Title: Tensor Network Holography and Deep Learning

Date: Nov 20, 2017 11:00 AM

URL: http://pirsa.org/17110111

Abstract: <p>Motivated by the close relations of the renormalization group with both the holography duality and the deep learning, we propose that the holographic geometry can emerge from deep learning the entanglement feature of a quantum many-body state. We develop a concrete algorithm, call the entanglement feature learning (EFL), based on the random tensor network (RTN) model for the tensor network holography. We show that each RTN can be mapped to a Boltzmann machine, trained by the entanglement entropies over all subregions of a given quantum many-body state. The goal is to construct the optimal RTN that best reproduce the entanglement feature. The RTN geometry can then be interpreted as the emergent holographic geometry. We demonstrate the EFL algorithm on 1D free fermion system and observe the emergence of the hyperbolic geometry (AdS_3 spatial geometry) as we tune the fermion system towards the gapless critical point (CFT_2 point).</p>

Tensor Network Holography and Deep Learning

Yi-Zhuang You
Harvard University

Y.-Z. You, Z. Yang, X.-L. Qi, arXiv: 1709.01223

Perimeter Institute, Nov. 2017

Outline

- Review of two kinds of networks:
  - **Tensor Network - Holographic Duality** (AdS/CFT)
  - **Neural Network - Deep Learning**
- Is there any connection between tensor network and neural network? What about holography and deep learning?
- **Entanglement Feature Learning (EFL)**
  - Holographic spacial geometry emerges from deep learning the entanglement features in a quantum many-body state.
  - Demonstration on 1D free fermion CFT.

Tensor Network

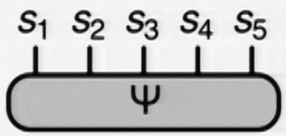
- Efficient representation of quantum many-body state
  - Quantum states represented by wave functions:

$$|\Psi\rangle = \sum_{[s_i]} \Psi(s_1, s_2, \ldots)\, |s_1\ s_2\ \ldots\rangle$$

← basis states of Hilbert space

↖ wave function ↖ for qubits: $s_i = 0, 1$

  - Wave function is also a tensor:



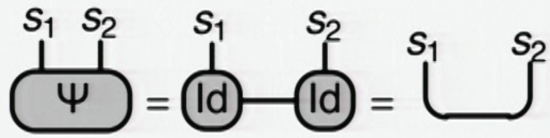
$= \Psi(s_1, s_2, \ldots) \in \mathbb{C}$ 2^N entries!

  - How can we store/represent these data efficiently?
    - Tensor networks are more structural and efficient way to represent a quantum many-body wave function than a single tensor.

Quantum Entanglement

- Example: Einstein-Podolsky-Rosen (EPR) pair

$$|\text{EPR}\rangle = \frac{1}{\sqrt{2}}(|00\rangle + |11\rangle)$$



- Quantify entanglement by Entanglement Entropy
  - Reduced Density Matrix $\rho_1 = \text{Tr}_{s_2} |\text{EPR}\rangle \langle \text{EPR}|$
  - (Renyi) Entanglement Entropy

$$S_E^{(n)} = \frac{1}{1-n} \ln \text{Tr}_{s_1} \rho_1^n \leftarrow \text{Renyi index } n$$

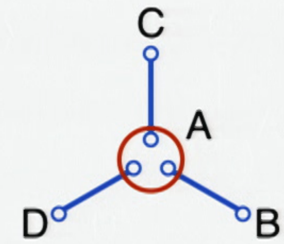
  - For EPR pair, entanglement entropy $S_E^{(n)} = \ln 2 = 1$ bit
  - Mutual Information

$$I^{(n)}(1,\, 2) = S_E^{(n)}(1) + S_E^{(n)}(2) - S_E^{(n)}(1 \cup 2) = 2 \text{ bit}$$

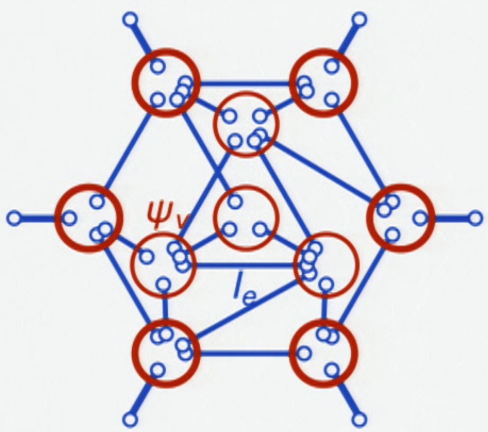
  - Measures the 2×log of the *effective* bond dimension (rank).

Tensor Network

- Building entangled many-body state from EPR pairs
  - Starting from $|\mathrm{EPR_{AB}}\rangle \otimes |\mathrm{EPR_{AC}}\rangle \otimes |\mathrm{EPR_{AD}}\rangle$
  - A measures the qubits $\rho_A \rightarrow |\psi_A\rangle\langle\psi_A|$
  - Now B, C, D are entangled



- Distribute EPR pairs + measurement → Entanglement formed among unmeasured qubits



- Tensor Network State (PEPS)
  - Given a graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$
  - Pure states $|\psi_v\rangle$ $(\forall\, v \in \mathcal{V})$
  - Entangled pairs $|I_e\rangle$ $(\forall\, e \in \mathcal{E})$

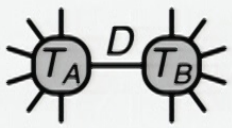
$$|\Psi_{\mathrm{TN}}\rangle = \bigotimes_{v,e \in \mathcal{G}} \langle \psi_v | I_e \rangle$$

Verstraete, Cirac, Murg (2009)

Tensor Network Holography

- Why are tensor networks interesting?
  - Useful in condensed matter physics: DMRG, MERA, PEPS
  - Entanglement structure visualized as Network geometry
    - Each link carries some amount of entanglement

      $\mathrm{rank}(\rho_A) \leq D$ or $S_E \leq \ln D$



    - Upper-bound of entropy in tensor network

      $S_E(A) \leq |\gamma_A| \ln D$ ← bond dimension

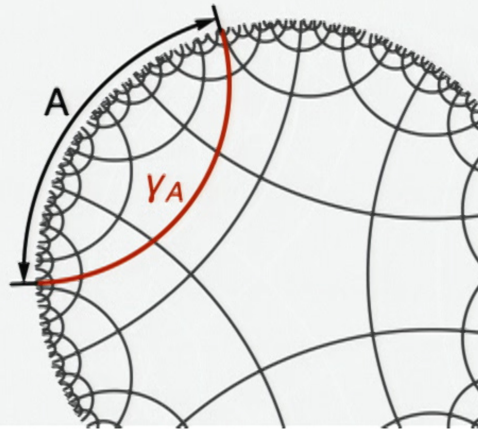
      ↖ minimal area (minimal cut)

- Holographic Duality (AdS/CFT)
  - Entanglement = Area

    $$S_E(A) = \frac{1}{4\,G_N} |\gamma_A|$$ Ryu, Takayanagi (2006)

  - Microscopic: network tiling the space

    Swingle (2009)



Tensor Network Holography

- If we want to be more serious with the idea of tensor network holography, the tensor network would better saturate the Ryu-Takayanaki (RT) bound:

$$S_E(A) = |\gamma_A| \ln D$$

- So what kind of tensor network can saturate the RT bound?
  - Early breakthrough: Perfect Tensor Network
    - holographic quantum error correction codes

    Pastawski, Yoshida, Harlow, Perskill (2015)

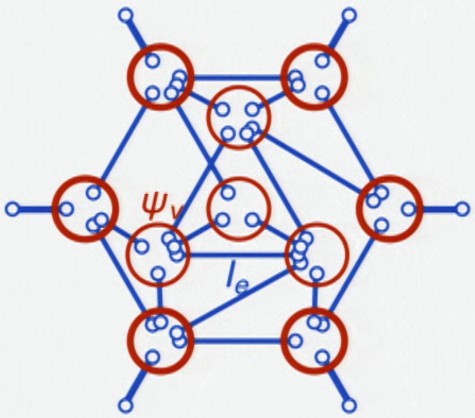
  - It turns out that the Random Tensor Network automatically satisfies the perfect conditions in the large bond dimension limit, and therefore saturates the RT bound.

$$S_E(A) \overset{D\to\infty}{\to} |\gamma_A| \ln D$$

Hayden, Nezami, Qi, Thomas, Walter, Yang (2016)

Random Tensor Network

- The Ryu-Takayanaki bound can be saturated $S_E(A) \overset{D\to\infty}{\to} |\gamma_A| \ln D$ if we consider Random Tensor Network.



- Each tensor is drawn randomly
  (Each site projects to random state)

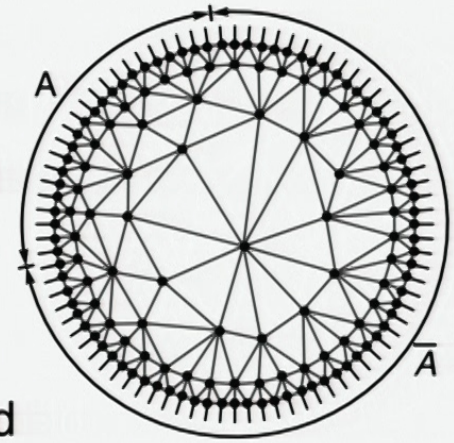
$$|\Psi_{\mathrm{RTN}}\rangle = \bigotimes_{v,e\in\mathcal{G}} \langle \psi_v \, | I_e \rangle \quad (|\psi_v\rangle \text{ random})$$

- Entanglement entropy on RTN states can be mapped to Ising models.

  Hayden, Nezami, Qi, Thomas, Walter, Yang (2016)

- Why Ising model? If entanglement = minimal cut …
  - Finding the minimal cut is a graph partition problem
  - Network science: minimize network modularity
    Physics: minimize Ising domain wall energy

Ising Model and Minimal Cut



- Given a graph with boundary
  - Introduce Ising spins $\sigma_v = \pm 1 \ (\forall\, v \in \mathcal{V})$
  - Consider a ferromagnetic Ising model

$$E = -J \sum_{e \in \mathcal{E}} \prod_{v \in \partial e} \sigma_v - h \sum_{v \in \mathcal{V}_\partial} \tau_v \, \sigma_v$$

  - Set up *opposite* boundary pinning field in regions A and Ā respectively
  - In the large J limit, Ising domain wall through the bulk automatically finds the minimal cut.
  - Cut length $|\gamma_A|$ ~ energy cost $E_A - E_0 = 2\, J\, |\gamma_A|$
- Compare to the entanglement entropy $S_E(A) = |\gamma_A| \ln D$

Conjecture: $J = \frac{1}{2} \ln D$?

Random Tensor Network

- Entanglement Entropy of Random Tensor Network
  - Reduced density matrix $\rho_A = \mathrm{Tr}_{\overline{A}} \, |\Psi_{\mathrm{RTN}}\rangle \langle \Psi_{\mathrm{RTN}}|$
  - Ensemble average over random tensors:

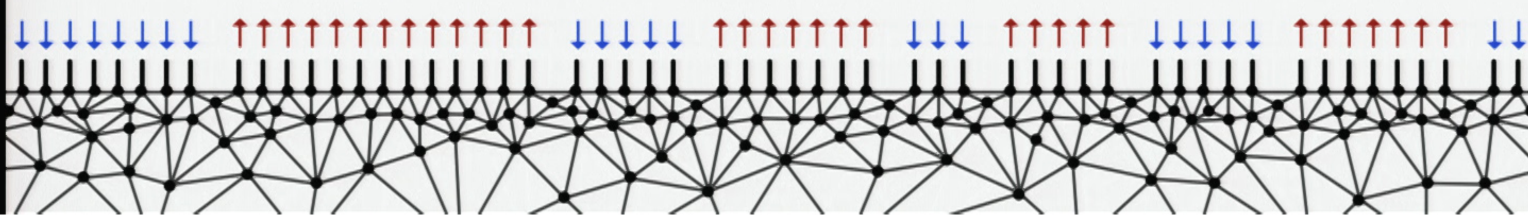
$$\mathbb{E}\,\mathrm{Tr}\,\rho_A^2 = \sum_{[\sigma]} e^{-E[\sigma,\tau]} = e^{-F[\tau]}$$

Hayden, Nezami, Qi, Thomas, Walter, Yang (2016)

  - Entanglement regions specified by the boundary field configuration $\tau_v = \begin{cases} +1 & v \in A \\ -1 & v \in \overline{A} \end{cases}$
  - Entanglement entropy = Free energy

$$S_E^{(2)}(A) = F[\tau] - F[\tau = +1]$$

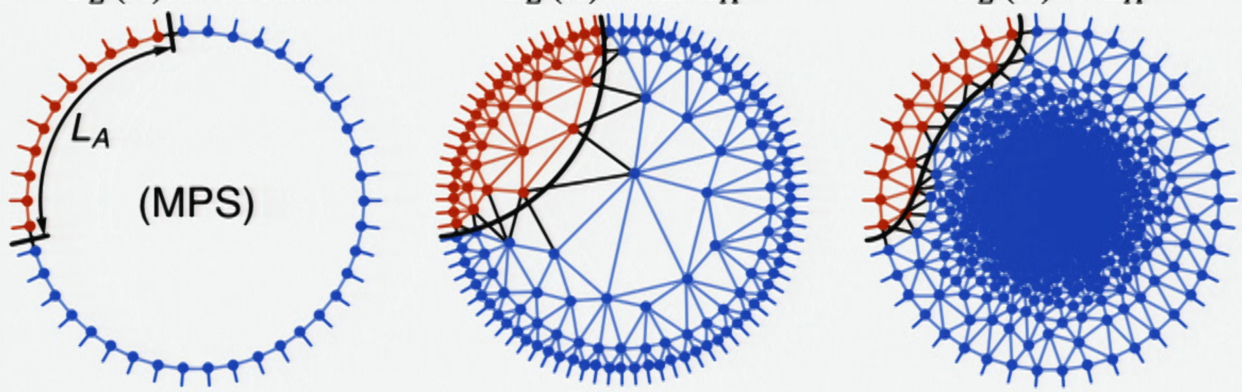
  - Coupling ∝ bound mutual information $J_e = I_e / 4 \;\; (\forall\, e \in \mathcal{E})$



Entanglement Scaling

- For a single (connected) region, how does entanglement scales with the region length?

Area-law	Logarithmic	Volume-law
$S_E(A) \sim \text{cosnt.}$	$S_E(A) \sim \ln L_A$	$S_E(A) \sim L_A$



- In principle, all graph geometries can be realized on a *complete graph* by changing $J_e = I_e / 4$ (the edge effectively disconnected when $J_e \to 0$)

Tensor Network Holography

- Holographic Duality
  - Entanglement structure ↔ Holographic geometry
  - Given bulk geometry → tensor network tiling → tensor network sate produces entanglement features
  - The inverse problem: given entanglement features, can we determine the optimal holographic geometry?
- Random Tensor Network Holography
  - J_e are the only tunable parameters: focus on network connectivity (graph geometry), ignore tensor contents
  - Quantum geometry
    - J_e label a set of coherent state basis $|\Psi_{\mathrm{RTN}}[J_e]\rangle$
    - Bidirectional isometry, quasi-orthogonality Qi, Yang, You (2017)
  - Classical geometry: treat as J_e classical parameters

A Machine Learning Problem?

- So we need to tune the set of couplings J_e in an Ising model in search of an optimal network connectivity …
- This sounds so familiar as optimizing a neural network → a Machine Learning problem?
  - **Supervised Learning**
    - Learn from labeled training data: (input x, label y) pairs
    - To establish a map from input to label $\hat{y}(x)$
    - Minimize the loss $\mathcal{L}[y, \hat{y}(x)]$
  - **Unsupervised Learning**
    - Learn from unlabeled data with empirical $P_{\mathrm{dat}}(x)$
    - To train model that generates data with $P_{\mathrm{mdl}}(x)$
    - Minimize the cross entropy $\mathcal{L} = -\sum_x P_{\mathrm{dat}}(x) \ln P_{\mathrm{mdl}}(x)$

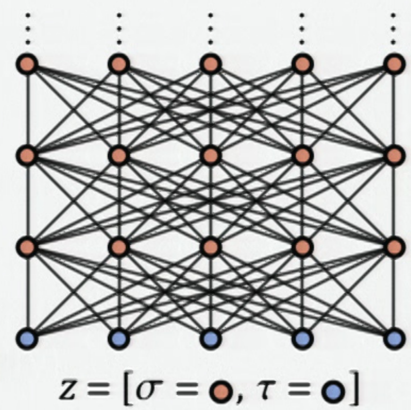
Deep Boltzmann Machine

- A deep neural network, each neuron is an Ising spin
- Visible units τ: boundary of the network
  Hidden units σ: bulk of the network
- Generative model

$$P[\tau] = \frac{1}{Z}\sum_{[\sigma]} e^{-E[\sigma,\tau]}$$ trace out hidden neurons

$$E[z] = -\sum_{e\in\mathcal{E}} J_e \prod_{v\in\partial e} z_v$$ (neglect bias terms, assume Z2 symmetry)

Salakhutdinov, Hinton (2009)



- Goal: tune network weights J_e, to optimize objective function
  - For example, the cross entropy $\mathcal{L} = -\sum_{[\tau]} P_{\text{dat}}[\tau] \ln P[\tau]$
  - Or more general, some function(al) of the free energy

$$\min_{[J]} \mathcal{L}[F[\tau; J]]$$

Entanglement Feature Learning

- Each sample in the training set is a partition of the quantum many-body system $\partial\mathcal{V} = A \cup \bar{A}$, each partition is labeled by a boundary pinning field configuration, or equivalent a set of visible spins.

$$[\tau]: \ \tau_v = \begin{cases} +1 & v \in A \\ -1 & v \in \bar{A} \end{cases}$$

Ā A
↑↑↑↓↓↓↓↑↑↑↑↓↓↑↑...

- Given a quantum many-body state $|\Psi\rangle$ (to learn), each partition [τ] is associated with an entanglement entropy

$$S^{(2)}_{E,\Psi}[\tau] \equiv S^{(2)}_{E,\Psi}(A) = -\ln \mathrm{Tr}_A \left(\mathrm{Tr}_{\bar{A}} |\Psi\rangle\langle\Psi|\right)^2$$

- **Goal**: fit the entanglement feature by that of a random tensor network (RTN)

Data $\left([\tau],\ S^{(2)}_{E,\Psi}[\tau]\right)$ **Model** $S^{(2)}_{E,\mathrm{RTN}}[\tau; J]$ **Parameters** J_e

Training Scheme (Supervised)

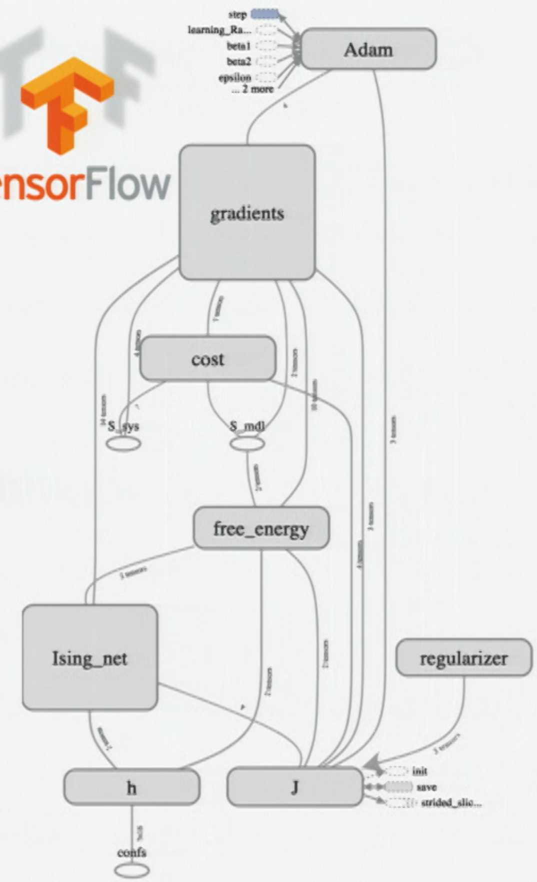
- From free fermion chain, prepare training set $([\tau], S_E[\tau])$
- From planar Ising model, predict the label $([\tau], F[\tau] - F[\tau_0])$
- Objective: mean square error

$$\mathcal{L} = \underset{[\tau]}{\text{avg}} \left(S_E[\tau] - (F[\tau] - F[\tau_0])\right)^2$$

- Optimizer: gradient descent (Adam)
- Regularization
  - Positivity $J_z > 0 \ (z = 1, 2, \dots)$
  - Coupling decay with depth

$$J_1 \geq J_2 \geq J_3 \geq \dots$$

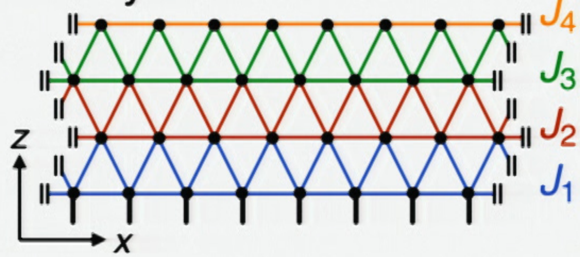
- Implementation: TensorFlow™



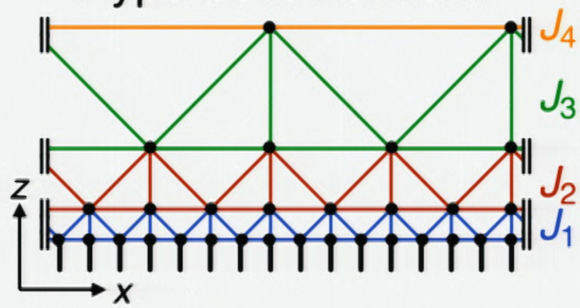
Graph Gauge Fixing

- Architecture designs
  - Layered structure (learning signal passed down from one layer to the next)
  - Uniform coupling within each layer (preserving translational symmetry as much as possible)
  - Periodic boundary condition
- Background geometry: cylindrical (flat) v.s. hyperbolic
- Boundary pinning field set by the physical bond dimension (given by the training set) $h = \frac{1}{2} \ln D$

Cylindrical network



Hyperbolic network

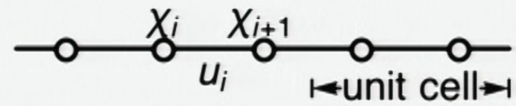


Training Set: 1D Free Fermion Chain

- Computing entanglement entropies for generic many-body system is hard.
- To test the idea of EFL, we start with the simplest free fermion chain

Majorana fermions

$$H = -\sum_{a=1}^{N}\sum_{i=1}^{L} i\, u_i\, \chi_{a,i}\, \chi_{a,i+1} \qquad u_i = 1 + m\,(-1)^i$$

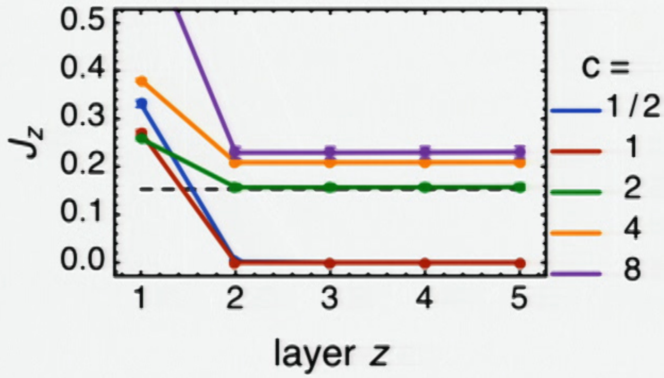


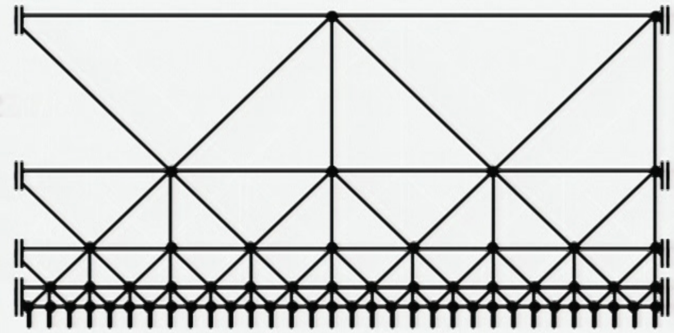
- Two key parameters:
  - Central charge $c = N/2$, counts fermion flavor number
  - Mass m: controls correlation length $\xi = 1/m$
- Entanglement entropy over any (multiple) regions can be calculated efficiently from the fermion Green's function

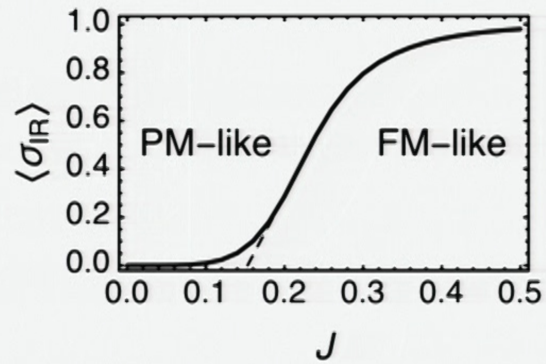
$$G_A = \frac{1}{2}\langle P_A\, \chi\, \chi^{\mathrm{T}}\, P_A\rangle \qquad S_E^{(2)}(A) = -\mathrm{Tr}\ln\left(G_A^2 + (1-G_A)^2\right)$$

Choosing the Central Charge

- Fix mass $m = 0$ (critical fermion chain), hyperbolic network



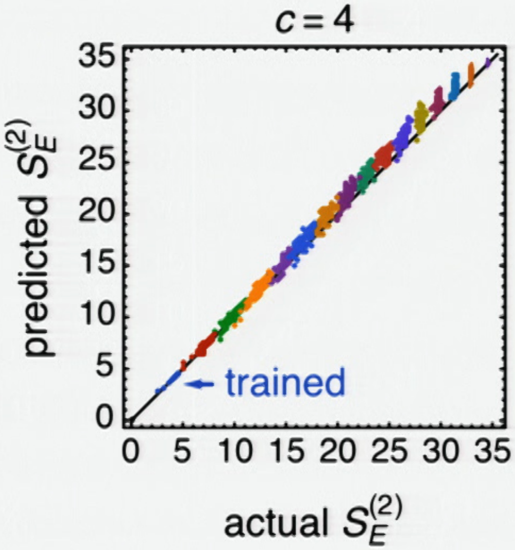




- Ising coupling J_z develops layer by layer during training
- J_z varies with central charge → a FM-PM crossover in the bulk
- AdS/CFT: $c = 3\,\ell/(2\,G_N)$
  small c = strong-coupling (quantum) gravity/geometry

Multi-Region Entanglement Entropy

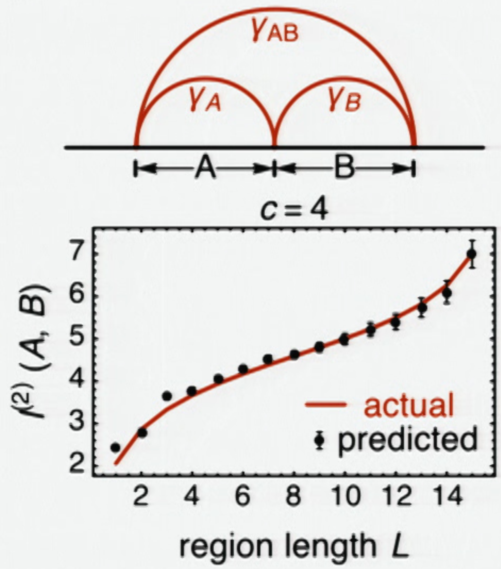
- Train a hyperbolic network using single-region entanglement entropies → Can the network predict multi-region entanglement entropies?
  - Different color - different number of entanglement regions
  - In the training phase, only single-region data is served
  - The trained machine was able to predict multi-region entanglement entropies (which was not in the training set) with accuracy ~ 95%



  - Not too surprising, as $S_E(A \cup B) = S_E(A) + S_E(B) - I(A, B)$
    The additive part is relatively easy to capture.
    What about the sub-additive (mutual information) part?

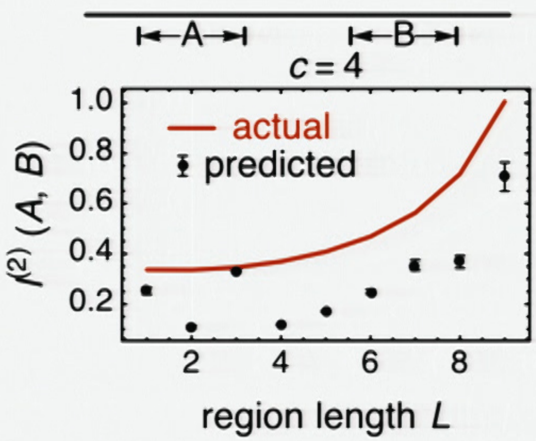
Mutual Information

- Adjacent regions: mutual information still fits well, because there is still a geometric interpretation



- Deep network at work!

- Separated regions: can not provide sufficient mutual information, need additional matter on top of the background geometry…



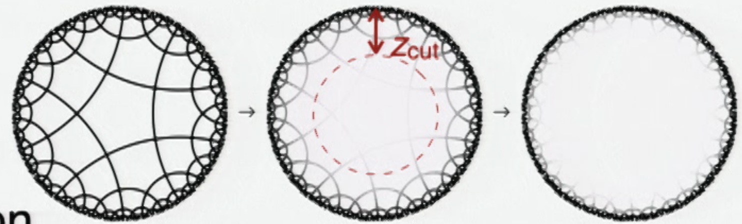
- Learnt the classical bulk geometry, nothing more

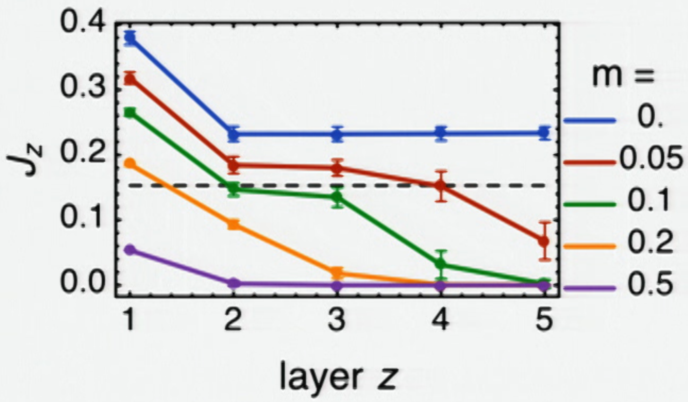
How Deep Could It Be?

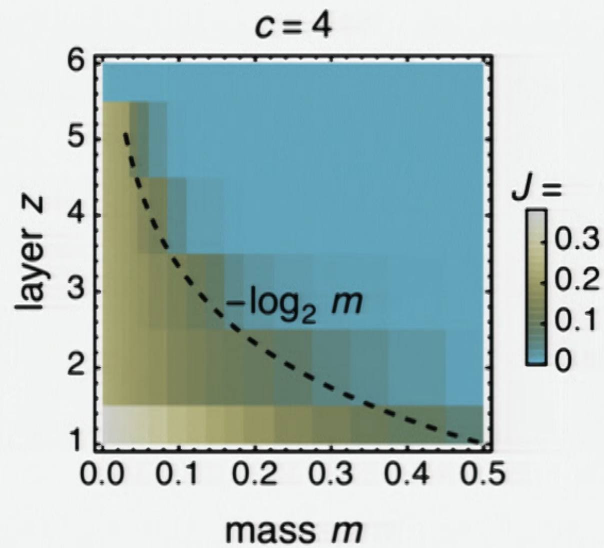
- Turn on the fermion mass *m*, the holographic bulk is cut off at the mass scale (in IR)

  $$z_{\text{cut}} \sim \ln \xi = -\ln m$$

  - Deep layers fade away
  - Quantum phase transition of the Majorana chain is signified by the peak of z_{cut}







Future Applications

- Modeling Thermalization Transition
  - Many-body localization (MBL): area law state
  - Eigenstate thermalization (ETH): volume law state
  - MBL-ETH transition: a transition of entanglement features, a transition of holographic bulk geometry
- Understanding Mandy-Body Entanglement
  - Unfolding the structure of entanglement at multiple scales
  - Classify phases of matter based on entanglement structure
  - Dynamics of entanglement formation/propagation

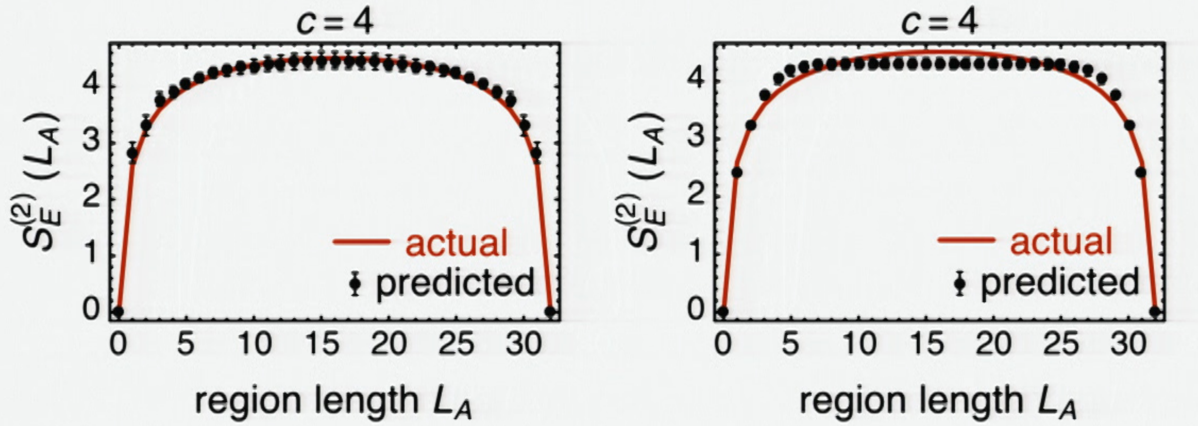
Thanks for your attention!

Single-Region Entanglement Entropy

- For critical fermion chain ($m = 0$), the single-region entanglement entropy follows the logarithmic scaling

$$S_E^{(n)}(A) = \frac{c}{6}\left(1 + \frac{1}{n}\right) \ln L_A$$ Calabrese, Cardy (2004)

Trained on hyperbolic net   Trained on cylindrical net



- For CFT states, hyperbolic network provides a better fit of the logarithmic scaling of the entanglement entropy.