geometry controls many of its quantum mechanical properties, namely entanglement. As a result, tensor networks could be a good playground to investigate the duality between quantum entanglement and geometry. Indeed, tensor networks holography of different flavors have been studied in detailed. In particular, the random tensor network (RTN) states were shown to satisfy the RT formula in the large bond dimension limit[HNQ$^+$16]. Based on that idea, a deep learning algorithm called Entanglement Feature Learning[YYQ18] was developed, pioneering the use of machine learning to establish a holographic bulk description from the entanglement structure. This sets the starting point of this work.

1.5   Contribution Of This Work

In prior approaches to RTN holography, the dual bulk network geometry is assumed to be static, analogous to the classical gravity dual in the context of AdS/CFT correspondence. While those models recover the RT formula, they share the same feature that they do not capture mutual information between disjoint regions effectively for non-holographic state.

In comparison, this work makes progress by including the gravitational fluctuation in the

model. Not only it would capture more sophisticated entanglement structure but also allow us to learn about the 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 statistically fluctuating spatial geometry as well as matter. Applying our approach to free-fermion CFT state with large central charge, we uncover a statistical gravity model governed by Weyl field fluctuations propagating on the hyperbolic background geometry. Such model with fluctuating geometry was found to be much more effective in mediating mutual information than its classical analogue. Furthermore we show that the Weyl field fluctuation has the emergent bulk locality by studying its bulk correlation. By analysing the spectrum and the leading collective modes of the emergent gravity theory, familiar excitation such as s-wave, p-wave and d-wave emerge.

Chapter 2

Random Tensor Network Holography

The idea of tensor network holography[HNQ$^+$16, YYQ18] provides a concrete microscopic model directly connects quantum states to an emergent geometries. In such a model, the quantum state is defined on the boundary of a tensor network whose connectivity can be interpreted as a discrete representation of space. Conventionally, the bond dimensions are assumed to be fixed and large, such that the 2nd Rényi entropy of the boundary quantum state is mapped to the free energy of an Ising model whose couplings is a function of the bond dimensions. In this work, we consider a generalised RTN whose bond dimensions (Ising coupling) takes a non-trial joint distribution. This results to a statistical gravity model with fluctuating matter and geometry.

2.1   Construction of RTN states

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[HNQ$^+$16] 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)[QYY17, VPYL18]. The generalized RTN model that we will consider in this work is defined as follows: (i) A planar graph $G = (V,E)$ is given to describe the background

two-dimensional spatial geometry, where V denotes the vertex set and E denotes the edge set. $V = V_{\text{blk}} \cup V_{\text{bdy}}$ is divided into two subsets: the bulk V_{blk} and the boundary V_{bdy} sets, see Figure 2.1(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 Figure 2.1(a). (iii) A random pure state $|\psi_v\rangle \in \mathcal{H}_v \equiv \bigotimes_{e \in \partial v} \mathcal{H}_v^e$ is defined on every bulk vertex $v \in V_{\text{blk}}$. (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 $\mathcal{E}_{\text{RTN}} = \{|\Psi\rangle\}$ of pure states in the boundary Hilbert space $\mathcal{H}_{\text{bdy}} \equiv \bigotimes_{v \in V_{\text{bdy}}} \mathcal{H}_v$ by taking a (partial) projection in the bulk Hilbert space $\mathcal{H}_{\text{blk}} \equiv \bigotimes_{v \in V_{\text{blk}}} \mathcal{H}_v$ as

$$|\Psi\rangle = \langle \psi | \phi \rangle : |\psi\rangle = \bigotimes_{v \in V_{\text{blk}}} |\psi_v\rangle, |\phi\rangle = \bigotimes_{e \in E} |\phi_e\rangle. \tag{2.1}$$

The probability measure of $|\Psi\rangle$ in the RTN ensemble $\mathcal{E}_{\text{RTN}}$ is given by $P(|\Psi\rangle) = P(|\psi\rangle) P(|\phi\rangle)$. The vertex state distribution $P(|\psi\rangle) = \prod_{v \in V_{\text{blk}}} P(|\psi_v\rangle)$ 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 $\mathcal{H}_v$) identically. Conventionally, the edge (link) state distribution $P(|\phi\rangle)$ is assumed to 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 $\mathcal{H}_{\text{bdy}}^{\otimes k}$, its expectation value in the product state $|\Psi\rangle^{\otimes k}$ is defined to be

$$\langle O^{(k)} \rangle = \mathbb{E}_{|\Psi\rangle \in \mathcal{E}_{\text{RTN}}} \frac{\text{Tr}\left((|\Psi\rangle\langle\Psi|)^{\otimes k} O^{(k)}\right)}{\langle \Psi | \Psi \rangle^k}. \tag{2.2}$$

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
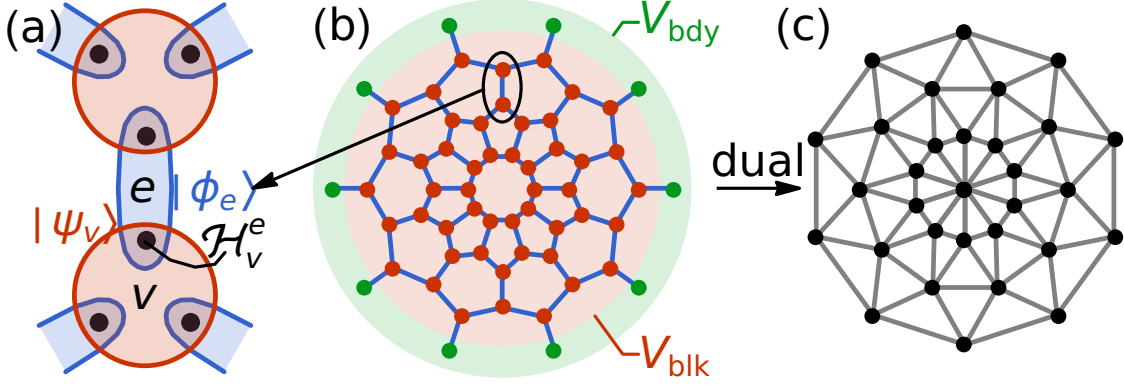
Figure 2.1: (a) Details of the RTN near an edge. (b) The planar graph G on which the RTN is defined. The vertices are classified into the bulk vertices (in red) and the boundary vertices (in green). (c) The dual graph $\tilde{G}$ of the RTN graph G.

approximate the ensemble average of the ratio by the ratio of separate averages,

$$\langle O^{(k)} \rangle \simeq \frac{1}{\mathcal{N}_k} \mathbb{E}_{|\Psi\rangle \in \mathcal{E}_{\mathrm{RTN}}} \mathrm{Tr} \left((|\Psi\rangle\langle\Psi|)^{\otimes k} O^{(k)} \right), \tag{2.3}$$

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),

$$e^{-S_A} \propto \mathbb{E}_{|\Psi\rangle \in \mathcal{E}_{\mathrm{RTN}}} \mathrm{Tr} \left((|\Psi\rangle\langle\Psi|)^{\otimes 2} X_A \right). \tag{2.4}$$

We will suppress the Rényi index throughout this work, 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.

10

2.2 Ising and Dual Ising Models

Evaluating the ensemble average in Eq. (2.4) following the approach developed in [HNQ$^+$16], 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

$$e^{-S_A} = \sum_{[\sigma,J]} P[\sigma|J]P[J]\,\delta[\sigma_{\text{bdy}} \Leftrightarrow A],\qquad(2.5)$$

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

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

$$E[\sigma|J] = -\sum_{e \in E}\left(\frac{J_e}{2}\prod_{v \in \partial e}\sigma_v\right),\qquad(2.6)$$

$$Z[J] = \sum_{[\sigma]}e^{-E[\sigma|J]}\delta[\sigma_{\text{bdy}} \Leftrightarrow \emptyset].$$

and $P[J]$ given by

$$P[J] = \int_{|\phi\rangle}\prod_{e \in E}\delta\big(J_e - S(|\phi_e\rangle)\big)P(|\phi\rangle).\qquad(2.7)$$

Here $\sigma_v = \pm 1$ is the Ising variable defined on every vertex $v \in V$, $J_e \geq 0$ 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,

$$\forall v \in V_{\text{bdy}} : \sigma_v = \begin{cases} +1 & v \notin A, \\ -1 & v \in A, \end{cases} \tag{2.8}$$

which is denoted as $\delta[\sigma_{\text{bdy}} \Leftrightarrow A]$ in Eq. (2.5). 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.

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 Ising model ($J_e \gg 1$). As a result, the boundary condition Eq. (2.2) implied by the entanglement region A creates disjoint domains that extend to the bulk. Thus, the free energy $F[A] = -\ln Z[J]$ is mainly controlled by the energy cost of the shortest domain walls $|\gamma_A|$ homologous to A. As a result, the RT formula

$$S_A = k|\gamma_A| \tag{2.9}$$

is recovered after identifying k the free energy cost of the domain wall per unit length as $\frac{1}{4G}$.

Given that G is a planar graph, we can use the Kramers-Wannier duality to rewrite the Ising model Eq. (2.5) on the dual lattice $\tilde{G} = (\tilde{V}, \tilde{E})$, as shown in Fig. 2.1(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

$$e^{-S_A} = \sum_{[\tilde{\sigma}, \tilde{J}]} \left(\prod_{\tilde{v} \in \partial A} \tilde{\sigma}_{\tilde{v}} \right) P[\tilde{\sigma}|\tilde{J}] P[\tilde{J}], \tag{2.10}$$

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

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

$$E[\tilde{\sigma}|\tilde{J}] = -\sum_{\tilde{e}\in\tilde{E}}\left(\frac{\tilde{J}_{\tilde{e}}}{2}\prod_{\tilde{v}\in\partial\tilde{e}}\tilde{\sigma}_{\tilde{v}}\right), \qquad (2.11)$$

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

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

$$P[\tilde{J}] = \left(\prod_e \frac{\partial J_e}{\partial\tilde{J}_{\tilde{e}}}\right)P[J]. \qquad (2.12)$$

Here $\tilde{\sigma}_{\tilde{v}} = \pm 1$ is the dual Ising variable and

$$\tilde{J}_{\tilde{e}} = -\ln\tanh(J_e/2) \qquad (2.13)$$

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). Therefore the purity of the RTN state can be interpreted as the boundary correlation of dual Ising variables $e^{-S_A} \propto \langle\prod_{\tilde{v}\in\partial A}\tilde{\sigma}_{\tilde{v}}\rangle$ in an Ising model with fluctuating couplings.
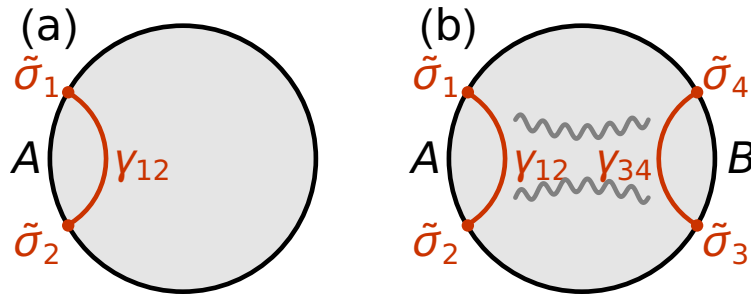
Figure 2.2: (a) Two-point correlation and (b) four-point correlation of dual Ising spins.

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 ($J_e \gg 1$). Equivalently the dual

Ising model is deep in the high temperature paramagnetic phase ($\tilde{J}_{\tilde{e}} \ll 1$). 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. 2.2(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 $e^{-S_{AB}} \propto \langle \tilde{\sigma}_1 \tilde{\sigma}_2 \tilde{\sigma}_3 \tilde{\sigma}_4 \rangle \sim e^{-\gamma_{12}/\xi} e^{-\gamma_{34}/\xi}$ in Fig. 2.2(b). Allowing the dual Ising coupling $\tilde{J}$ to fluctuate collectively will introduce perturbations to the geodesic distance $\gamma_{xy} \rightarrow \bar{\gamma}_{xy} + \delta\gamma_{xy}$ in a correlated manner, such that

$$
\begin{aligned}
e^{-S_{AB}} &\sim \mathbb{E}\left(e^{-\gamma_{12}/\xi} e^{-\gamma_{34}/\xi}\right) \\
&\sim \mathbb{E}e^{-\gamma_{12}/\xi} \mathbb{E}e^{-\gamma_{34}/\xi} e^{\frac{1}{2\xi^2} \mathbb{E}\delta\gamma_{12}\delta\gamma_{34}} \\
&\sim e^{-S_A} e^{-S_B} e^{I_{A:B}}.
\end{aligned}
\tag{2.14}
$$

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.

2.3 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[1] in the continuum limit

$$S[\phi|g] = \frac{1}{2} \int d^2x \sqrt{g} (g^{ij} \partial_i \phi \partial_j \phi + m^2 \phi^2), \tag{2.15}$$

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). 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[Reg61] to discretize the action,

$$S[\phi|\omega] = \sum_{\langle xy \rangle \in \tilde{E}} \frac{A_{xy}}{2} \left(\frac{\phi_x - \phi_y}{\ell_{xy}} \right)^2 + \sum_{x \in \tilde{V}} \frac{m^2 A_x}{2} \phi_x^2, \tag{2.16}$$

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.

A explicit relation between ℓ_{xy} and $\tilde{J}_{xy}$ can be found by comparing the correlation functions $\langle \tilde{\sigma}_x \tilde{\sigma}_y \rangle$ and $\langle \phi_x \phi_y \rangle$. By assumption $\tilde{J}_{xy} \ll 1$, it is therefore advantageous to employ the high-temperature tanh expansion

$$\langle \tilde{\sigma}_x \tilde{\sigma}_y \rangle = \frac{\sum_{\tilde{\sigma}} \tilde{\sigma}_x \tilde{\sigma}_y \prod_{\langle ij \rangle} e^{\frac{\tilde{J}_{ij}}{2} \tilde{\sigma}_i \tilde{\sigma}_j}}{\sum_{\sigma} \prod_{\langle ij \rangle} e^{\frac{\tilde{J}_{ij}}{2} \tilde{\sigma}_n \tilde{\sigma}_m}}$$

$$= \frac{\sum_{\tilde{\sigma}} \tilde{\sigma}_x \tilde{\sigma}_y \prod_{\langle ij \rangle} (1 + \tilde{\sigma}_i \tilde{\sigma}_j \tanh \frac{\tilde{J}_{ij}}{2})}{\sum_{\tilde{\sigma}} \prod_{\langle ij \rangle} (1 + \tilde{\sigma}_i \tilde{\sigma}_j \tanh \frac{\tilde{J}_{ij}}{2})}. \tag{2.17}$$

---

[1]For this mapping to make sense, we had assumed that the original network of the RTN model was chosen such that distances can be defined consistently throughout the bulk in the dual Ising model, via $\text{dist}(x, y) \propto -\ln \langle \tilde{\sigma}_x \tilde{\sigma}_y \rangle$. Otherwise, zero probability $P[\tilde{J}]$ will be assigned to such graph.

In the simplest case where there is a unique shortest path γ_{xy} joining sites x and y, we have

$$\langle \tilde{\sigma}_x \tilde{\sigma}_y \rangle = \prod_{\langle ij \rangle \in \gamma_{xy}} \tanh \frac{\tilde{J}_{ij}}{2} (1+...) \tag{2.18}$$

where the ellipses denotes corrections of higher order in $\tanh \tilde{J}_{ij}$ from longer paths. On the other hand, the scalar field correlations decays exponentially with geodesic distances as the leading behavior

$$\langle \phi_x \phi_y \rangle \approx K \exp \left(-m \sum_{\langle ij \rangle \in \gamma_{xy}} \ell_{ij} \right) = K \prod_{\langle ij \rangle \in \gamma_{xy}} e^{-m\ell_{ij}} \tag{2.19}$$

where K is a non-universal proportionality constant. As a result, by demanding the correlation functions $\langle \tilde{\sigma}_x \tilde{\sigma}_y \rangle$ and $\langle \phi_x \phi_y \rangle$ to match each other (up to the proportionality constant) for any arbitrary sites x and y, one conclude that

$$e^{-J_e} = \tanh \frac{\tilde{J}_{xy}}{2} \approx e^{-m\ell_{xy}} \tag{2.20}$$

where the duality relation (which has the same form as it inverse) Eq (2.13) is used in the first equality. One concludes

$$J_e = -m\ell_{xy}. \tag{2.21}$$

Thus, by trading the dual spin variables $\tilde{\sigma}_x$ to real scalar variables ϕ_y, we arrive at the following model

$$e^{-S_A} = \mathbb{E}_{\tilde{J}} \left\langle \prod_{x \in \partial A} \tilde{\sigma}_x \right\rangle_{\tilde{J}}$$

$$\rightarrow \mathbb{E}_{\ell} \left\langle \prod_{x \in \partial A} \phi_x \right\rangle_{\ell} = \int P[\ell] \left\langle \prod_{i \in \partial A} \phi_x \right\rangle_{\ell} \tag{2.22}$$

where $\langle \cdot \rangle_{\ell_{ij}}$ denotes the expectation value with respect to a set of fixed ℓ_{ij}. In the continuum limit, where the average over $\{\ell_{xy}\}$ is ammped to average over g, this model becomes by a statistical

16

gravity model

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

$$P[\phi|g] = \frac{e^{-S[\phi|g]}}{Z[g]}, \tag{2.23}$$

$$Z[g] = \int_{[\phi]} e^{-S[\phi|g]},$$

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. For brevity, the proportionality constant K in eq (2.19) has been absorbed in the definition of of the scalar field $\phi \to \phi/\sqrt{K}$.

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 $\bar{g}_{ij}(x)$

$$g_{ij}(x) = e^{2\omega(x)} \bar{g}_{ij}(x), \tag{2.24}$$

such that each Weyl field configuration represents a physically distinct geometry. In this gauge the continuum scalar field action is given by

$$S[\phi|\omega] = \frac{1}{2} \int d^2x \sqrt{g} (g^{ij} \partial_i \phi \partial_j \phi + m^2 e^{2\omega} \phi^2). \tag{2.25}$$

As a result, the integration $\int_{[g]} P[g]$ can be replaced by $\int_{[\omega]} P[\omega]$ in Eq. (2.23). The unknown joint distribution $P[\omega]$ will be what we aim to learn from the entanglement entropy data. The corresponding discrete action reads

$$S[\phi|\omega] = \sum_{\langle xy \rangle \in \tilde{E}} \frac{A_{xy}}{2} \left(\frac{\phi_x - \phi_y}{\ell_{xy}} \right)^2 + \sum_{x \in \tilde{V}} \frac{m^2 A_x}{2} e^{2\omega_x} \phi_x^2, \tag{2.26}$$

17

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[2] 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

$$
\begin{aligned}
e^{-S_A} &= \int_{[\phi, \omega]} \left(\prod_{x \in \partial A} \phi_x \right) P[\phi | \omega] P[\omega], \\
P[\phi | \omega] &= \frac{e^{-S[\phi | \omega]}}{Z[\omega]}, \\
Z[\omega] &= \int_{[\phi]} e^{-S[\phi | \omega]}
\end{aligned}
\tag{2.27}
$$

which is the underlying lattice model that will be used in the machine learning approach.

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.

This chapter is currently being prepared for submission for publication of the material. Lam, Jonathan C. C.; Yi-Zhuang, You. The dissertation author was the primary investigator and author of this material.

---

[2]The use of fixed ℓ_{xy} and fluctuating ω_x is, in the continuum limit, equivalent to saying ℓ_{xy} fluctuates in the gauge-fixed form $\ell_{xy} = e^{\omega_x} \bar{\ell}_{xy} e^{\omega_y}$ where $\bar{\ell}$ is a fixed background.

Chapter 3

Learning the Distribution of Geometries

Having interpreted RTNs with fluctuating bond dimensions as an ensemble average over geometry, a natural question follows: What could be the corresponding bulk geometric distribution $P[\ell]$ (or parameterised by $P[\omega]$ in the conformal gauge) given a set of entanglement features $e^{-S_{\text{data}}(A)}$ from an arbitrary quantum state living on the boundary? The problem of modelling high-dimensional probability distributions from data is not unique to physics. In fact, it has been one of the most challenging problems in statistics and machine learning and goes by the name *generative models*[GBC16]. In this chapter we review a conventional approach used in statistical physics to infer $P[\ell]$ from data. Then an alternative approach using neural networks — the route taken in this work, is discussed.

3.1   A Statistical Physics Approach

We want to find a possible $P[\ell]$ such that entropy produced by the holographic model matches the data as close as possible

$$e^{-S_{\text{model}}(A)} \overset{!}{=} e^{-S_{\text{data}}(A)} \tag{3.1}$$

where $e^{-S_{\text{model}}(A)} = \int D\ell P[\ell] \langle \prod_{x\in\partial A} \phi_x \rangle_\ell$ is the model defined in the previous chapter. To do so, a more conventional approach could be to postulate an effective Hamiltonian/action $S[\ell]$ such that ℓ's follows the Boltzmann distribution

$$P[\ell] = \frac{e^{-S[\ell]}}{Z} \tag{3.2}$$

where $Z = \int D\ell e^{-S[\ell]}$. When constructing $S[\ell]$, a common approach is to combine possible terms which are consistent with all the symmetries or gauge structure known to the system

$$S[\ell] = g_1 S_1[\ell] + g_2 S_2[\ell] + \cdots \tag{3.3}$$

where g_i are the coupling constants. Moreover, as a form of inductive bias, local terms are generally preferred, and organized by relevance — number of derivatives. Generally speaking, depending on the kind of question interested, a model with only a few terms may suffice. For example, to describe the long distance spin-spin correlation in paramganetic phase, an Ising model with only nearest neighbor coupling is good enough.

In Regge calculus' approach to quantum gravity[Reg61, Ham08], a smooth manifold is replaced by a triangulation made of simplices. The edges $\{\ell_{xy}\}$ between neighboring vertices play the role of the metric $ds^2 = g_{ij}dx^i dx^j$ in the smooth manifold. In this formalism, most geometric quantities derived from the metric have a lattice analogue in terms of $\{\ell_{xy}\}$. For example, curvature can be expressed in terms of *deficit angle* at a vertex, which can be computed using the edge variables $\{\ell_{xy}\}$. This is analogous to the computation of the Riemann curvature tensor using g_{ij}. As a result, it is possible to construct the lattice analogue of the Einstein-Hilbert action and any higher-derivative of Ricci scalar. A number of Monte Carlo study have been done along this direction with Euclidean path integral.

With this technology in hand, a natural, systematic approach to model $P[\ell]$ could be adding terms with increasing order of Ricci scalar and derivatives of it. To determine the

coupling constants g_i's, one may for example resort to techniques such as variational Monte Carlo to determine the best-fitted value for a given boundary entropy data. While this approach often produce models that can be interpreted easily, the approach has some drawbacks: It is rarely the case that expected value $\mathbb{E}_\ell O[\ell]$ of an observable $O[\ell]$ of interest can be computed exactly. To estimate it, one has to perform some kind of Markov Chain Monte Carlo (MCMC) sampling. A major difficulty in deploying such energy-based model of the form $P[\ell] = e^{-S[\ell]}/Z$ is the intractability of the partition function Z. A solution is to use sampling schemes akin to Metropolis-Hasting algorithms. However as the energy-action landscape gains complexity — multimodality, such system becomes frustrated and the acceptance ratio vanishes[1]. Therefore it becomes computationally prohibitive to make estimation with finite error bar as it is becoming harder to obtain uncorrelated samples. A famous example is the phenomenon of critical slow-down in Ising model. Therefore, traditional approaches to generative models often face the dichotomy between either (i) to use a simple model that is easy to sample or (ii) to use an expressive (e.g. energy-based) model but is expensive to sample.

In this work, we explored an alternative approach to model the geometric distribution parameterised by $P[\omega]$ in the conformal gauge. Instead of constructing an effective Hamiltonian and model the distribution by Boltzmann weights, we deployed a deep generative model (DGMs).

3.2   Deep Generative Models

Driven by the explosion of big data, there has been increasing interests in generative modelling in machine learning. In particular, with the rise of deep leaning, a new family of techniques known as deep generative models (DGM) has gained considerable attentions[2]. One of

---

[1]For continuous variables, Hamiltonian Monte Carlo algorithm (also known as hybrid Monte Carlo) could be a solution to low acceptance ratio due to rough energy landscape.

[2]Some authors also consider energy-based model such as Deep Boltzmann Machine as a form of deep generative model. However, we choose to reserve DGM to describe generative model using a feed-forward architecture, excluding energy-based models.

the major advantages of this approach is that it is easy to sample while being very expressive —
in stark contrast to the traditional energy-based model.

3.2.1   Basics of Deep Learning

The term neural network (NN) encompasses a large class of models and learning methods.
Here we describe the most 'vanilla' kind, which is also known as deep feedforward networks or
multilayer perceptrons (MLPs). The basic goal of such a model to approximate some function f^*.
These models are typically formed by composing layers of simpler functions $f^{(n)}$

$$f_\theta(x) = f^{(N)} \circ f^{(N-1)} \circ \cdots f^{(1)}(x) \tag{3.4}$$

where θ denotes the set of parameters that defines the transformation. The number of layers N
used gives the depth of the model, thus the name 'deep learning'. Typically, each layer is a simple
affine transformation followed an element-wise nonlinearity

$$f^{(n)} = \psi(W_n x + \beta_n) \tag{3.5}$$

where W_n is the weight matrix, β_n is the bias. Typical choices of nonlinearity function ψ include
the Rectified Linear Unit (ReLU) and sigmoid. It is natural to ask if any function f^* can be
approximated this way. The answer, originally due to Kolmogorov and Arnold, and continuously
refined by generations of machine learning theorists, is 'yes'. This collections of closely related
results are known as universal approximation theorems[Cyb89, HSW89].

Training a NN amounts to finding the optimal parameters $\{W_n\}$ and $\{\beta_n\}$ such that f best
approximates f^*. This is done is first constructing a loss function[3] $\mathcal{L}$ that measures, in some
sense, the difference between f and f^*. For example,given a collection of labels y_i's and features
x_i's, the data set $\mathcal{D} = \{(y_i, x_i)\}$ implicitly defines the target function f^* via $y_i = f^*(x_i)$. A simple

---

[3]Also known as cost or utility function.

example of possible loss function is the mean squared error (MSE)

$$\mathcal{L} = \sum_i |y_i - f(x_i)|^2. \tag{3.6}$$

Then by computing the gradient of $\partial \mathcal{L}$ with respect to the model parameters, the model can be updated via stochastic gradient descent (SGD) and its modern variants. The specific algorithm that computes the gradient with respect the parameters layer-by-layer via chain rule is called backpropagation[RHW86], which is also known as the reverse mode of automatic differentiation.

As a concluding remark, the use of deep neural network is not new. In fact the first working learning algorithm for deep neural network was published in 1967 known as group method of data handling[Sch15, IL67]. However, the hardware available at that time simply could not support such endeavor. It was until last decade or so, the availability of hardware accelerators such as graphical processing unit (GPU) and large scale data set (e.g. ImageNet[DDS$^+$09]) has revitalized the field. Furthermore, the publication of open-sourced deep learning frameworks such as TensorFlow[AAB$^+$15] and PyTorch[PGM$^+$19], in particular their implementations of backpropagation, have enabled the creations of many more sophisticated models.

3.2.2 Generative Neural Networks

The heart of deep generative models is that, instead of modelling the functional form of a probability distribution directly, it models a transformation from a (latent) variable z sampled from a known, tractable distribution[JM15, DSB16, KSJ$^+$16, PPM17]. It is still a field of active research. Popular examples of such kind of model include, to name a few, generative adversarial networks (GANs)[GPAM$^+$14], variational autoencoders (VAEs)[KW13] and autoregressive models such as PixelRNN[VOKK16].

More concretely, let $z \in \mathbb{R}^n$ a random variable drawn from a standard Gaussian distribution

$P_Z[z]$. Then by applying a neural network f_θ, a transformed random variable

$$x = f_\theta(z) \tag{3.7}$$

is obtained. An elementary result[4] from probability theory asserts that the transformed variable follows the distribution

$$P_X[x] = \left| \frac{\partial f_\theta^{-1}}{\partial x} \right| P_Z[z = f_\theta^{-1}(x)] \tag{3.8}$$

where $\left| \frac{\partial f_\theta^{-1}}{\partial x} \right|$ is the Jacobian of the transformation. As an example, the standard way to produce correlated Gaussian random variables x characterized by the mean μ and covariance Σ is to use a simple network with only one layer of affine transformation to a standard Gaussian z such that

$$x = \mu + Wz \tag{3.9}$$

where W is the Cholesky decomposition of Σ such that $\Sigma = WW^T$. Depending on the specific application, sometimes it is desirable to evaluate $P_X[x]$ explicitly. For example, when the training involves a loss function that contains the log-likelihood $\ln P_X[x_i]$ where x_i is a training sample. In that case, it is crucial to design the neural network such that it has tractable inverse and Jacobian one can see from eq (3.8). Such class of models are known as normalizing flows. It is also possible to design a loss function to avoid using explicit evaluation of $P_X[x]$, which is the approach taken in this work.

As a result, the task of 'learning a distribution P_X' is now rephrased as 'learning a transformation $x = f_\theta(z)$'. In most applications, such task can be archived by designing a loss function which involves averaging $g(x)$ over the target distribution P_X, for some function g that measure the 'goodness' of the distribution of X. Multiplying both sides by $g(x)$ and integrate

---

[4]In one-dimension, $|P_X(x)dx| = |P_Z(z)dz|$.

against x on eq (3.8), one concludes

$$\mathbb{E}_{X|\theta}g(x) = \mathbb{E}_Z g(f_\theta(z)) \qquad (3.10)$$

where the notation $\mathbb{E}_{X|\theta}$ is used to emphasize the fact that the distribution of x implicitly depends on the neural network parameters θ via eq (3.8). As a result, to compute the gradient of the loss function, one has to evaluate $\partial_\theta \mathbb{E}_{X|\theta}g(x)$. One of the the most appealing feature of DGMs is that, the r.h.s of eq (3.10), and its derivatives, can be estimated efficiently by sample mean as uncorrelated samples of z can be generated readily without the need to run MCMC. Namely

$$\mathbb{E}_Z g(f_\theta(z)) \approx \frac{1}{N} \sum_{i=1}^{N} g(f_\theta(z_i)) \qquad (3.11)$$

$$\partial_\theta \mathbb{E}_Z g(f_\theta(z)) \approx \frac{1}{N} \sum_{i=1}^{N} \partial_\theta g(f_\theta(z_i)) \qquad (3.12)$$

where N is the number of independent samples used in the estimation. The advances in both hardware and software has made such computation very practical. For one, independent samples of $g(f_\theta(z_i))$ can be generated and evaluated in parallel very efficiently. Moreover, the derivative $\partial_\theta g(f_\theta(z_i))$ can be computed readily by backpropagation, as implemented in any open-sourced deep learning frameworks. This approach to computing gradients of an expectation is often called the 'reparameterization trick'[KB14]. With this technique, we now have a self-contained algorithm to learn the holographic geometric distributions $P[\omega]$ from the set of purity $\{e^{-S_{\text{data}}(A)}\}$ of any generic quantum state.

3.2.3 Symmetry and Architecture

It is both practical and theoretically motivated to enforce the neural network's architecture such that the learned distribution will respects certain symmetries, i.e. to construct an 'equivariant neural network'[CW16]. This can be achieved by restricting the functional form — by properly

designing the architecture of the network.

More specifically, let Q represents a symmetry transformation that we wish to impose on the target distribution P_X. We wish to strictly enforce that

$$P_X[Qx] = P_X[x]. \tag{3.13}$$

Recall a typical generative neural network consists of consecutive layers of simple transformations

$$x = f_N \circ \cdots f_2 \circ f_1(z) \tag{3.14}$$

to a latent variable z which is sampled from a simple distribution such as a standard Gaussian. A sufficient condition for eq (3.13) to hold is to require the same condition holds at each layer. In other words, $P[Qz_n] = P[z_n]$ where z_n is the output of the nth layer. It turns out that this is guaranteed if the two conditions (i) $P[Qz_{n-1}] = P[z_{n-1}]$ and (ii) $Qf_n(z_{n-1}) = f_n(Qz_{n-1})$ are satisfied. To see that, using eq (3.8),

$$
\begin{aligned}
P[Qz_n] &= \left| \frac{\partial f_n(Qz_{n-1})}{\partial Qz_{n-1}} \right|^{-1} P[Qz_{n-1}] \\
&= \left| \frac{\partial f_n(Qz_{n-1})}{\partial Qz_{n-1}} \right|^{-1} P[z_{n-1}] \qquad \text{[by (i)]} \\
&= \left| \frac{\partial Qf_n(z_{n-1})}{\partial Qz_{n-1}} \right|^{-1} P[z_{n-1}] \qquad \text{[by (ii)]} \\
&= \left| \frac{\partial f_n(z_{n-1})}{\partial z_{n-1}} \right|^{-1} P[z_{n-1}] = P[z_n].
\end{aligned}
\tag{3.15}
$$

Thus, by appropriately restricting the functional form of each layer one can strictly enforce certain symmetry in the learned distributions. For example, the use of convolutional layers results in translation symmetry.

3.3 General Algorithm

Given a set of entanglement features $\{e^{-S_{\text{data}}(A)}\}$ of various choices of subregion A, our goal is to find a possible $P[\omega]$ such that entropy produced by the holographic model matches the data as close as possible $e^{-S_{\text{model}}(A)} \stackrel{!}{=} e^{-S_{\text{data}}(A)}$. Here we describe the procedure to do so. An overview of the algorithm is illustrated in Figure 3.1.
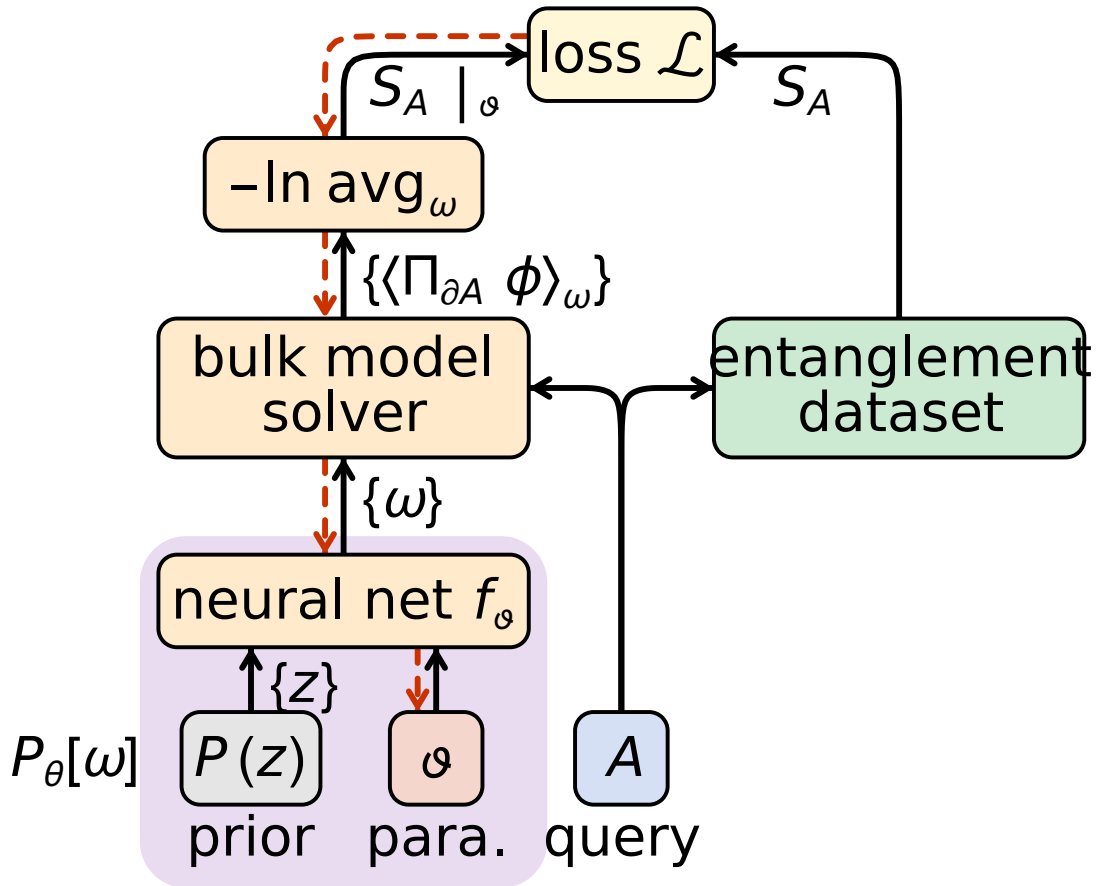
Figure 3.1: Flow diagram of the machine learning algorithm. Black arrows denotes the forward evaluation of the loss function. Red dashed arrows denotes the gradient back propagation to train the parameter.

To recap, we have a holographic model of a fluctuating space Σ, which is characterised by the background $\{\ell_{xy}\}$ and Weyl field $\{\omega_x\}$. The model describes the entanglement entropy of

system defined on the boundary $\partial\Sigma$ via

$$e^{-S_{\text{model}}(A)} = \mathbb{E}_\omega \left\langle \prod_{x\in\partial A} \phi_x \right\rangle_\omega = \int P[\omega] \left\langle \prod_{x\in\partial A} \phi_x \right\rangle_\omega. \tag{3.16}$$

Therefore, to compute the $S_{\text{model}}(A)$, one has to first compute the scalar field correlation $\left\langle \prod_{x\in\partial A} \phi_x \right\rangle_\omega$ at fixed Weyl field configuration and then average over the Weyl field ω.

3.3.1 Scalar Field Correlation Functions

For any given Weyl field configuration $\{\omega_i\}$, the bulk scalar field is Gaussian

$$S[\phi|\omega] = \frac{1}{2}\left\{ \sum_{\langle xy\rangle} A_{xy}\left(\frac{\phi_x - \phi_y}{\ell_{xy}}\right)^2 + \sum_x m^2 V_i e^{2\omega_x}\phi_x^2 \right\}$$

$$=: \sum_{x,y} \phi_x [L_{xy} + \text{diag}(m^2 V_k e^{2\omega_k})]\phi_y \tag{3.17}$$

where

$$L_{xy} = \begin{cases} -A_{xy}/\ell_{xy}^2 & \text{if } x,y \text{ are neighbor} \\ \sum_{k\in\text{neighbor}(x)} V_{xk}/\ell_{xk}^2 & \text{if } x = y \\ 0 & \text{otherwise} \end{cases} \tag{3.18}$$

is the graph Laplacian. Thus, the 2-point functions can be computed via matrix inverse

$$\left\langle \phi_x \phi_y \right\rangle_\omega = [L + \text{diag}(m^2 V_k e^{2\omega_k})]_{xy}^{-1}. \tag{3.19}$$

Higher-order correlations follow from Wick's theorem. For example,

$$\left\langle \phi_1\phi_2\phi_3\phi_4 \right\rangle_\omega = \left\langle \phi_1\phi_2 \right\rangle_\omega \left\langle \phi_3\phi_4 \right\rangle_\omega + \left\langle \phi_1\phi_3 \right\rangle_\omega \left\langle \phi_2\phi_4 \right\rangle_\omega + \left\langle \phi_1\phi_4 \right\rangle_\omega \left\langle \phi_2\phi_3 \right\rangle_\omega. \tag{3.20}$$

3.3.2 Weyl Field Generator

Using the techniques described in session 3.2.2, the problem of learning the distribution $P[\omega]$ is now recast into finding an optimal transformation implemented by a neural network. In particular, the average over geometries now reads

$$\mathbb{E}_\omega \left\langle \prod_{x\in\partial A} \phi_x \right\rangle_\omega = \mathbb{E}_z \left\langle \prod_{x\in\partial A} \phi_x \right\rangle_{\omega=f_\theta(z)} \tag{3.21}$$

where f_θ is a neural network parameterised by θ and z is a standard Gaussian random variable.

3.3.3 Loss Function

A possible way to measure similarity between the data and the model prediction is the mean squared relative error (MSRE) loss:

$$\mathcal{L}_{\text{data}} = \frac{1}{|\mathcal{A}|} \sum_{A\in\mathcal{A}} \left| \frac{e^{-S_{\text{data}}(A)} - e^{-S_{\text{model}}(A)}}{e^{-S_{\text{data}}(A)}} \right|^2 \tag{3.22}$$

which is approximately equal to $\frac{1}{|\mathcal{A}|} \sum_{A\in\mathcal{A}} |S_{\text{data}}(A) - S_{\text{model}}(A)|^2$. In general, the set $\mathcal{A}$ of all possible choices of entanglement region A grows exponentially as $2^\mathcal{N}$ where $\mathcal{N}$ is the size of the quantum system of interest. The standard approach to such big data set is to only select a random subset of sample A in each iteration of gradient descent. This creates a sources of randomness and lead to stochastic gradient descent. Alternatively one can also simply use a truncated data set throughout. E.g. a set of A's that only contains at most two disjoint regions. This is the approach we have taken in this work. In case of degeneracy where there are more than one geometric distribution $P_\theta[\omega]$ giving the same $\mathcal{L}_{\text{data}}$, regularization can be introduced by adding appropriate perturbations to $\mathcal{L}_{\text{data}}$. See section 4.7.2 for more details.

With a concrete loss function in hand, one can optimize of the parameters of the network θ, as well as other parameters in the model such as the normalization constant in eq (2.19) by

gradient descent using the 'reparameterization trick' introduced in section 3.2.2, i.e. computing by sampling derivative of the l.h.s of eq (3.21). Notice that the statistical noise from estimated gradient is also a source of randomness. Upon convergence, a neural network that generates samples of geometry $\{\omega_x\}$ is at our disposal. Questions regarding various statistical properties of $\{\omega_x\}$ can be answered by interrogating the trained network.

  This chapter is currently being prepared for submission for publication of the material. Lam, Jonathan C. C.; Yi-Zhuang, You. The dissertation author was the primary investigator and author of this material.

Chapter 4

Numerical Results

We apply the method described in the previous chapter to a free fermion system, which does not have a classical gravity dual. We found that bulk geometric fluctuation is an effective way to mediate mutual information. Upon training completion, the learned geometric distribution was analyzed. In particular, locality of Weyl field correlation emerged, suggesting the existence of of a local effective action.

4.1   Entanglement Entropy Data

While efficient experimental approaches[BEJ$^+$19, HKP20] have been developed to estimate Rényi entropies from randomized measurements, which enables the acquisition of large amount of entanglement data to drive the entanglement feature learning, preparing a 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 the ground state of a free fermion system for demonstration purpose. Consider N copies of the (1+1)D massless Majorana fermion

chain, described by the Hamiltonian

$$H = \sum_{a=1}^{N} \sum_{j} i\chi_{j,a}\chi_{j+1,a} \tag{4.1}$$

where $\{\chi_{n,a}, \chi_{m,b}\} = \delta_{nm}\delta_{ab}$. Each unit cell contains two sites. The long distance behavior of this critical system is described of a $(1+1)$D CFT with central charge $c = N/2$. Asymptotically, the nth Renyi entropy of a single-region A reads[CC04]

$$S_A^{(n)} = \frac{c}{6}\left(1 + \frac{1}{n}\right)\ln\left(\frac{|A|}{\delta}\right) \tag{4.2}$$

where δ is a non-universal constant acting as a UV-cutoff. For arbitrary choice of subregion A (e.g. when A consists of multiple disjoint subregions), the 2nd-Renyi entropy can be efficiently computed from the fermion correlation function[PE09]. More specifically, let $[C_A]_{ij} = \langle\chi_i\chi_j^\dagger\rangle_A$ be the ground state correlation matrix restricted to the subregion A. The 2nd Rényi entropy reads[PE09]

$$S_A^{(2)} = \frac{1}{2}\operatorname{Tr}\ln\left(C_A^2 + (1 - C_A)^2\right). \tag{4.3}$$

When collecting the entropy, entanglement cuts are always placed between the unit cells. In other words, the region A always contain complete unit cells. In the following, a chain of 32 sites (16 unit cells) with periodic boundary condition is used.

4.2 Model Specifications

4.2.1 Background Geometry

Since we intend to apply our approach to CFT data, it is natural to choose the two-dimension hyperbolic geometry (the spatial slice of AdS$_3$) as the background geometry. We employed the following global coordinate metric $s^2 = d\rho^2 + \sinh^2\rho\, d\theta^2$ where $0 \le \theta < 2\pi$ and

$0 \leq \rho \leq \rho_{bdy}$. Here ρ_{bdy} sets the IR cutoff of the hyperbolic space which is related to the UV cutoff scale of the holographic CFT. The geodesic distance between any two points on the boundary $\rho = \rho_{bdy}$ separated by θ is given by

$$\gamma(\theta) = \text{arccosh}(1 + 2\sinh^2\rho_{bdy}\sin^2(\theta/2)) \xrightarrow{e^{\rho_{bdy}} \gg 1} 2\left(\ln\sin(\theta/2) + \rho_{bdy}\right). \qquad (4.4)$$

Without loss of generality, we chose to discretize space with a triangular lattice with periodic
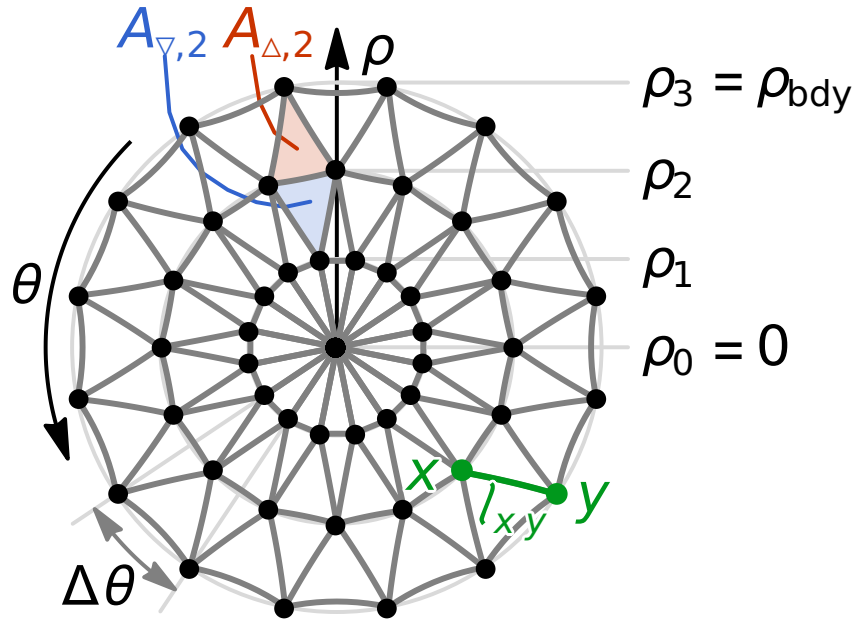
Figure 4.1: Triangular lattice in (ρ, θ) coordinate.

boundary condition along the θ-direction. Each layer contains 16 vertices matching the number of unit cells in the fermion chain used. All vertices in the same layer are of the same ρ-coordinate and their θ-coordinates are uniformly spaced, see Figure 4.1. The geodesic distance ℓ_{xy} between two vertices x and y is given by

$$\cosh\ell_{xy} = \cosh\rho_x\cosh\rho_y - \sinh\rho_x\sinh\rho_y\cos(\theta_x - \theta_y). \qquad (4.5)$$

The area of an elementary triangle in the ith layer reads

$$\tan \frac{A_{\triangle,i}}{4} = \tanh\left(\frac{\rho_{i+1} - \rho_i}{2}\right) \tanh\left(\frac{b_i}{4}\right), \tag{4.6}$$

$$\tan \frac{A_{\triangledown,i}}{4} = \tanh\left(\frac{\rho_i - \rho_{i-1}}{2}\right) \tanh\left(\frac{b_i}{4}\right), \tag{4.7}$$

$$\cosh b_i = \cosh^2 \rho_i - \sinh^2 \rho_i \cos \Delta\theta, \tag{4.8}$$

which defines the vertex and edge areas in the barycentric scheme[Ham08]. Specifically, the

vertex area A_x is given by

$$A_x = \frac{1}{3}(2A_{\triangle,i} + 2A_{\triangledown,i} + A_{\triangle,i-1} + A_{\triangledown,i+1}), \tag{4.9}$$

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

$$A_{xy} = \begin{cases} \frac{1}{3}(A_{\triangle,i} + A_{\triangledown,i}) & \rho_x = \rho_y = \rho_i; \\ \frac{1}{3}(A_{\triangle,i} + A_{\triangledown,i+1}) & \rho_x = \rho_i, \rho_y = \rho_{i+1}; \\ \frac{1}{3}(A_{\triangle,i-1} + A_{\triangledown,i}) & \rho_x = \rho_i, \rho_y = \rho_{i-1}. \end{cases} \tag{4.10}$$

These equations defines the choice of ℓ_{xy}, A_x and A_{xy}, which all relies on the values of ρ_i in

each layer. Thus the discretization scheme in the radial dimension is specified by how ρ_i is

spaced from ε to ρ_{bdy} where $\varepsilon = 0.1$ is used instead of strictly zero for better numerical stability.

Different choice of the discretization scheme is just a coordinate re-parametrization which should

not matter in the continuum limit. However, for a finite lattice $\{\rho_i\}$ would to fine-tuned according

to the entropy data in the following way.

For concreteness, we have prepared a data set $\{e^{-S_{\text{data}}}(A)\}$ using the free Majorana chain

with $c = 8$ which contains only single region A's. Then a 'classical' version of our model

described in eq (3.16), with static Weyl field $\omega_i \equiv 0$

$$e^{-S_{\text{model}}(A)} = \left\langle \prod_{i \in \partial A} \phi_i \right\rangle_{\omega \equiv 0} \tag{4.11}$$

is used to predict the purity $e^{-S_{\text{model}}}(A)$ for all choices of single region A. Together with radial coordinates $\{\rho_i\}$, the normalization constant K and the the scalar field mass parameter m will also be treated as trainable parameters of the model. Using the MSRE loss as defined in eq (3.22), the gradient with respect to each trainable parameters were computed and then used to update their values. These procedures are implemented in the TensorFlow[AAB$^+$15] framework using the Adam[KB14], a modern variant of SGD. Upon convergence, the single-region MSRE loss of the trained model has decreased to $\sim 10^{-6}$. This indicates a classical model without any geometric fluctuations, is capable to describe the single region entropy as predicted by the RT formula. See Figure (4.2). As a benchmark, the double-region RMSE was evaluated to be $\sim 10^{-1}$ with the trained model. Notice that the model did not use any double-region data in training for this benchmark.

The situation becomes very different once multi-region entanglement entropies are introduced to the training data set. As we argued, static bulk geometry is not a very efficient way to encode mutual information. This is seen by training the same classical model with a data set $\{e^{-S_{\text{data}}}(A)\}$ containing both single- and double-region A's. Upon convergence, the both single- and double-region RMSE have decreased to $\sim 10^{-2}$. In particular, the single-region RMSE is orders of magnitude larger than before. This can be seen as a compromise of the classical model trying to match both kinds of data as much as possible. As a comparison, the full model with fluctuating Weyl (which is discussed in the next sections) does not face such dilemma. See Figure (4.3). The implication of mutual information is discussed in section 4.4. A summary of these results using a classical model can be found in Table 4.1, together with those from a model with fluctuating Weyl field, which will be discussed in the next section.
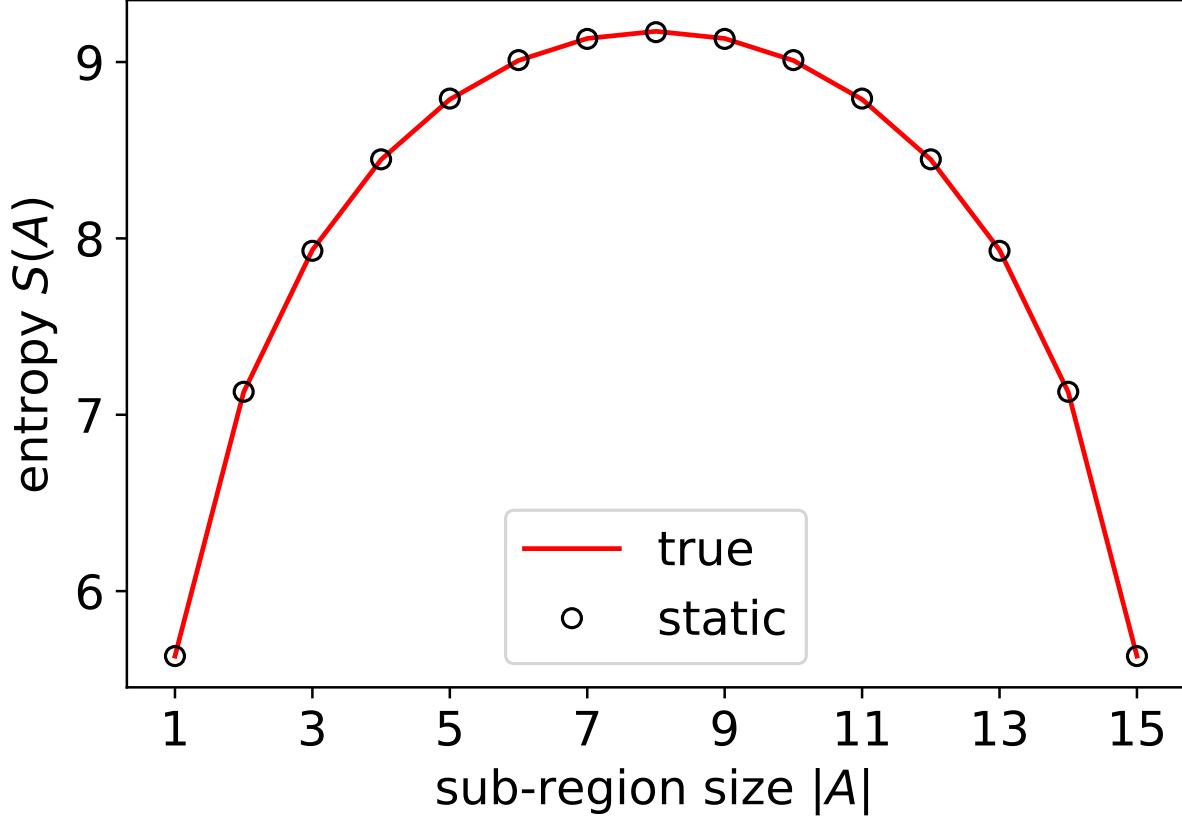
Figure 4.2: True vs predicted single-region second Rényi entropy. The 'classical' model with static bulk geometry is trained on a data set of a $c = 8$ free Majorana fermion chain with 16 unit cells consisting all possible choices of single sub-region.

Table 4.1: A summary of three numerical experiments. As shown in column (a), the classical model with static geometry is trained on single sub-region data of a $c = 8$ free Majorana fermion chain with 16 unit cells. The trained model is then asked to predict (i) 1 sub-region (ii) 2 sub-regions and (iii) 3 sub-regions entropy yielding the respective RMSE. The same model is then then trained on both single sub-region and two single sub-regions data as shown in column (b). In column (c), the full model with fluctuating Weyl field ω_x sampled from from the generative network is trained on the same data set of in column (b).

Model	(a) static geo.	(b) static geo.	(c) fluctuating geo.
Traning RMSE	1 sub-region	1, 2 sub-region(s)	1, 2 sub-region(s)
(i) 1 sub-region	8.7×10^{-6}	2.1×10^{-2}	1.5×10^{-3}
(ii) 2 sub-regions	1.1×10^{-1}	3.9×10^{-2}	5.7×10^{-3}
(iii) 3 sub-regions	7.5×10^{-1}	6.0×10^{-1}	3.1×10^{-1}
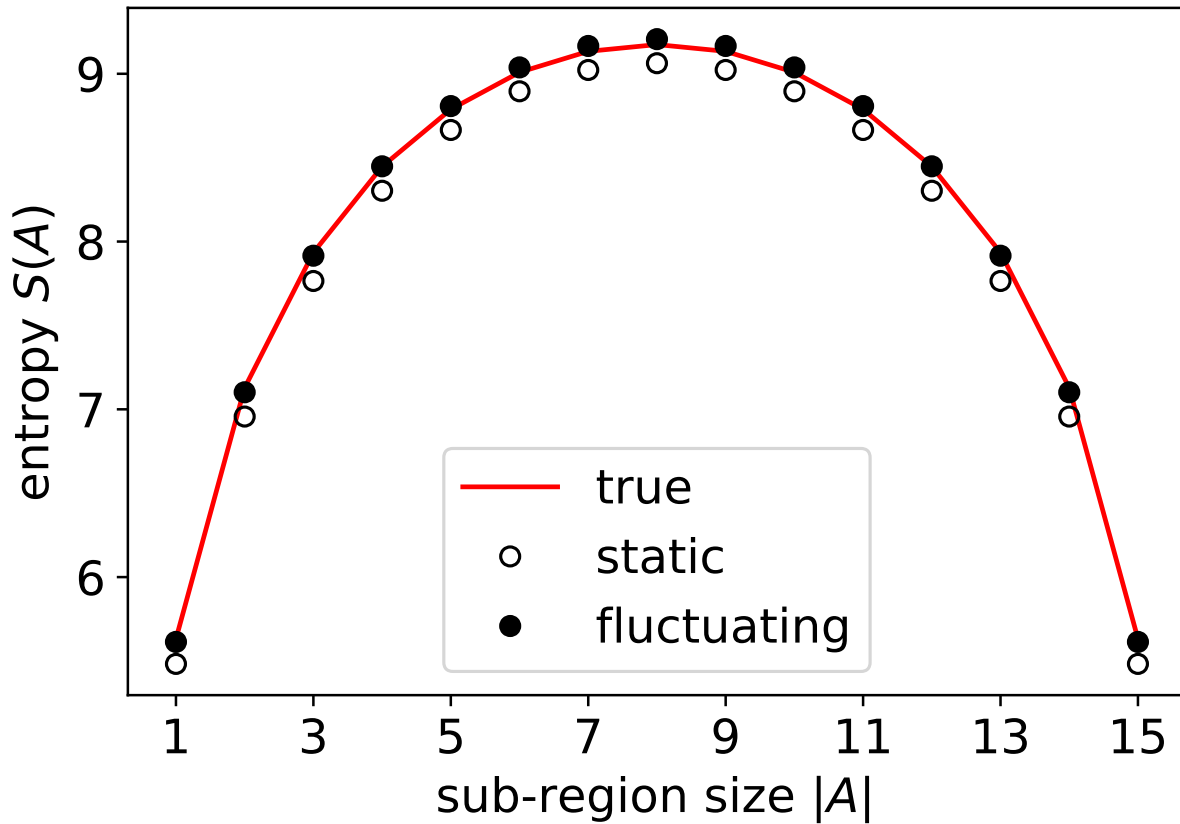
Figure 4.3: True vs predicted single-region second Rényi entropy. Both the 'classical' model with static geometry and the full model with fluctuating Weyl field are trained on the same data set of a $c = 8$ free Majorana fermion chain with 16 unit cells consisting both all possible choices of single sub-region and two disjoint sub-regions.

4.2.2 Network Architecture

When designing the generative network for the Weyl field ω, we have strictly enforced $P[\omega]$ to be invariant under the two symmetries

$$\omega(\theta, \rho) \rightarrow \quad \omega(\theta + a, \rho) \qquad \text{[translation]} \tag{4.12}$$

$$\omega(\theta, \rho) \rightarrow \quad \omega(-\theta, \rho) \qquad \text{[reflection]} \tag{4.13}$$

in response to the Hamiltonian eq (4.1) used to generate boundary data. As discussed in section (3.2.3), this can enforced layer-wise by adopting an appropriate architecture such as a convolutional layer. The standard 1-D convolutional[1] layer computes the following:

$$x[i, j] = \sum_{i', j'} f[i', j'] z[i + i', j'] \tag{4.14}$$

where z is the input array and f is a filter array with trainable elements. Notice that only the first dimension (length) of the input array is convolved while second one (channels) are mixed. As a result the output distribution preserves translation symmetry in the first dimension, for example

$$\mathbb{E}x[i, j]x[n, m] = \mathbb{E}x[i + k, j]x[n + k, m] \tag{4.15}$$

assuming $\mathbb{E}z[i, j]z[n, m] = \mathbb{E}z[i + k, j]z[n + k, m]$ for any k, while allowing more general dependence along the second dimension. As a fundamental building block of many computer-vision deep learning models, the GPU implementations of convolutional layers have been highly optimized. It is therefore of practical advantage to adapt such operations the Weyl field generator. Had the Weyl field been defined on a square lattice, one can directly use this elementary operation by identifying the first dimension as the θ- while the second one as the ρ-coordinate, such that the $P[\omega]$ to be invariant under translation $\theta \rightarrow \theta + a$. A complication arose due to the use of

---

[1]which actually computes the cross-correlation which amounts to redefining the filter to its mirror image.

a triangular lattice instead of a square lattice. This was solved by relabelling the sites on the triangular lattice along a zigzag path along the ρ-direction to have the same first array index before invoking the standard convolution. A mirrored operation is then applied, which is equivalent to having a symmetric filter $f[i', j'] = f[-i', j']$.

Thus, by stacking convolutional layers, one obtain a generative network with the desired symmetry. Theoretically, the deeper network, the more expressive it would be. But it is also more expensive to train. Thus, the depth of a network is often treated as a hyperparameter which modellers might find a balance between parsimony and performance. Incidentally, we found that a single convolution layer without non-linearity, suffice to illustrate the role of geometric fluctuations in mediating mutual information. Consequently, in the following the Weyl field is generated by convolving a standard Gaussian z. I.e. $\omega = Wz$ where W is the matrix that defines the convolution in the triangular lattice as described above.

4.3   Training

Having obtained the background geometry characterised by $\{\ell_{ij}\}$ from the last section, we would now introduce the complete model with a fluctuating Weyl field. From now on, the background geometry characterised by $\{\ell_{ij}\}$ as well the scalar field normalization K are frozen. However, the scalar field mass m could acquire correction due to fluctuating geometry. Thus it would be allowed to adjust as a part of the optimization. The renormalization of m would be discussed in detail in the section 4.6. More training details such as use of regularisation are discussed in section 4.7.

4.4 Mutual Information

Upon convergence, the single- and double-region RMSE loss converge to $\sim 10^{-4}$ and 10^{-3} respectively, clearly outperforming the 'classical' version described in section 4.2.1. The utility of geometric fluctuation can be seen even more clearly by computing the mutual information between two disjoint sub-regions, as shown in Figure (4.4). This is in accordance with our expectation that geometric fluctuation is an effective way to mediate mutual information.

## 4.5 Properties of the Learned Geometric Distribution

With the trained generative network, one can study various statistical properties of the learned Weyl field distribution $P[\omega_i]$. In the simplest case where the distribution is a Gaussian with zero mean, it suffices to study the covariance matrix

$$\Sigma_{\text{model}} = \mathbb{E}\omega\omega^T. \tag{4.16}$$

Firstly, from the diagonal entries we saw the fluctuations $\mathbb{E}\omega_i^2$ grows as approaching the boundary as shown in Figure 4.5. As a passing note we remark that a similar phenomenon is also known for massive scalar field in AdS space. Secondly, the correlation between Weyl field in the same layer closely follows an exponential decay

$$C_{xy} := \frac{\mathbb{E}\omega_x\omega_y}{\sqrt{\mathbb{E}\omega_x^2\mathbb{E}\omega_y^2}} \sim \exp(-\Delta|\gamma_{xy}|) \tag{4.17}$$

where Δ is the inverse correlation length, $|\gamma_{xy}|$ is the geodesic distance between. This is an unequivocal sign of locality. See Figure 4.7. In other words, the machine has come to the conclusion that a local Weyl field is the optimal way to encode holographic entanglement entropy. This is remarkable in light of locality was never explicitly given to the generative model during
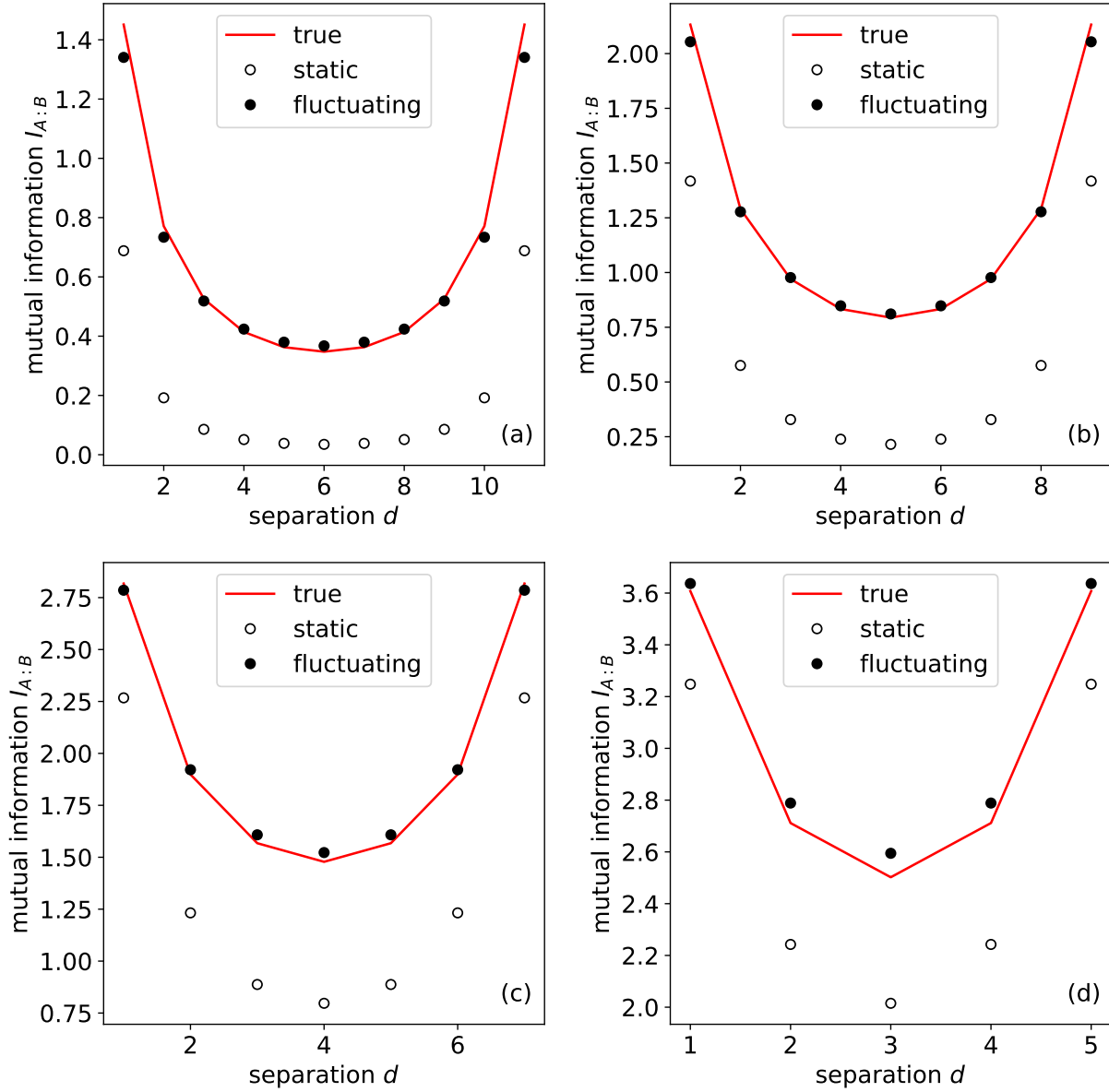
Figure 4.4: Mutual information between two disjoint sub-regions A and B of equal size (a) $|A| = |B| = 2$, (b) $|A| = |B| = 3$, (c) $|A| = |B| = 4$ and (d) $|A| = |B| = 5$ separated by distance d. Both the 'classical' model with static geometry and the full model with fluctuating Weyl field are trained on the same data set of a $c = 8$ free Majorana fermion chain with 16 unit cells consisting both one sub-region and two disjoint sub-regions. The distance d is measured in unit cells.
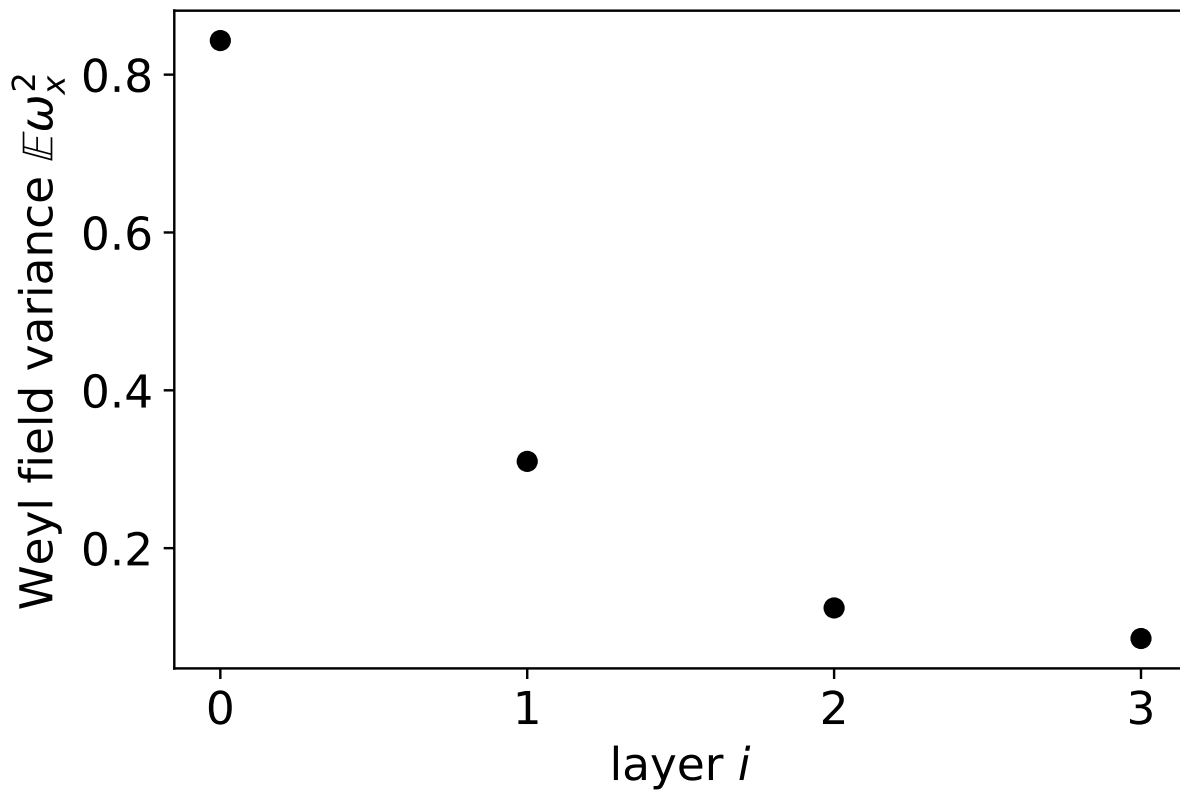
Figure 4.5: Weyl field variance $\mathbb{E}\omega_x^2$ of the learned distribution as a function of layer i.
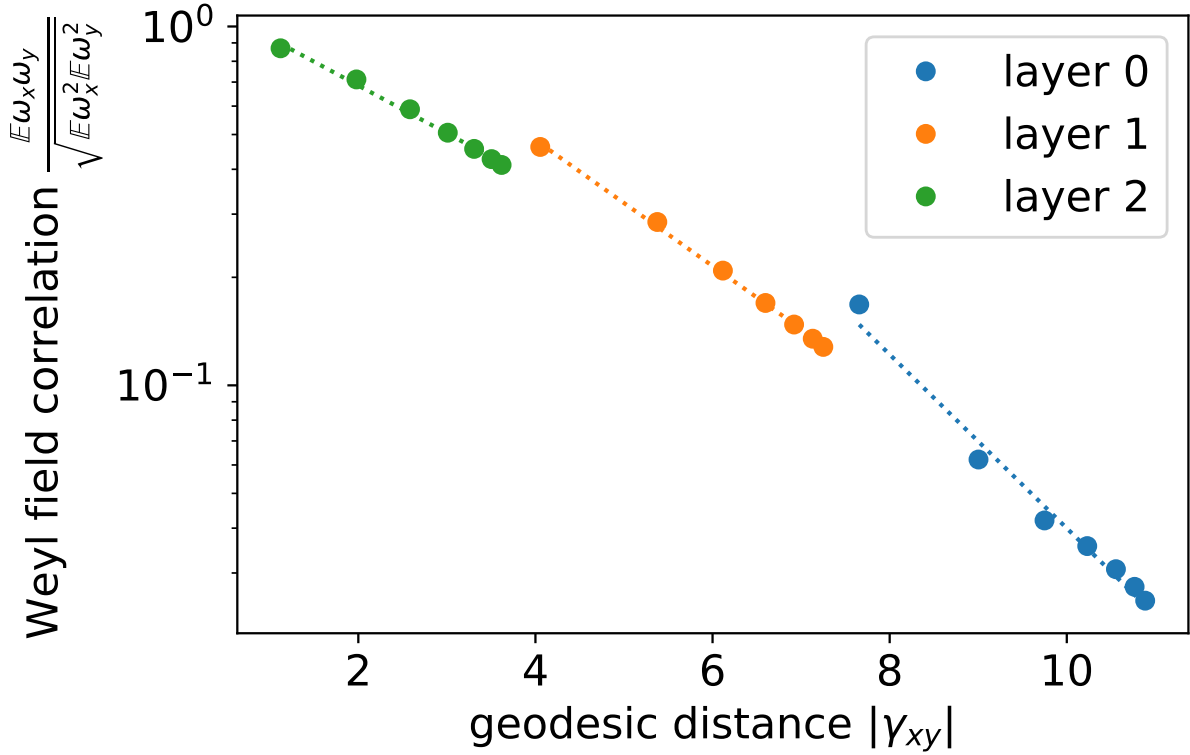
Figure 4.6: The learned Weyl field correlation $\mathbb{E}\omega_x\omega_y/\sqrt{\mathbb{E}\omega_x^2\mathbb{E}\omega_y^2}$ between two sites $x \neq y$ from the same layer i is plotted against the geodesic distance $|\gamma_{xy}|$. Due to the hyperbolic nature of space, the distance between neighboring points within the same layer is larger when closer to the boundary layer 0. A best fitting exponential decay is shown (dotted) for each layer, with inverse correlation length Δ found to be (i) 0.56 (layer 0), (ii) 0.40 (layer 1) and (iii) 0.31 (layer 2), indicating a larger decay rate for larger distance. As similar effect is also known for massive scalar field. In flat space the inverse correlation length Δ is simply given by the mass m. However in hyperbolic space i.e. spatial slice of AdS_d of unit radius, it receives a correction such that for large distance $\Delta = (d + \sqrt{d^2 + 4m^2})/2 > m$, while for short distance it remains $\Delta \sim m$ as curvature of space is not noticeable.
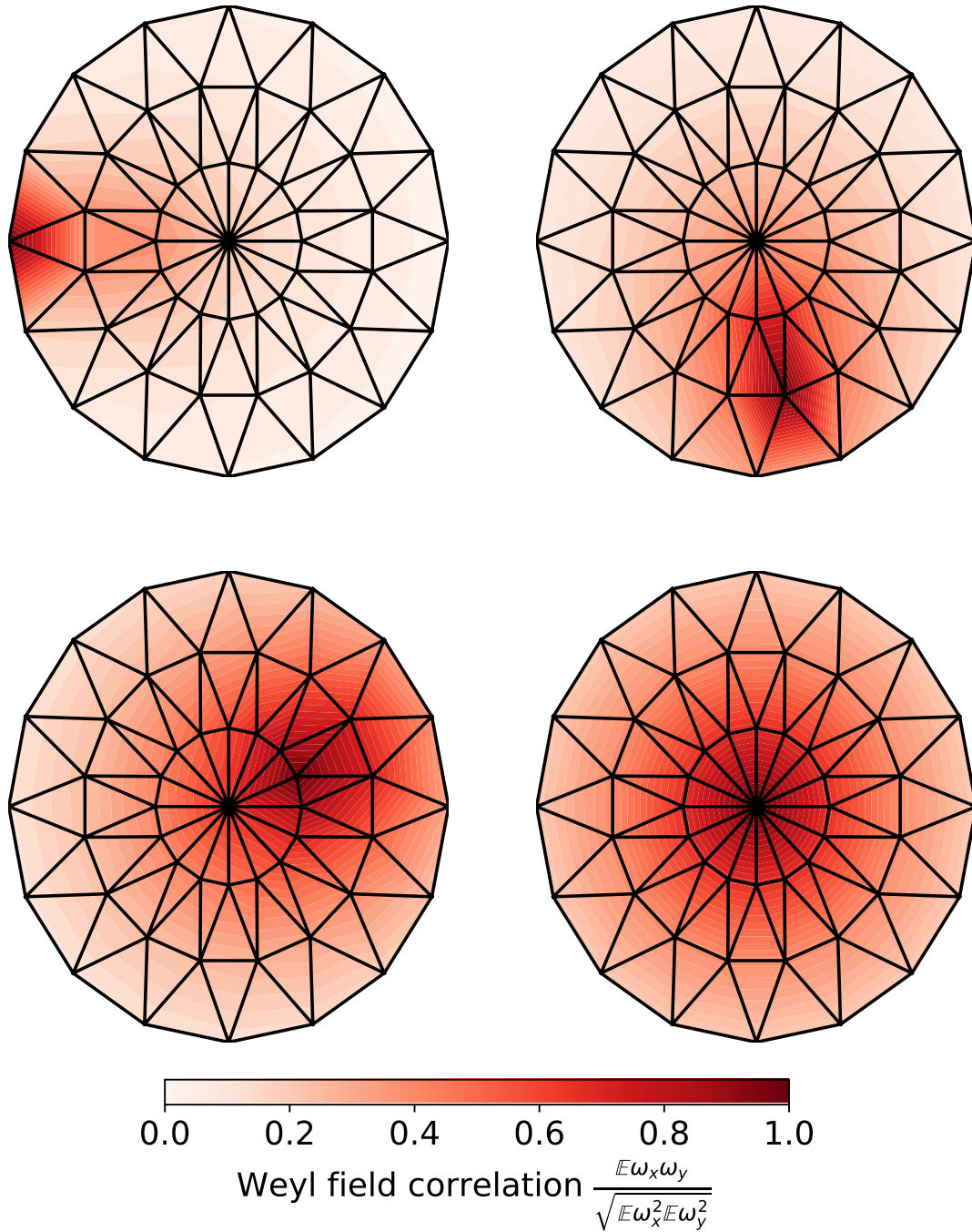
Figure 4.7: Slices of the learned Weyl field correlation matrix $C_{xy} := \frac{\mathbb{E}\omega_x\omega_y}{\sqrt{\mathbb{E}\omega_x^2\mathbb{E}\omega_y^2}}$. In this normalization, the largest possible value is the correlation between itself which is defined to be unity. In each subplot, a row of C_{xy} of a fixed x is shown.

training[2].

Furthermore, it is curious to see what elementary excitations are present in the learned distribution. By computing the spectrum decomposition of

$$C_{xy} = \sum_{\alpha} \lambda_{\alpha} \phi_{\alpha}(x) \phi_{\alpha}(y) \qquad (4.18)$$

where λ_{α} is the eigenvalue of the αth eigenmode ϕ_{α}, some familiar patterns like s-wave and p-wave emerge. For example see Figure 4.8 and 4.9.
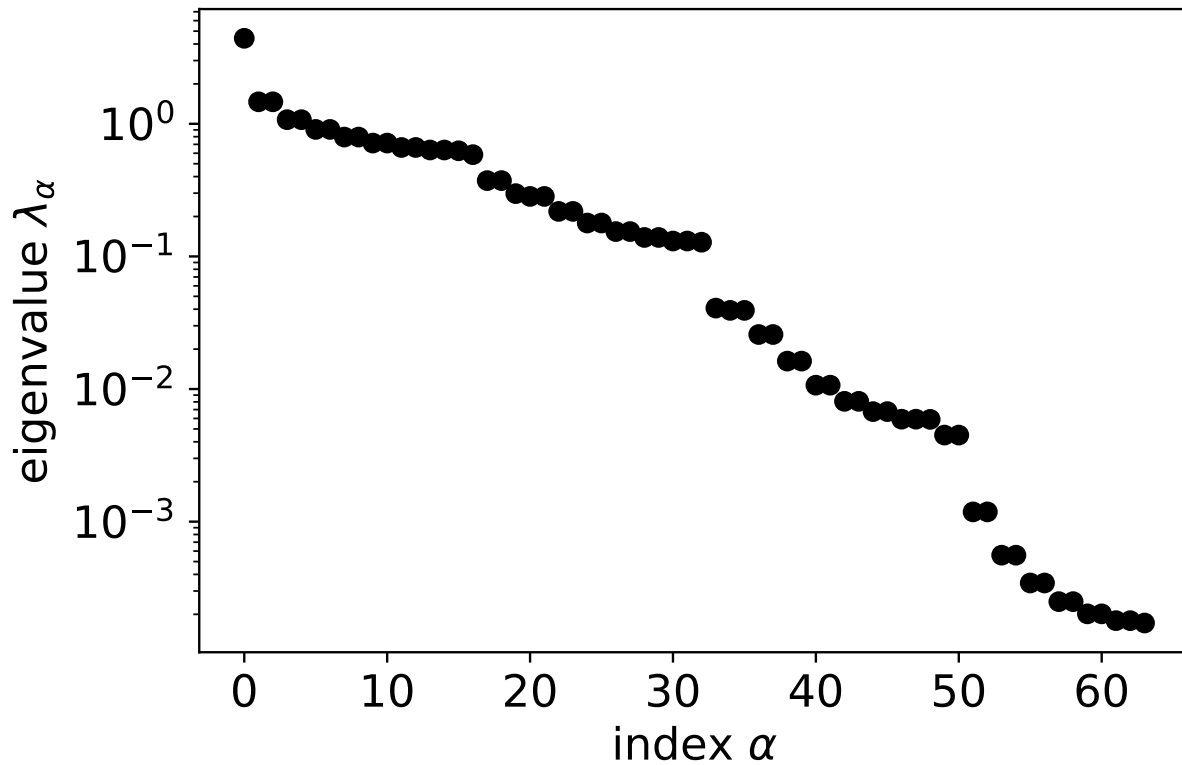
Figure 4.8: Eigenvalues λ_{α} of the learned Weyl covariance matrix C_{xy}.

---

[2]It is true that the use of L2 norm regularization encourage sparsity, but not necessarily locality which requires knowledge of bulk geometry.
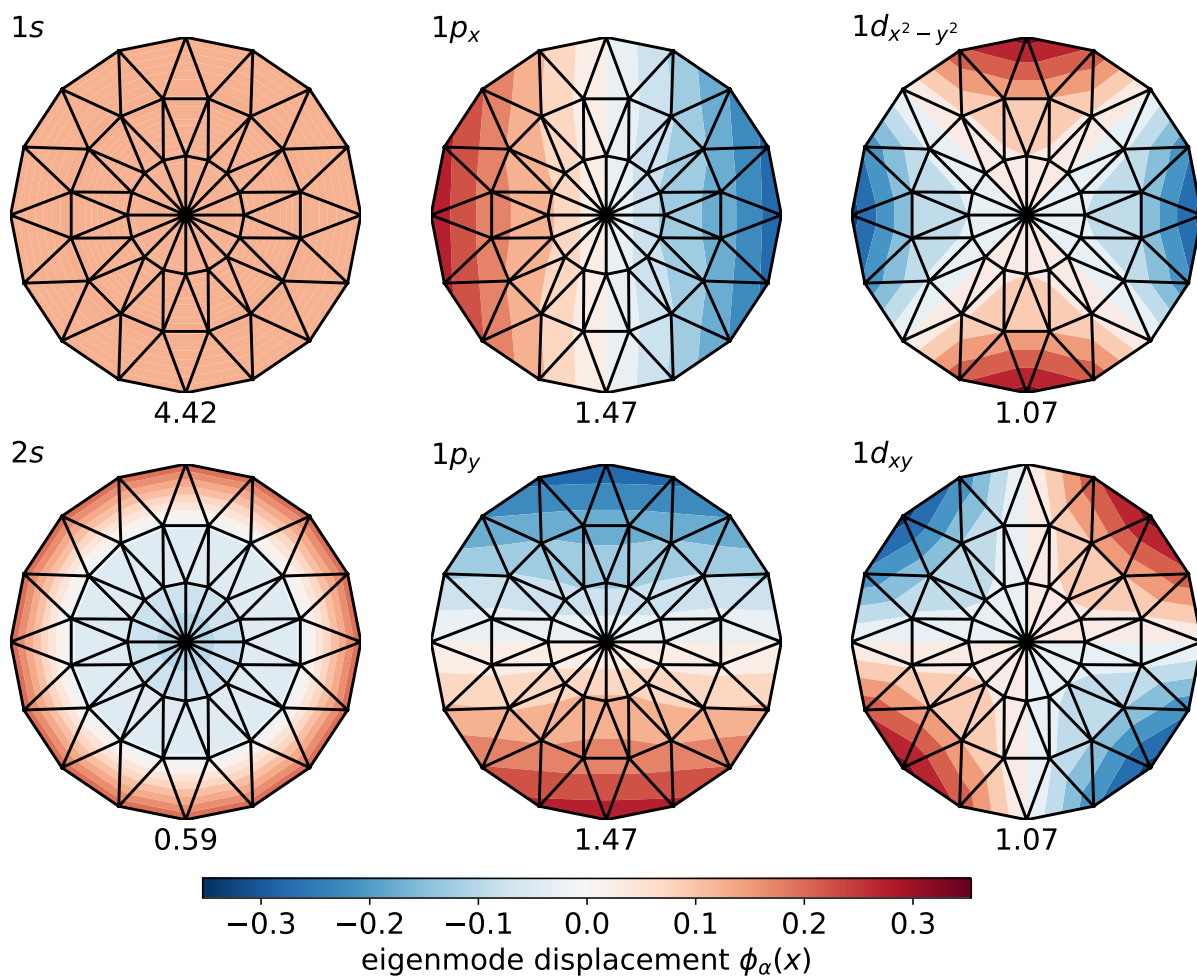
Figure 4.9: Selected eigenmodes ϕ_α and their corresponding eigenvalues, organised by their symmetry properties (angular momentum).

4.6 Mass Renormalization

Geometric fluctuation effectively introduces interaction between the bulk scalar field. As a result, the bare mass m needs to be renormalized. Remarkably, a similar phenomenon is observed in our holographic model. We trained a 'classical' model without Weyl fluctuations on single-region data, followed by a full model with fluctuating Weyl field trained on both single- and double-region data. In both cases the scalar field mass m were trainable. For the range of central charge $5.5 \leq c \leq 8$ studied, we saw the trained values of the (bare) mass in the full model are systematically larger than that of a static geometry. See Figure 4.10. This effect can
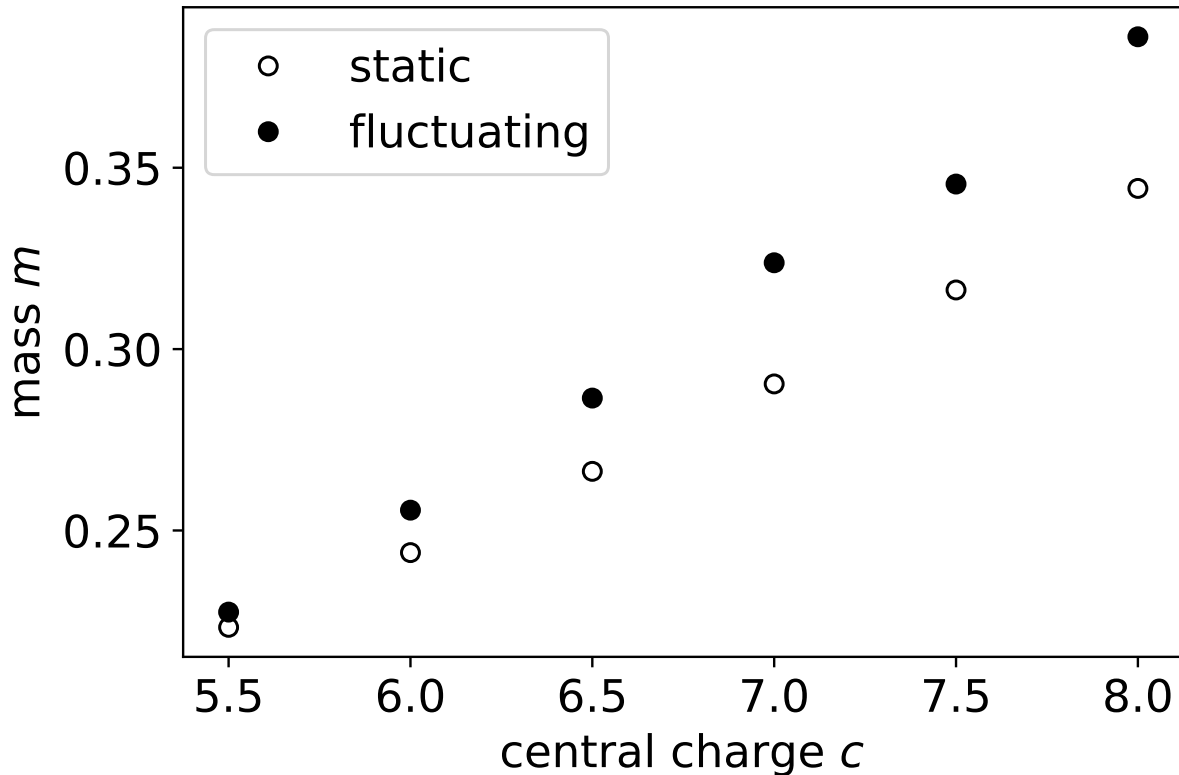
Figure 4.10: Trained values of the scalar field mass m for different central charge c. The 'classical' model with static geometry is trained on single sub-region entropy, while full model is trained on both 1 and 2 sub-regions. These correspond to column (a) and (c) in Table 4.1 respectively.

also be understood heuristically by considering the single-region purity: with a static geometry,

$e^{-S_{\text{model}}} = \langle \phi\phi \rangle \sim e^{-m_0\gamma_0}$ where γ_0 is the geodesic connecting the entanglement cut through the bulk. With Gaussian geometric fluctuation, $\gamma_\omega = \gamma_0 + \delta\gamma_\omega$, then

$$e^{-S_{\text{model}}} =\sim \mathbb{E}_\omega e^{-m\gamma_\omega} = e^{-m(\gamma_0 - \frac{m}{2}\mathbb{E}_\omega(\delta\gamma_\omega)^2)}. \tag{4.19}$$

For these two models to produce the same result, one concludes $m > m_0$, in qualitative agreement with what we observed.

4.7 More Training Details

4.7.1 Pretraining

Even with a linear Weyl field generator, the optimization problem is non-convex. It is because non-linearity entered the loss function at several stages — exponentiation of the Weyl field, matrix inversion at eq (3.19) and wick contractions such as eq (3.20). We found that it is more efficient to initialize the network by pretraining it with a simpler task. Namely instead of asking the network to predict $e^{-S(A)}$, it was first fitted to produce a target Weyl covarinace Σ_0 by minimizing the MSE loss

$$\mathcal{L}_{\text{pretrain}} = |\Sigma_{\text{model}} - \Sigma_0|^2. \tag{4.20}$$

where $\Sigma_{\text{model}} := \mathbb{E}\omega\omega^T$ can be estimated by sample mean. However, for a linear model that consist a single convolutional layer $\omega = Wz$ applied to standard Gaussian $z \sim \mathcal{N}(0,I)$, the generated Weyl field is simply another Gaussian with covariance

$$\Sigma_{\text{model}} = WW^T. \tag{4.21}$$

A convenient choice of Σ_0 could be just the scalar field propagator

$$\Sigma_0 = C[L + \text{diag}(m^2 A_x)]^{-1} \tag{4.22}$$

for some value of m and C where L and A_x are defined in eq (3.18). With such initialization, the model is trained by the MRSE as defined in eq(3.22). The MRSE loss was found to converge much quicker. We also found that with proper regularization, the precise value of m and C used in pretraining does not affect the final result significantly. See section 4.7.2.

4.7.2   Regularization

A natural question is about the uniqueness of bulk geometric distribution parameterised by $P[\omega]$. To address this question empirically, one may repeat the training procedures with different initialization and compare the final learned distributions. For linear network, it suffices to study the learned covariance defined by eq (4.21). It was observed the learned covariances vary significantly from trials to trials, while all producing very similar RMSE loss.

Guided by the principle of Oscam's razor, we then asked: what is the *simplest* bulk geometric distribution $P[\omega]$ that can explain the boundary entropy data? Quantitatively, we propose to regularize the optimization problem by adding two terms to the loss defined in eq (3.22) so that

$$\mathcal{L}_{\text{total}} = \mathcal{L}_{\text{data}} + \varepsilon_{\text{smooth}} \mathcal{L}_{\text{smooth}} + \varepsilon_{\text{norm}} \mathcal{L}_{\text{norm}} \tag{4.23}$$

$$\mathcal{L}_{\text{smooth}} = \left\| \frac{L\Sigma_{\text{model}}}{\Sigma_{\text{model}}} \right\|_{\text{F}} \tag{4.24}$$

$$\mathcal{L}_{\text{norm}} = \|\Sigma_{\text{model}}\|_{\text{F}} \tag{4.25}$$

where L is the graph Laplacian defined in eq (3.18), $\varepsilon_{\text{smooth}}$ and $\varepsilon_{\text{smooth}}$ are some small numbers, $||\cdot||_F$ denotes the Frobenius matrix norm and element-wise division is used in eq (4.24). We

emphasized that the neither $\mathcal{L}_{smooth}$ nor $\mathcal{L}_{norm}$ explicitly encourage/discourage locality of Σ_{model}.

A standard practice in machine learning is to tune hyperparameters like ε_{smooth} and ε_{smooth} by optimizing the out-of-sample performance. Due to limited resources, we did not tune those parameters systematically. Instead we used fixed values $\varepsilon_{smooth} = 10^{-4}$ and $\varepsilon_{smooth} = 10^{-3}$. The addition of regularisation losses had negligible effect on model's performance in terms of $\mathcal{L}_{data}$. However, we found the over a number of trials ($N = 10$), with (i) random convolutional filter initialization (ii) random pretraining parameters within ranges $0.1 < m < 0.3$ and $0.1 < C < 0.3$, the converged models produce the same Σ_{model} within 10%. In this sense, this regularised inverse problem — inferring bulk from boundary seems to have a unique solution.

This chapter is currently being prepared for submission for publication of the material. Lam, Jonathan C. C.; Yi-Zhuang, You. The dissertation author was the primary investigator and author of this material.

Chapter 5

Conclusion

The RT formula has offered an explicit quantitative relation between boundary entanglement and bulk geometry assuming geometric fluctuations are negligible. One implication of such assumption is vanishing mutual information on the boundary, which is generally not true. In order to build a holographic description for a wider class of boundary quantum systems, we have incorporated the effect of bulk geometric fluctuations. More specifically, by considering a generalized RTN model with fluctuating bond dimensions, an emergent statistical gravity model emerged as an effective description. With the help of modern machine learning techniques, the nature of bulk gravitational fluctuations were uncovered from boundary entanglement data. We believe the hybrid approach that we have taken — physics informed machine learning could also benefit other fields in theoretical physics.

Bibliography

[AAB+15]   Martín Abadi, Ashish Agarwal, Paul Barham, Eugene Brevdo, Zhifeng Chen, Craig
Citro, Greg S. Corrado, Andy Davis, Jeffrey Dean, Matthieu Devin, Sanjay Ghe-
mawat, Ian Goodfellow, Andrew Harp, Geoffrey Irving, Michael Isard, Yangqing
Jia, Rafal Jozefowicz, Lukasz Kaiser, Manjunath Kudlur, Josh Levenberg, Dan
Mané, Rajat Monga, Sherry Moore, Derek Murray, Chris Olah, Mike Schuster,
Jonathon Shlens, Benoit Steiner, Ilya Sutskever, Kunal Talwar, Paul Tucker, Vin-
cent Vanhoucke, Vijay Vasudevan, Fernanda Viégas, Oriol Vinyals, Pete Warden,
Martin Wattenberg, Martin Wicke, Yuan Yu, and Xiaoqiang Zheng. TensorFlow:
Large-scale machine learning on heterogeneous systems, 2015. Software available
from tensorflow.org.

[BCC+14]   Vijay Balasubramanian, Borun D. Chowdhury, Bartłomiej Czech, Jan de Boer, and
Michal P. Heller. Bulk curves from boundary data in holography. *Phys. Rev. D*,
89:086004, Apr 2014.

[BCCd13]   Vijay Balasubramanian, Borun D. Chowdhury, Bartlomiej Czech, and Jan de Boer.
The entropy of a hole in spacetime. *Journal of High Energy Physics*, 2013:220, Oct
2013.

[BEJ+19]   Tiff Brydges, Andreas Elben, Petar Jurcevic, Benoît Vermersch, Christine Maier,
Ben P. Lanyon, Peter Zoller, Rainer Blatt, and Christian F. Roos. Probing Rényi
entanglement entropy via randomized measurements. *Science*, 364(6437):260–263,
Apr 2019.

[Bek20]   Jacob D Bekenstein. Black holes and the second law. In *JACOB BEKENSTEIN:
The Conservative Revolutionary*, pages 303–306. World Scientific, 2020.

[BH86]   J. D. Brown and Marc Henneaux. Central charges in the canonical realization of
asymptotic symmetries: an example from three-dimensional gravity. *Comm. Math.
Phys.*, 104(2):207–226, 1986.

[BHM+14]   Vijay Balasubramanian, Patrick Hayden, Alexander Maloney, Donald Marolf, and
Simon F. Ross. Multiboundary Wormholes and Holographic Entanglement. *arXiv
e-prints*, page arXiv:1406.2663, Jun 2014.

[BHRT15]   Jyotirmoy Bhattacharya, Veronika E. Hubeny, Mukund Rangamani, and Tadashi Takayanagi. Entanglement density and gravitational thermodynamics. , 91(10):106009, May 2015.

[BMM13]   Alexandre Belin, Alexander Maloney, and Shunji Matsuura. Holographic phases of Rényi entropies. *Journal of High Energy Physics*, 2013:50, Dec 2013.

[BPSW19]   Ning Bao, Geoffrey Penington, Jonathan Sorce, and Aron C. Wall. Holographic Tensor Networks in Full AdS/CFT. *arXiv e-prints*, page arXiv:1902.10157, Feb 2019.

[CC04]   Pasquale Calabrese and John Cardy. Entanglement entropy and quantum field theory. *Journal of Statistical Mechanics: Theory and Experiment*, 2004(06):P06002, 2004.

[CCM17]   ChunJun Cao, Sean M. Carroll, and Spyridon Michalakis. Space from Hilbert space: Recovering geometry from bulk entanglement. , 95(2):024031, January 2017.

[CHH12]   Robert Callan, Jianyang He, and Matthew Headrick. Strong subadditivity and the covariant holographic entanglement entropy formula. *Journal of High Energy Physics*, 2012:81, Jun 2012.

[CL14]   Bartłomiej Czech and Lampros Lamprou. Holographic definition of points and distances. *Phys. Rev. D*, 90:106005, Nov 2014.

[CW16]   Taco S. Cohen and Max Welling. Group equivariant convolutional networks. *CoRR*, abs/1602.07576, 2016.

[Cyb89]   George Cybenko. Approximation by superpositions of a sigmoidal function. *Mathematics of control, signals and systems*, 2(4):303–314, 1989.

[DDS$^+$09]   Jia Deng, Wei Dong, Richard Socher, Li-Jia Li, Kai Li, and Li Fei-Fei. Imagenet: A large-scale hierarchical image database. In *2009 IEEE conference on computer vision and pattern recognition*, pages 248–255. Ieee, 2009.

[DQSY20]   Xi Dong, Xiao-Liang Qi, Zhou Shangnan, and Zhenbin Yang. Effective entropy of quantum fields coupled with gravity. *Journal of High Energy Physics*, 2020(10):52, October 2020.

[DSB16]   Laurent Dinh, Jascha Sohl-Dickstein, and Samy Bengio. Density estimation using Real NVP. *arXiv e-prints*, page arXiv:1605.08803, May 2016.

[EV14]   G. Evenbly and G. Vidal. Class of Highly Entangled Many-Body States that can be Efficiently Simulated. *Physical Review Letters*, 112(24):240502, June 2014.

[EW15]   Netta Engelhardt and Aron C. Wall. Quantum extremal surfaces: holographic entanglement entropy beyond the classical regime. *Journal of High Energy Physics*, 2015:73, January 2015.

[FLM13]   Thomas Faulkner, Aitor Lewkowycz, and Juan Maldacena. Quantum corrections to holographic entanglement entropy. *Journal of High Energy Physics*, 2013:74, Nov 2013.

[GBC16]   Ian Goodfellow, Yoshua Bengio, and Aaron Courville. *Deep Learning*. MIT Press, 2016.

[GKP98]   S. S. Gubser, I. R. Klebanov, and A. M. Polyakov. Gauge theory correlators from non-critical string theory. *Physics Letters B*, 428:105–114, May 1998.

[GPAM⁺14] Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. Generative adversarial nets. In Z. Ghahramani, M. Welling, C. Cortes, N. Lawrence, and K. Q. Weinberger, editors, *Advances in Neural Information Processing Systems*, volume 27. Curran Associates, Inc., 2014.

[Ham08]   Herbert W Hamber. *Quantum gravitation: The Feynman path integral approach*. Springer Science & Business Media, 2008.

[Haw75]   Stephen W Hawking. Particle creation by black holes. *Communications in mathematical physics*, 43(3):199–220, 1975.

[Hea10]   Matthew Headrick. Entanglement Rényi entropies in holographic theories. *Physical Review D*, 82(12):126010, Dec 2010.

[Hea19]   Matthew Headrick. Lectures on entanglement entropy in field theory and holography. *arXiv e-prints*, page arXiv:1907.08126, Jul 2019.

[HKP20]   Hsin-Yuan Huang, Richard Kueng, and John Preskill. Predicting many properties of a quantum system from very few measurements. *Nature Physics*, 16(10):1050–1057, June 2020.

[HNQ⁺16]   Patrick Hayden, Sepehr Nezami, Xiao-Liang Qi, Nathaniel Thomas, Michael Walter, and Zhao Yang. Holographic duality from random tensor networks. *Journal of High Energy Physics*, 2016(11):9, Nov 2016.

[HRT07]   V. E. Hubeny, M. Rangamani, and T. Takayanagi. A covariant holographic entanglement entropy proposal. *Journal of High Energy Physics*, 7:062, July 2007.

[HSW89]   Kurt Hornik, Maxwell Stinchcombe, and Halbert White. Multilayer feedforward networks are universal approximators. *Neural networks*, 2(5):359–366, 1989.

[HT07]   Matthew Headrick and Tadashi Takayanagi. Holographic proof of the strong subadditivity of entanglement entropy. *Physical Review D*, 76(10):106013, Nov 2007.

[IL67]   Alekseĭ Grigorevich Ivakhnenko and Valentin Grigorevich Lapa. Cybernetics and forecasting techniques. 1967.

[JK13]   Kristan Jensen and Andreas Karch. Holographic dual of an einstein-podolsky-rosen pair has a wormhole. *Phys. Rev. Lett.*, 111:211602, Nov 2013.

[JM15]   Danilo Jimenez Rezende and Shakir Mohamed. Variational Inference with Normalizing Flows. *arXiv e-prints*, page arXiv:1505.05770, May 2015.

[KB14]   D. P. Kingma and J. Ba. Adam: A Method for Stochastic Optimization. *ArXiv e-prints*, December 2014.

[KSJ$^+$16] Diederik P. Kingma, Tim Salimans, Rafal Jozefowicz, Xi Chen, Ilya Sutskever, and Max Welling. Improving Variational Inference with Inverse Autoregressive Flow. *arXiv e-prints*, page arXiv:1606.04934, June 2016.

[KW13]   Diederik P Kingma and Max Welling. Auto-encoding variational bayes. *arXiv preprint arXiv:1312.6114*, 2013.

[LRSV15]   Nima Lashkari, Charles Rabideau, Philippe Sabella-Garnier, and Mark Van Raamsdonk. Inviolable energy conditions from entanglement inequalities. *Journal of High Energy Physics*, 2015:67, Jun 2015.

[Mal99]   J. Maldacena. The Large-N Limit of Superconformal Field Theories and Supergravity. *International Journal of Theoretical Physics*, 38:1113–1133, 1999.

[MS13]   J. Maldacena and L. Susskind. Cool horizons for entangled black holes. *Fortschritte der Physik*, 61(9):781–811, Sep 2013.

[PE09]   Ingo Peschel and Viktor Eisler. Reduced density matrices and entanglement entropy in free lattice models. *Journal of physics a: mathematical and theoretical*, 42(50):504003, 2009.

[PGM$^+$19] Adam Paszke, Sam Gross, Francisco Massa, Adam Lerer, James Bradbury, Gregory Chanan, Trevor Killeen, Zeming Lin, Natalia Gimelshein, Luca Antiga, Alban Desmaison, Andreas Kopf, Edward Yang, Zachary DeVito, Martin Raison, Alykhan Tejani, Sasank Chilamkurthy, Benoit Steiner, Lu Fang, Junjie Bai, and Soumith Chintala. Pytorch: An imperative style, high-performance deep learning library. In H. Wallach, H. Larochelle, A. Beygelzimer, F. d'Alché-Buc, E. Fox, and R. Garnett, editors, *Advances in Neural Information Processing Systems 32*, pages 8024–8035. Curran Associates, Inc., 2019.

[PPM17]   George Papamakarios, Theo Pavlakou, and Iain Murray. Masked Autoregressive Flow for Density Estimation. *arXiv e-prints*, page arXiv:1705.07057, May 2017.

[Qi13]   X.-L. Qi. Exact holographic mapping and emergent space-time geometry. *ArXiv e-prints*, September 2013.

[QYY17]   Xiao-Liang Qi, Zhao Yang, and Yi-Zhuang You. Holographic coherent states from random tensor networks. *Journal of High Energy Physics*, 8:60, August 2017.

55

[Reg61]   T. Regge. General relativity without coordinates. *Nuovo Cim.*, 19:558–571, 1961.

[RHW86]   David E Rumelhart, Geoffrey E Hinton, and Ronald J Williams. Learning representations by back-propagating errors. *nature*, 323(6088):533–536, 1986.

[RT06a]   Shinsei Ryu and Tadashi Takayanagi. Aspects of holographic entanglement entropy. *Journal of High Energy Physics*, 2006(8):045, August 2006.

[RT06b]   Shinsei Ryu and Tadashi Takayanagi. Holographic Derivation of Entanglement Entropy from the anti de Sitter Space/Conformal Field Theory Correspondence. , 96(18):181602, May 2006.

[Sch15]   Jürgen Schmidhuber. Deep learning in neural networks: An overview. *Neural networks*, 61:85–117, 2015.

[Sus14]   Leonard Susskind. ER=EPR, GHZ, and the Consistency of Quantum Measurements. *arXiv e-prints*, page arXiv:1412.8483, Dec 2014.

[Swi12a]   B. Swingle. Constructing holographic spacetimes using entanglement renormalization. *ArXiv e-prints*, September 2012.

[Swi12b]   Brian Swingle. Entanglement renormalization and holography. *Physical Review D*, 86(6):065007, 2012.

[Van09]   Mark Van Raamsdonk. Comments on quantum gravity and entanglement. *arXiv e-prints*, page arXiv:0907.2939, July 2009.

[van10]   Mark van Raamsdonk. Building up spacetime with quantum entanglement. *General Relativity and Gravitation*, 42(10):2323–2329, Oct 2010.

[VOKK16]   Aaron Van Oord, Nal Kalchbrenner, and Koray Kavukcuoglu. Pixel recurrent neural networks. In *International Conference on Machine Learning*, pages 1747–1756. PMLR, 2016.

[VPYL18]   Romain Vasseur, Andrew C. Potter, Yi-Zhuang You, and Andreas W. W. Ludwig. Entanglement Transitions from Holographic Random Tensor Networks. *arXiv e-prints*, page arXiv:1807.07082, Jul 2018.

[Wit98a]   E. Witten. Anti-de Sitter Space, Thermal Phase Transition, And Confinement In Gauge Theories. *Adv. Theor. Math. Phys.*, 2:505–532, March 1998.

[Wit98b]   Edward Witten. Anti-de Sitter space and holography. *Advances in Theoretical and Mathematical Physics*, 2:253–291, Jan 1998.

[YYQ18]   Yi-Zhuang You, Zhao Yang, and Xiao-Liang Qi. Machine learning spatial geometry from entanglement features. *Phys. Rev. B*, 97:045153, Jan 2018.