Tensor network
vs
Machine learning

Song Cheng （程嵩）

IOP, CAS

physichengsong@iphy.ac.cn

Outline

- Tensor network in a nutshell
- TN concepts in machine learning
- TN methods in machine learning

Outline

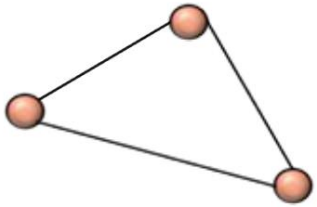
- Tensor network in a nutshell
- TN concepts in machine learning
- TN methods in machine learning

vector $\vec{v}$

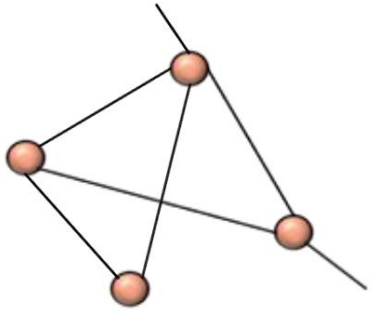
matrix A

matrix product AB

trace of matrix product $\mathrm{tr}(ABC)$



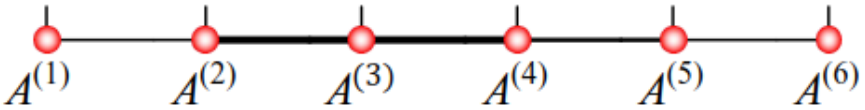
tensor contraction $f(A,B,C,D)$

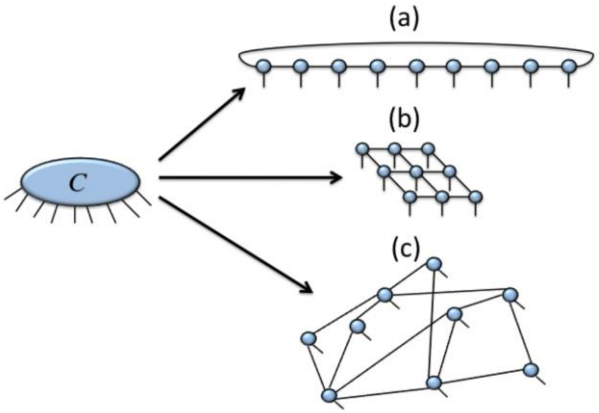


Orus, R. (2014). A Practical Introduction to Tensor Networks: Matrix Product States and Projected Entangled Pair States. *Annals of Physics*, *349*, 117–158. https://doi.org/10.1016/j.aop.2014.06.013

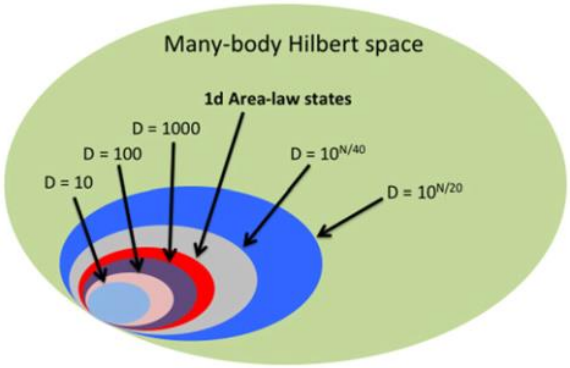
$$\Psi(v) = \sum_{ijlk\ldots\ldots} c_{ijlk\ldots\ldots}\, \varphi_i(v_1) \otimes \varphi_j(v_2) \otimes \varphi_k(v_3) \otimes \varphi_l(v_4) \otimes \ldots\ldots$$

$$\Psi_{\mathrm{MPS}}(v) = \mathrm{Tr} \prod_i A^{(i)}[v_i],$$





Tensors are local building blocks for the quantum state(like LEGO)



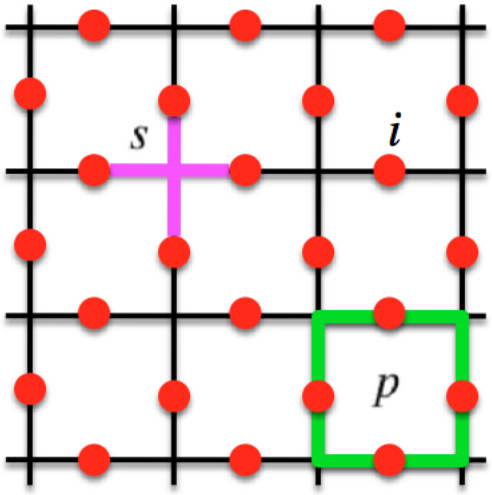
Locality lead to
low rank approximation

Orus, R. (2014). A Practical Introduction to Tensor Networks: Matrix Product States and Projected Entangled Pair States. *Annals of Physics*, *349*, 117–158. https://doi.org/10.1016/j.aop.2014.06.013

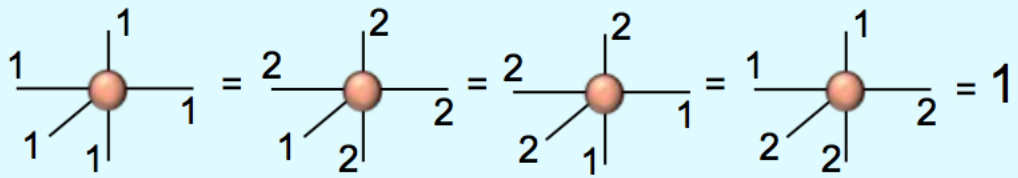
$$H = -J\sum_s A_s - J\sum_p B_p$$

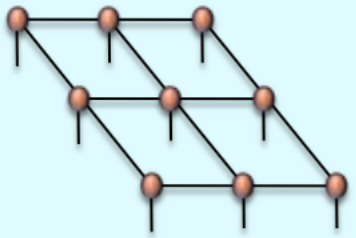
$A_s = \prod_{i\in s}\sigma_i^x$ *star operator*

$B_p = \prod_{i\in p}\sigma_i^z$ *plaquette operator*



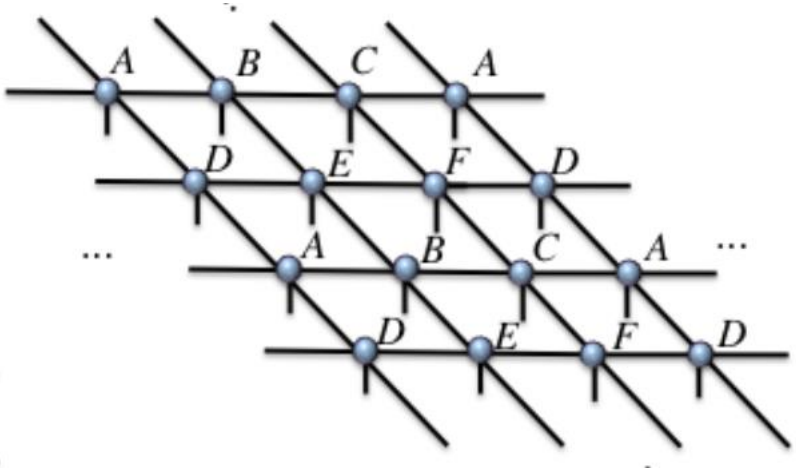
Simplest known model with “topological order”



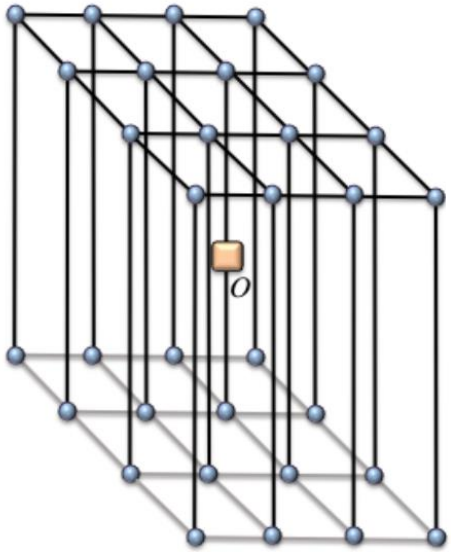


And another tensor rotated 90˚

1, Find right TN representation



2, Contract TN to calculate physics quantity



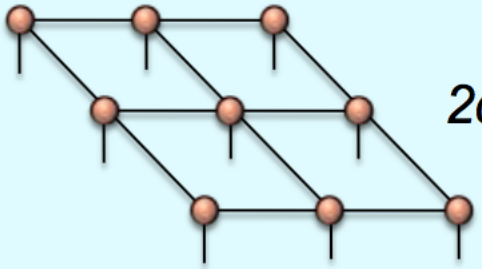
Matrix Product States (MPS)

1d systems

DMRG, Power Wave Function Renormalization Group, Time Evolving Block Decimation…

c.f., Miles Stoudenmire & Ulrich Schollwöck's talk

Projected Entangled Pair States (PEPS), Tensor Product States (TPS)

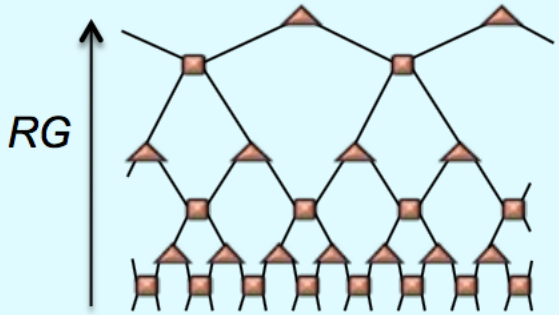


2d,3d… systems

Tensor Product Variational Approach, PEPS, Infinite-PEPS algorithm, Tensor-Entanglement Renormalization, TRG/SRG/HOTRG/HOSRG…

Multiscale Entanglement Renormalization Ansatz (MERA)

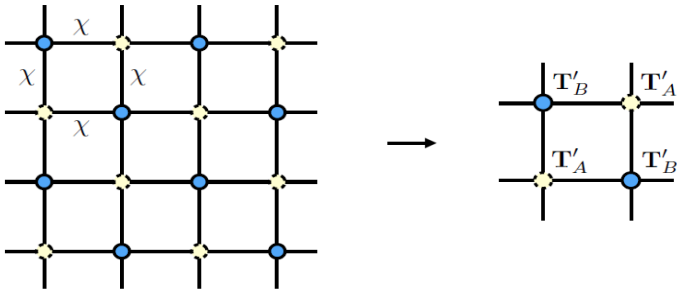
Entanglement Renormalization



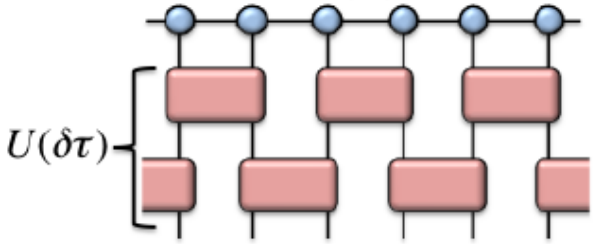
any-d scale invariant systems

Other: MPO, MPDO, TTN, PEPO…

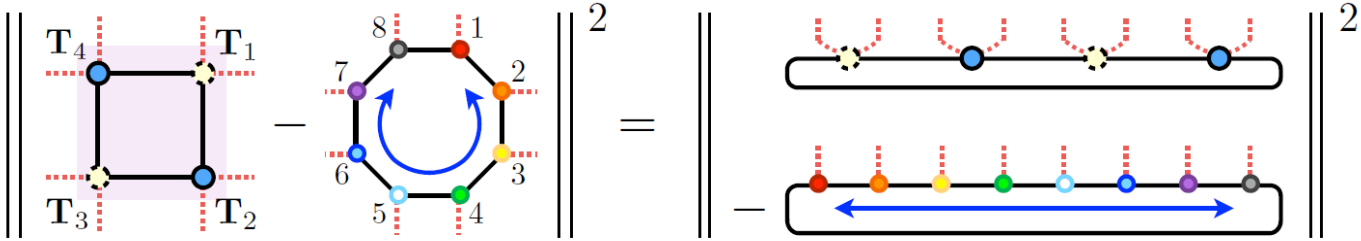
Coarse-graining



Projection



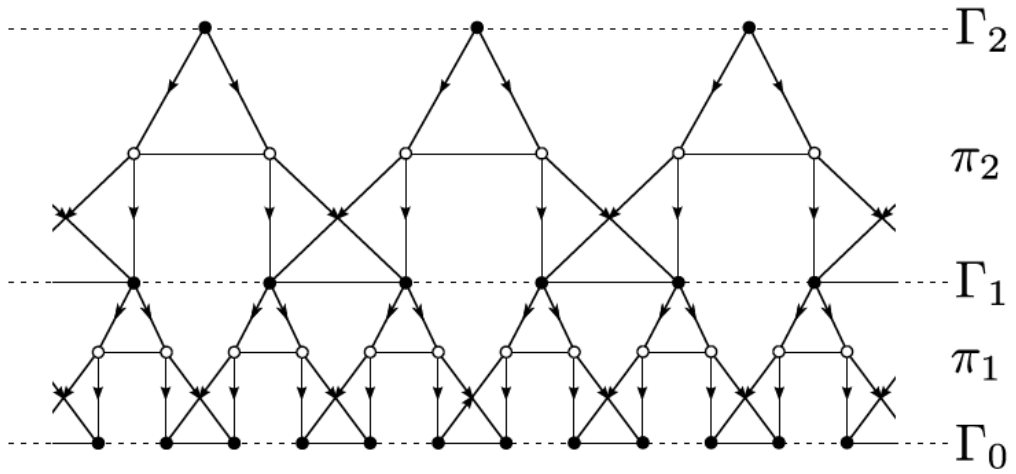
Variation



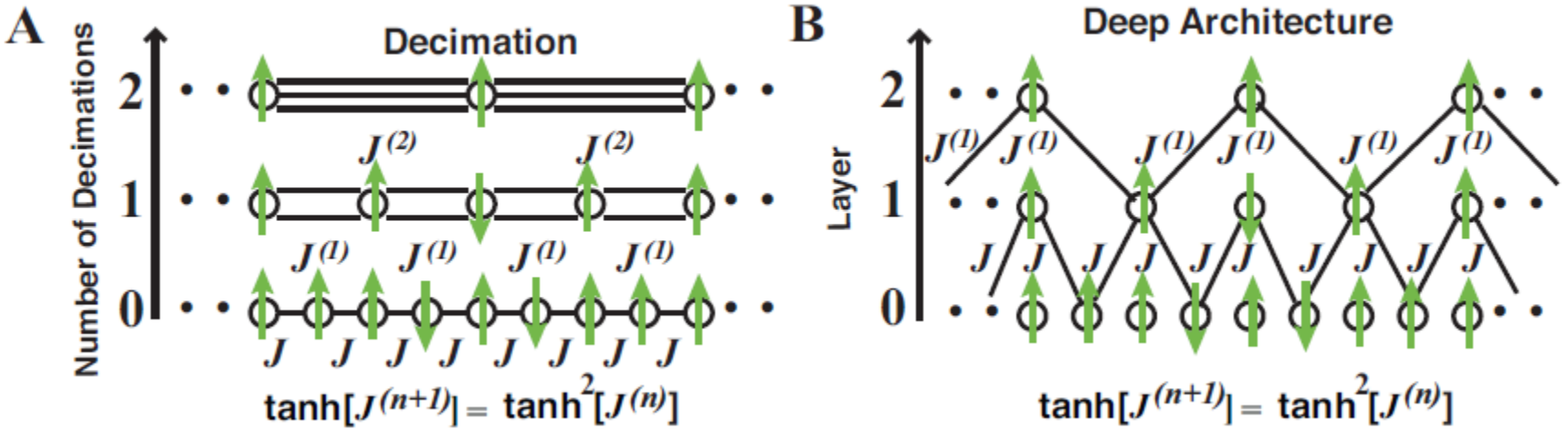
Outline

- Tensor network in a nutshell
- TN concepts in machine learning
- TN methods in machine learning

First contact: 2013



Bény, C. (2013). Deep learning and the renormalization group. *arXiv:1301.3124 [Quant-Ph]*. Retrieved from http://arxiv.org/abs/1301.3124



Mehta, P., & Schwab, D. J. (2014). An exact mapping between the variational renormalization group and deep learning. *arXiv Preprint arXiv:1410.3831*. Retrieved from http://arxiv.org/abs/1410.3831

Machine Learning	Quantum Physics
Nth-order tensor	rank-N tensor
high/low-order tensor	tensor of high/low dimension
ranks of TNs	bond dimensions of TNs
unfolding, matricization	grouping of indices
tensorization	splitting of indices
core	site
variables	open (physical) indices
ALS Algorithm	one-site DMRG or DMRG1
MALS Algorithm	two-site DMRG or DMRG2
column vector $\mathbf{x} \in \mathbb{R}^{I \times 1}$	ket $\|\Psi\rangle$
row vector $\mathbf{x}^T \in \mathbb{R}^{1 \times I}$	bra $\langle\Psi\|$
inner product $\langle \mathbf{x}, \mathbf{x} \rangle = \mathbf{x}^T \mathbf{x}$	$\langle\Psi\|\Psi\rangle$
Tensor Train (TT)	Matrix Product State (MPS) (with Open Boundary Conditions (OBC))
Tensor Chain (TC)	MPS with Periodic Boundary Conditions (PBC)
Matrix TT	Matrix Product Operators (with OBC)
Hierarchical Tucker (HT)	Tree Tensor Network State (TTNS) with rank-3 tensors

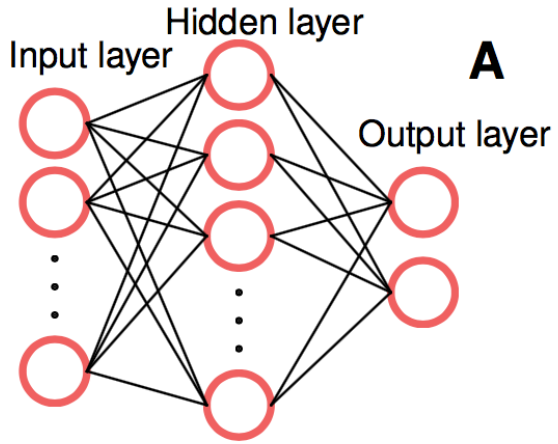
Cichocki, A. (2014). Tensor Networks for Big Data Analytics and Large-Scale Optimization Problems. *arXiv:1407.3124 [Cs, Math]*. Retrieved from http://arxiv.org/abs/1407.3124

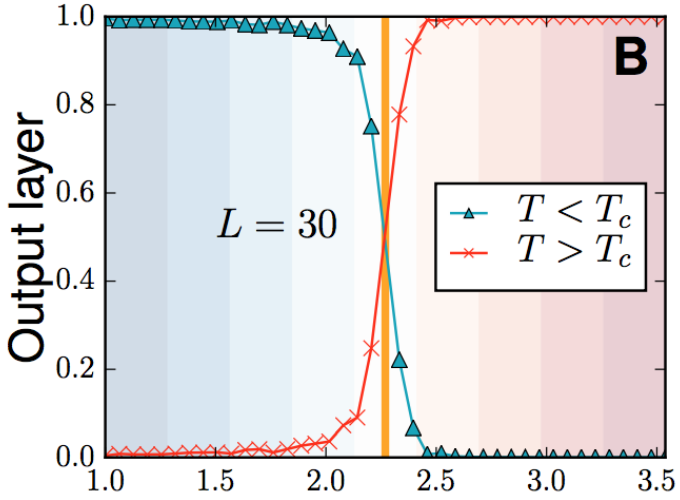
Tensor Networks	Neural Networks and Graphical Models in ML/Statistics
TT/MPS	Hidden Markov Models (HMM)
HT/TTNS	Deep Learning Neural Networks, Gaussian Mixture Model (GMM)
PEPS	Markov Random Field (MRF), Conditional Random Field (CRF)
MERA	Wavelets, Deep Belief Networks (DBN)
ALS, DMRG/MALS Algorithms	Forward-Backward Algorithms, Block Nonlinear Gauss-Seidel Methods

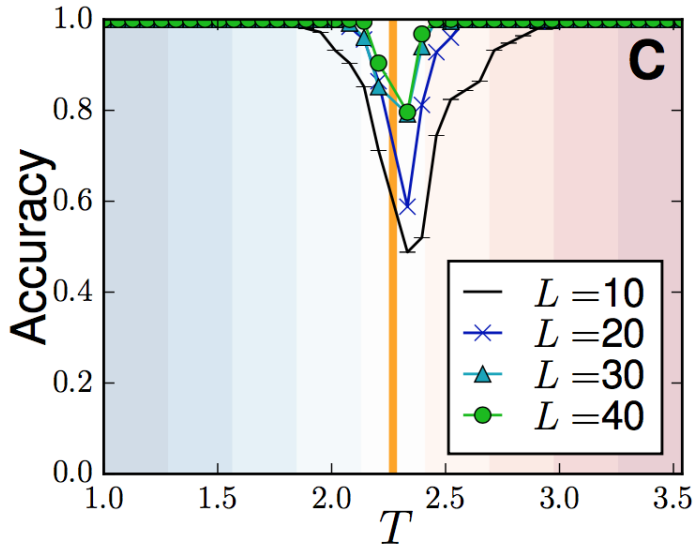
1. Cichocki, A. *et al.* Low-Rank Tensor Networks for Dimensionality Reduction and Large-Scale Optimization Problems: Perspectives and Challenges PART 1. *Foundations and Trends® in Machine Learning* **9,** 249–429 (2016).

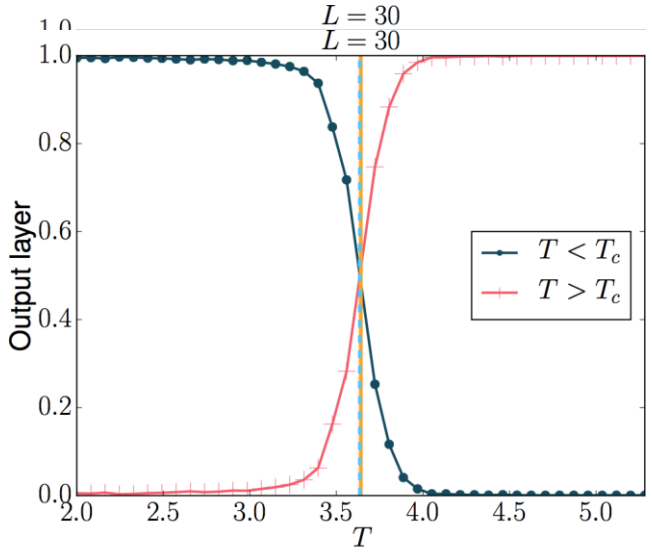
Fashionable: 2015


KBA
Korea Baduk Association
Google DeepMind
Challenge Match
8 - 15 March 2016
AlphaGo
ALPHAGO
01:58:37
AlphaGo
Lee Sedol
Google DeepMind
Challenge Match
8 - 15 March 2016

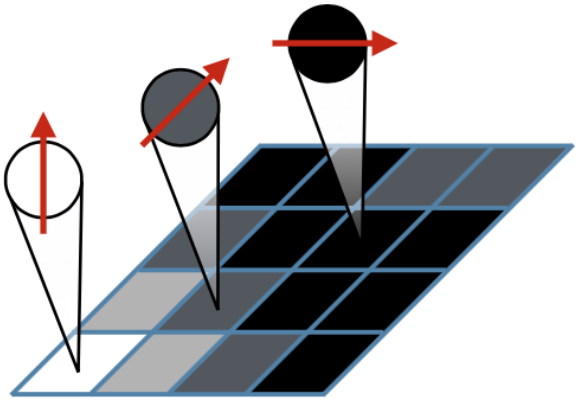


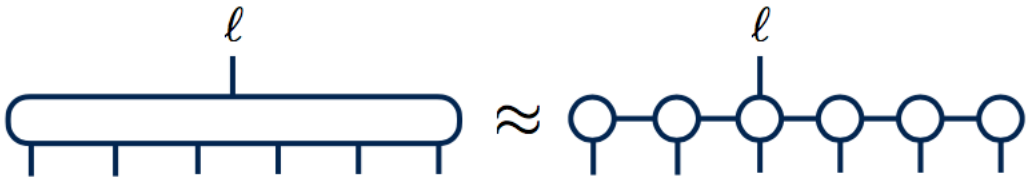




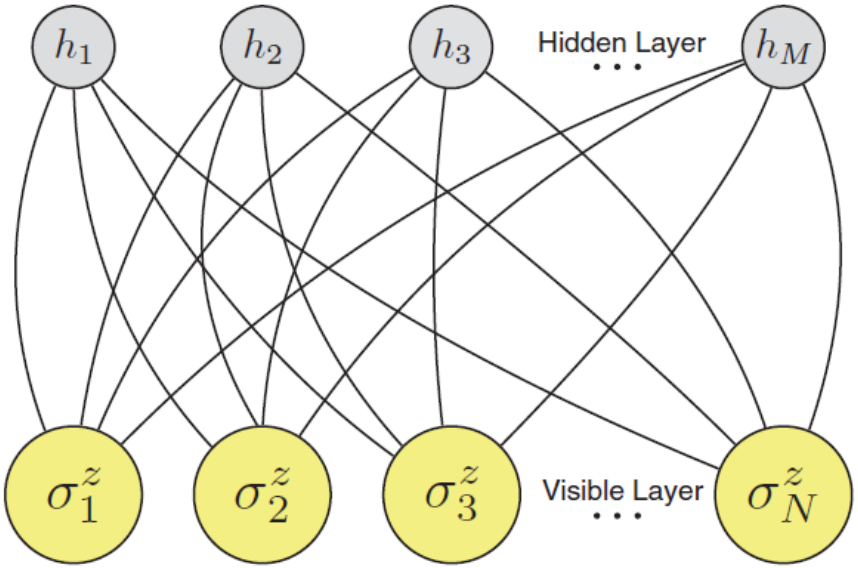


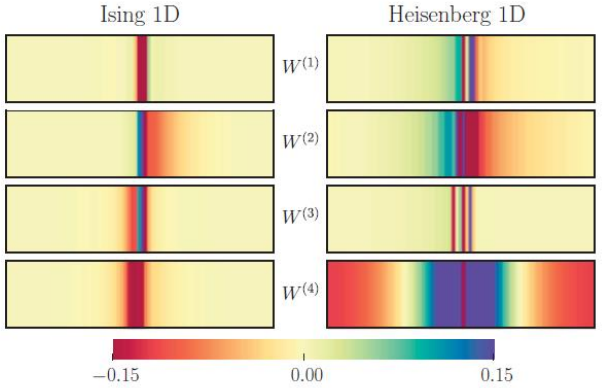
Carrasquilla, J., & Melko, R. G. (2016). Machine learning phases of matter. *arXiv Preprint arXiv:1605.01735*. Retrieved from http://arxiv.org/abs/1605.01735

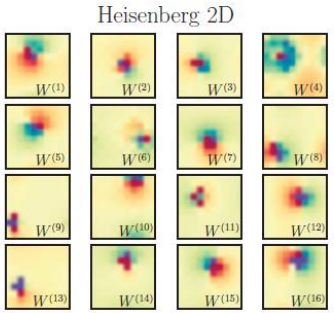


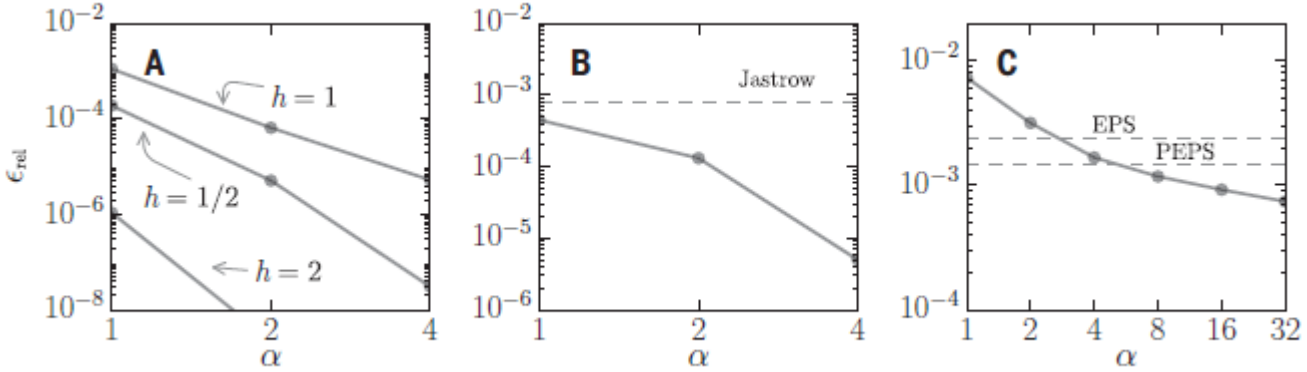


Stoudenmire, E. M., & Schwab, D. J. (2016). Supervised Learning with Quantum-Inspired Tensor Networks. *arXiv:1605.05775 [Cond-Mat, Stat]*. Retrieved from http://arxiv.org/abs/1605.05775









Carleo, G., & Troyer, M. (2017). Solving the quantum many-body problem with artificial neural networks. *Science*, *355*(6325), 602–606. https://doi.org/10.1126/science.aag2302

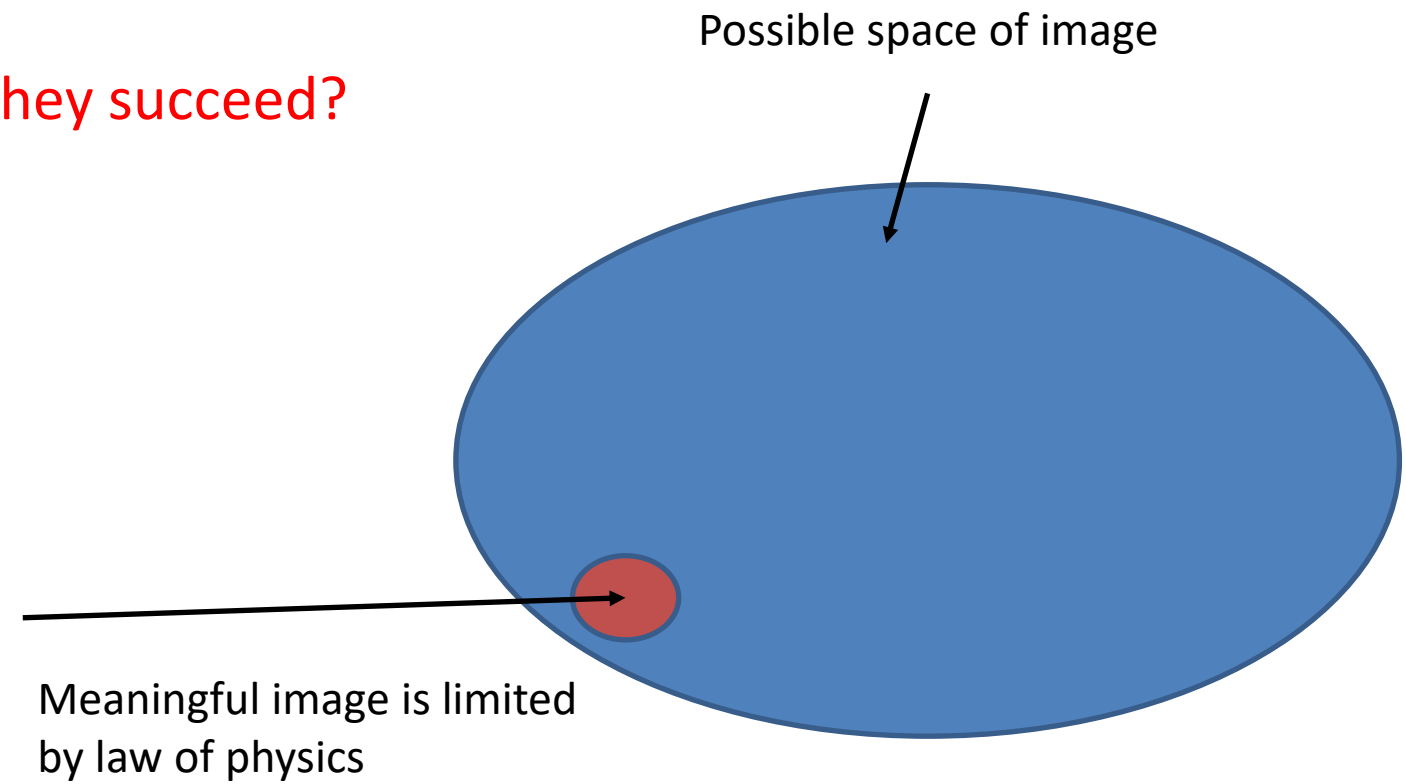
Reasoning: 2016-

Number of atoms in universe: 10^{78}

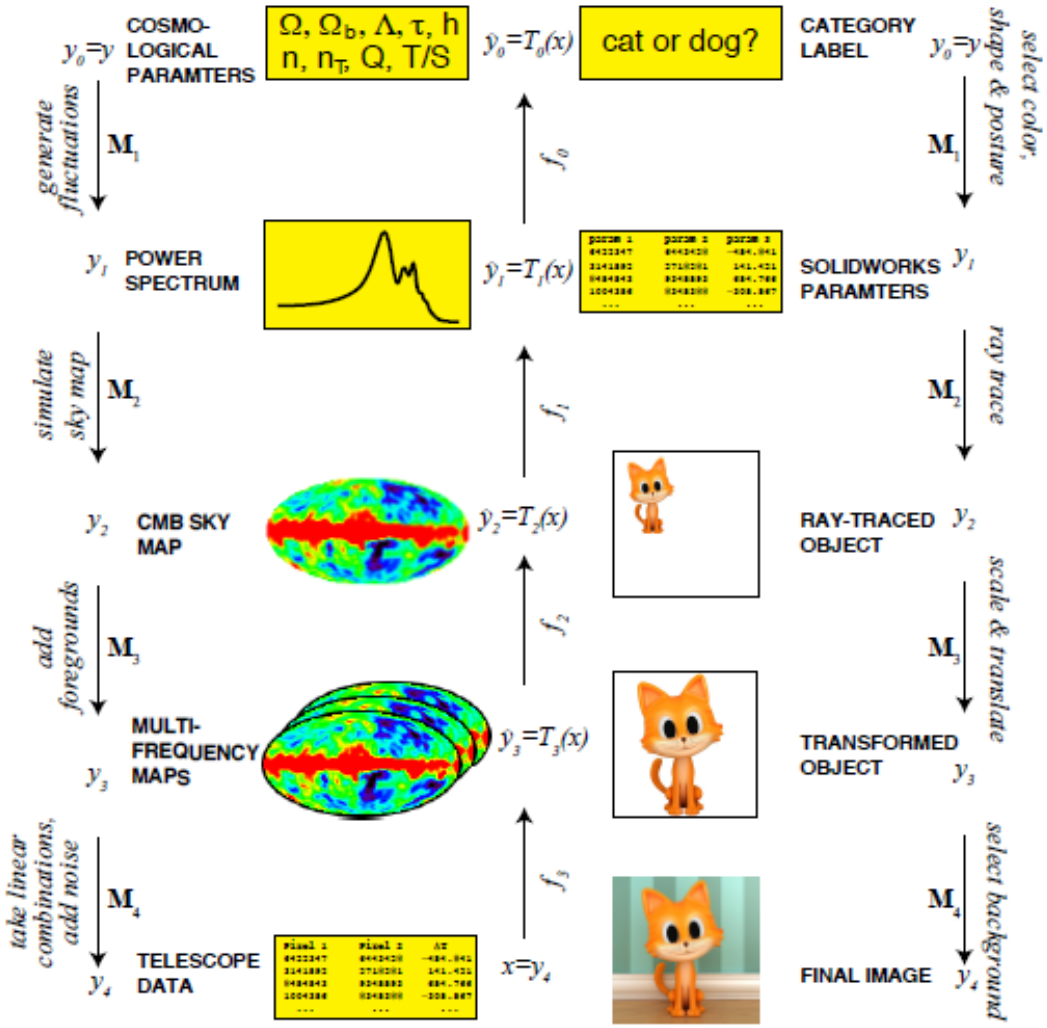
Number of possible samples for 28 * 28 gray image: 256^{784}

Number of parameters in RBM : $10^4 \sim 10^8$

Why they succeed?



H. W. Lin and M. Tegmark, Why does deep and cheap learning work so well， arXiv:1608.08225



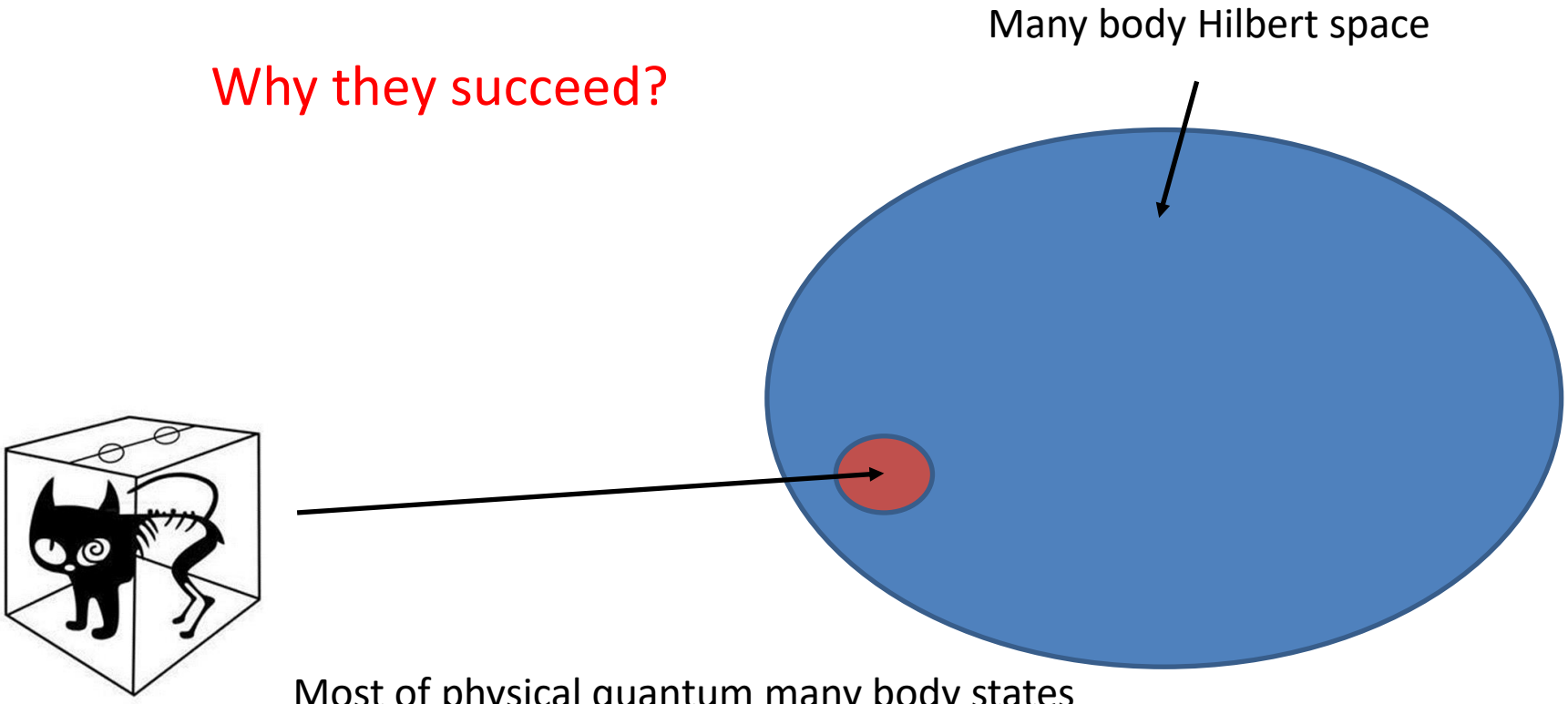
Lin, H. W., & Tegmark, M. (2016). Why does deep and cheap learning work so well? *arXiv:1608.08225 [Cond-Mat, Stat]*. Retrieved from http://arxiv.org/abs/1608.08225

Number of atoms in universe: 10^{78}

Number of states in manybody Hilbert space: tremendous!

Number of parameters in manybody numerical algorithm: negligible

Why they succeed?



Most of physical quantum many body states fulfill the area law.

H. W. Lin and M. Tegmark, Why does deep and cheap learning work so well， arXiv:1608.08225

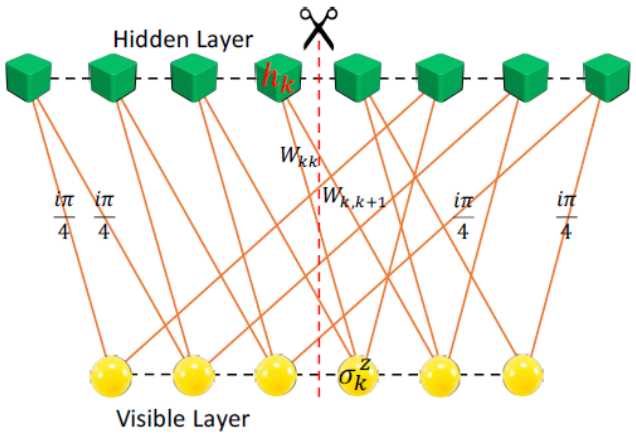
Similarities between
machine learning and quantum many body physics:

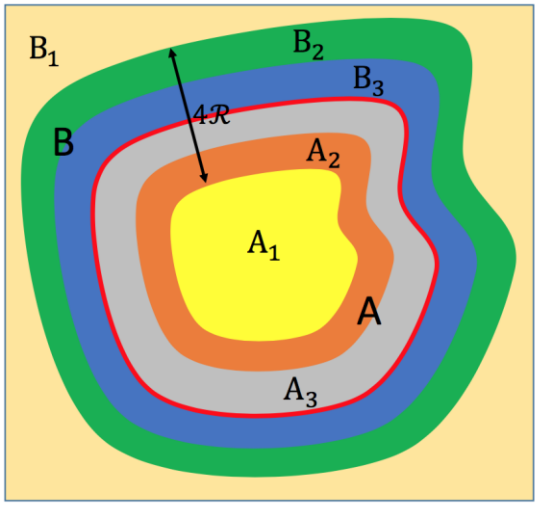
Using few parameters to approximate
exponentially large number of probabilities of data.

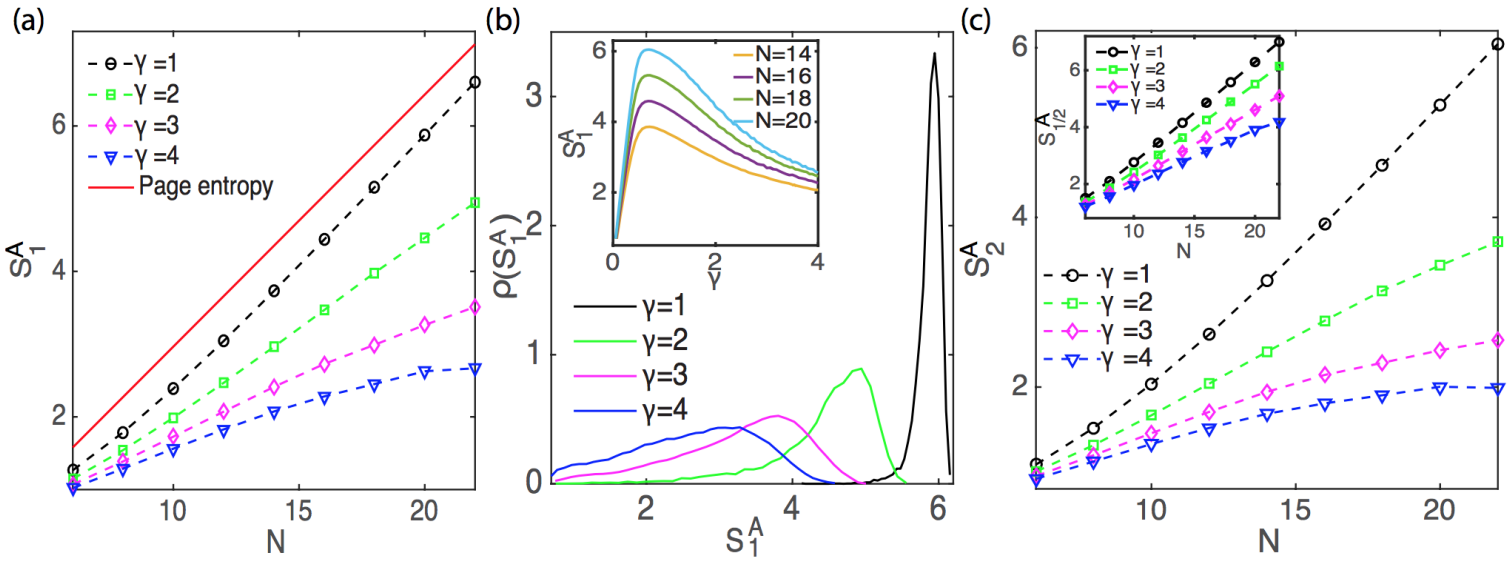
Physics	Machine learning
Hamiltonian	Surprisal $-\ln p$
Simple H	Cheap learning
Quadratic H	Gaussian p
Locality	Sparsity
Translationally symmetric H	Convnet
Computing p from H	Softmaxing
Spin	Bit
Free energy difference	KL-divergence
Effective theory	Nearly lossless data distillation
Irrelevant operator	Noise
Relevant operator	Feature

TABLE I: Physics-ML dictionary.

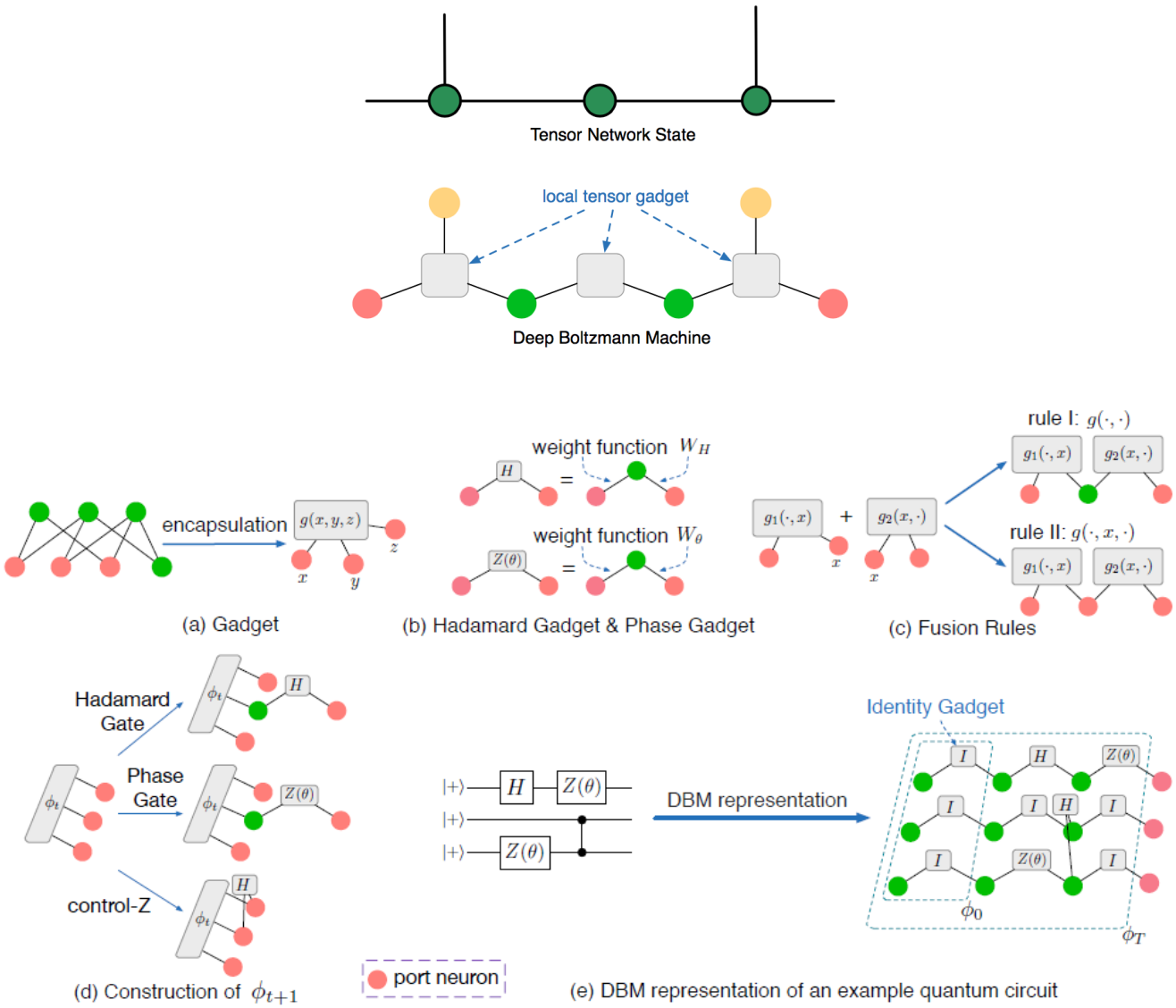
H. W. Lin and M. Tegmark, Why does deep and cheap learning work so well， arXiv:1608.08225



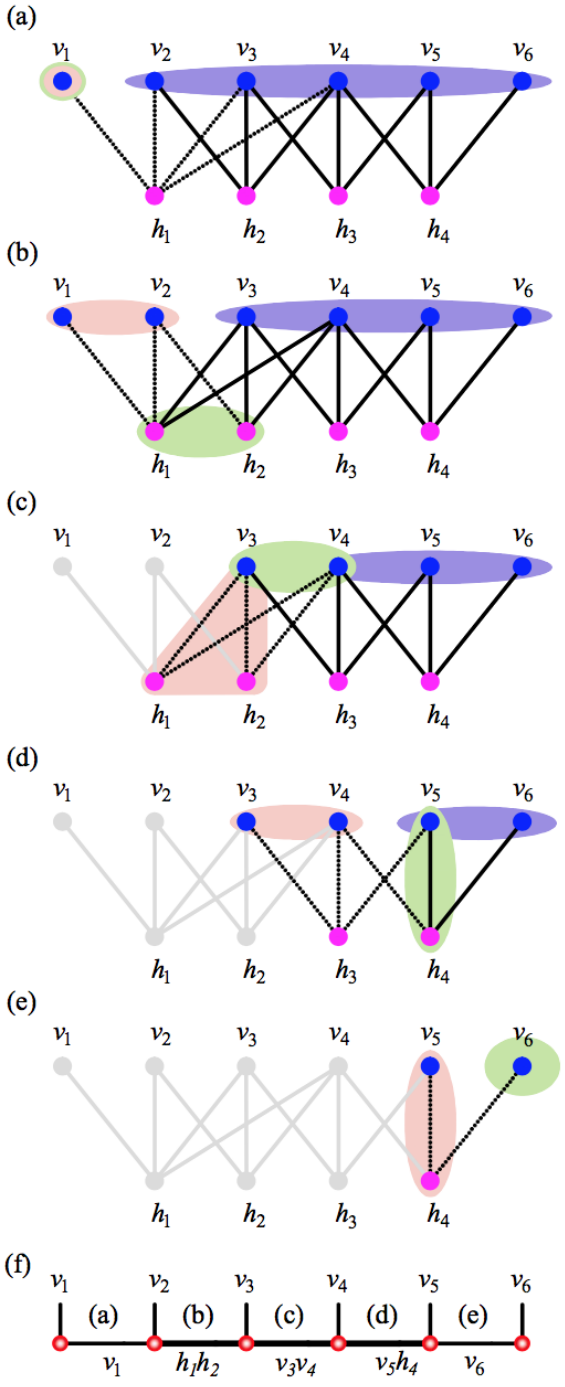


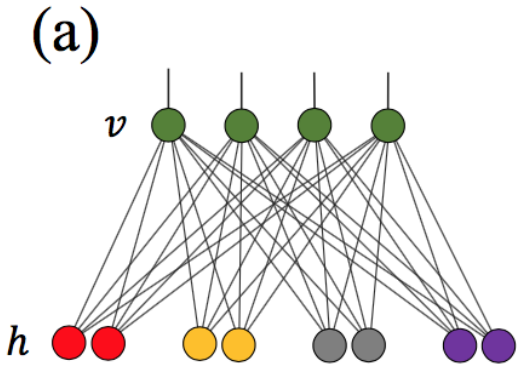


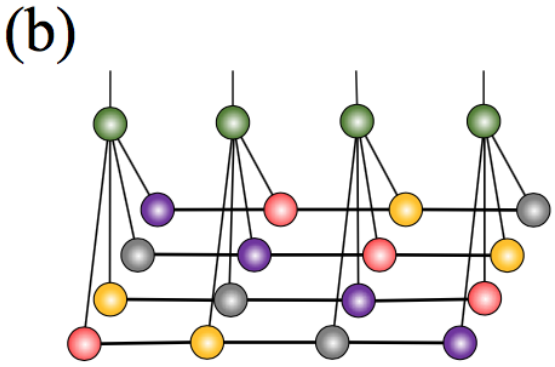
Deng, D.-L., Li, X., & Sarma, S. D. (2017). Quantum Entanglement in Neural Network States. *Physical Review X*, *7*(2). https://doi.org/10.1103/PhysRevX.7.021021



Gao, X., & Duan, L.-M. (2017). Efficient Representation of Quantum Many-body States with Deep Neural Networks. *arXiv:1701.05039 [Cond-Mat, Physics:quant-Ph]*. Retrieved from http://arxiv.org/abs/1701.05039





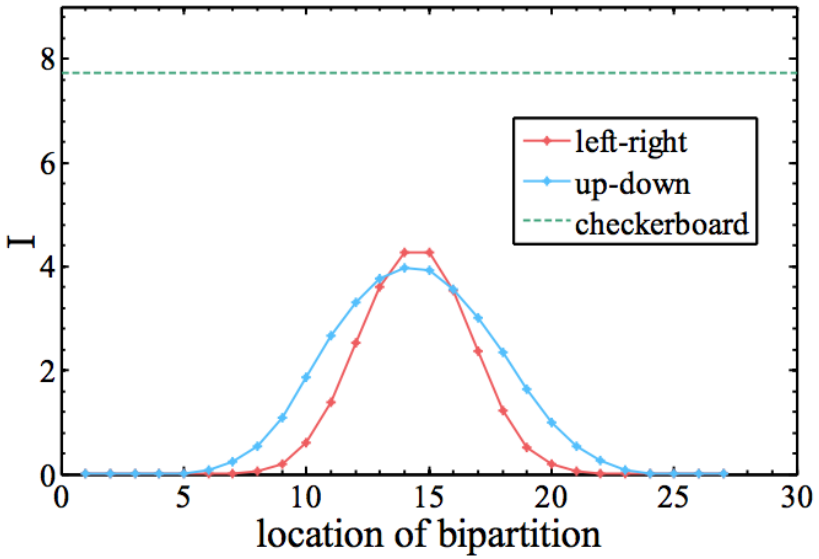


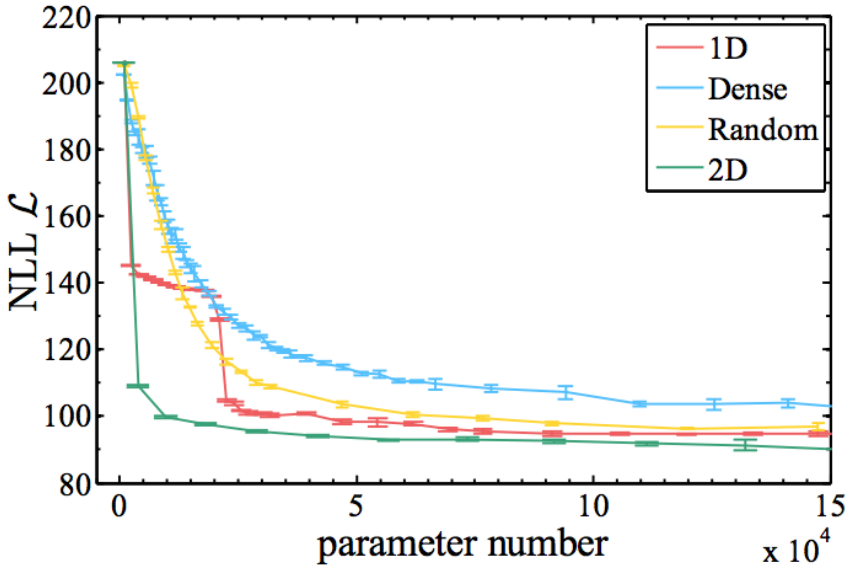
Chen, J., Cheng, S., Xie, H., Wang, L., & Xiang, T. (2017). On the Equivalence of Restricted Boltzmann Machines and Tensor Network States. *arXiv:1701.04831 [Cond-Mat, Physics:quant-Ph, Stat]*. Retrieved from http://arxiv.org/abs/1701.04831

$$I(\mathcal{X}:\mathcal{Y}) = -\left\langle \ln \left\langle \frac{\pi(x,y')\pi(x',y)}{\pi(x',y')\pi(x,y)} \right\rangle_{x',y'} \right\rangle_{x,y}, \qquad (5)$$

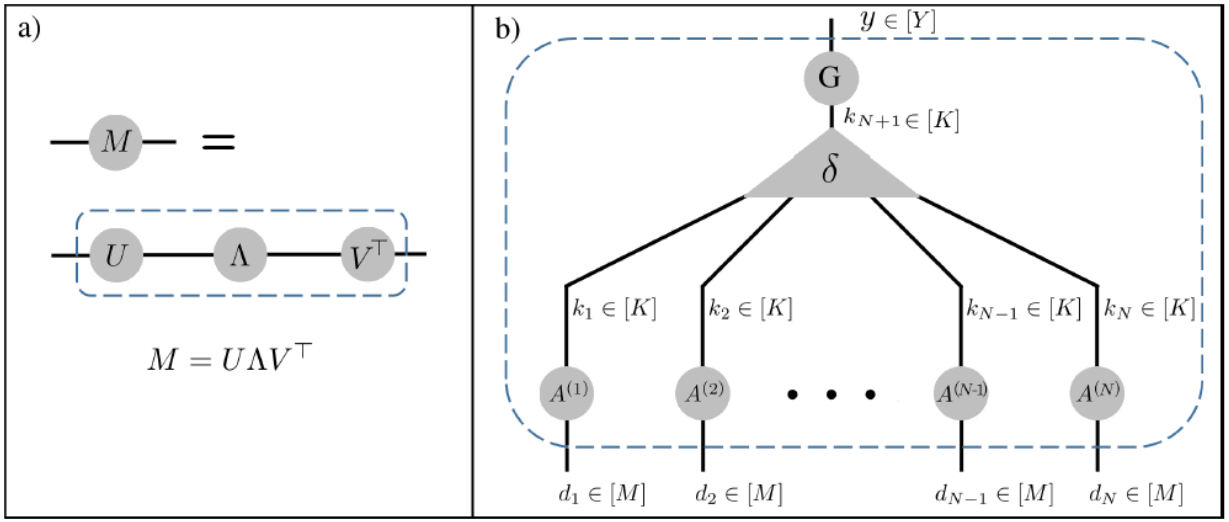
d the second Rényi entropy as

$$S^{\mathrm{R}} = -\ln \left\langle \left\langle \frac{\Psi(x,y')\Psi(x',y)}{\Psi(x',y')\Psi(x,y)} \right\rangle_{x',y'} \right\rangle_{x,y}, \qquad (6)$$





Cheng, S., Chen, J. & Wang, L. Information Perspective to Probabilistic Modeling: Boltzmann Machines versus Born Machines. *arXiv:1712.04144 [cond-mat, physics:physics, physics:quant-ph, stat]* (2017).



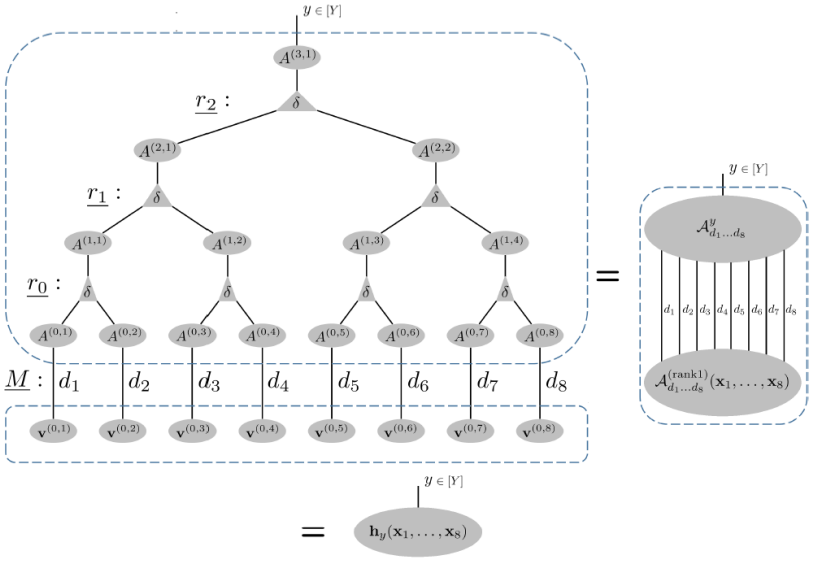
$$\phi^{1,j,\gamma} = \sum_{\alpha=1}^{r_0} a_\alpha^{1,j,\gamma} \mathbf{a}^{0,2j-1,\alpha} \otimes \mathbf{a}^{0,2j,\alpha}$$

$$\cdots$$

$$\phi^{l,j,\gamma} = \sum_{\alpha=1}^{r_{l-1}} a_\alpha^{l,j,\gamma} \underbrace{\phi^{l-1,2j-1,\alpha}}_{\text{order } 2^{l-1}} \otimes \underbrace{\phi^{l-1,2j,\alpha}}_{\text{order } 2^{l-1}}$$

$$\cdots$$

$$\mathcal{A}^y = \sum_{\alpha=1}^{r_{L-1}} a_\alpha^{L,y} \underbrace{\phi^{L-1,1,\alpha}}_{\text{order } \frac{N}{2}} \otimes \underbrace{\phi^{L-1,2,\alpha}}_{\text{order } \frac{N}{2}}.$$



Levine, Y., Yakira, D., Cohen, N., & Shashua, A. (2017). Deep Learning and Quantum Entanglement: Fundamental Connections with Implications to Network Design. *arXiv:1704.01552 [Quant-Ph]*. Retrieved from http://arxiv.org/abs/1704.01552

Neural network quantum states
vs
tensor network states

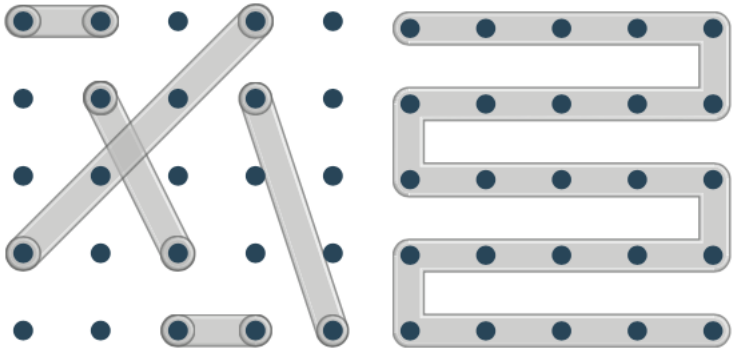
$$\psi = \frac{1}{Z} \sum_{\{s_1, s_2, \ldots, s_{|V|+|H|}\} \in \{\pm 1\}^{\otimes(|V|+|H|)}} e^{-\sum_j h_j s_j - \sum_{j,k} w_{jk} s_j s_k} \left| \{s_1, s_2, \ldots, s_{|V|}\} \right\rangle$$

$$\psi = \frac{1}{Z} \left(\int \mathrm{d}\xi_{|V|+1} \mathrm{d}\xi_{|V|+2} \cdots \mathrm{d}\xi_{|V|+|H|} e^{-\sum_{j,k} w_{jk} \xi_j \xi_k} \right) |0\rangle$$

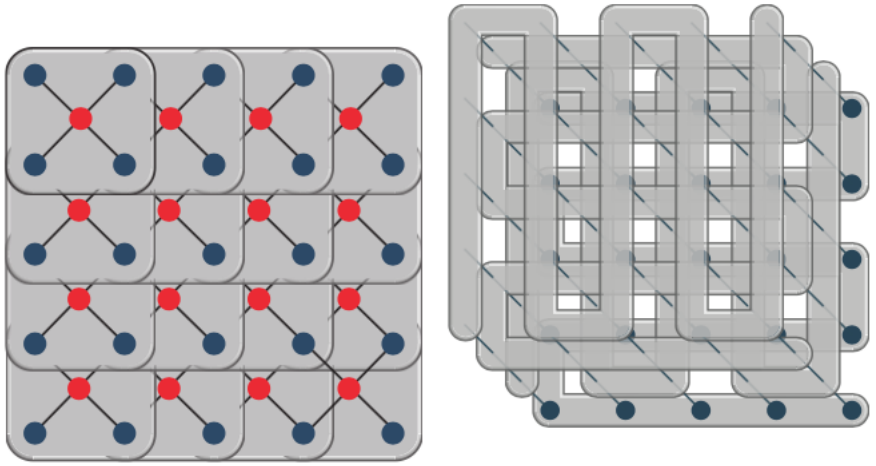
$$H = \sum_{\vec{x} \in \mathbf{Z}^2} \left(c^\dagger_{\vec{x}+\vec{i}} c_{\vec{x}} + c^\dagger_{\vec{x}+\vec{j}} c_{\vec{x}} + c^\dagger_{\vec{x}+\vec{i}} c^\dagger_{\vec{x}} + i c^\dagger_{\vec{x}+\vec{j}} c^\dagger_{\vec{x}} + \mathrm{h.c.} \right) - 2\mu \sum_{\vec{x} \in \mathbf{Z}^2} c^\dagger_{\vec{x}} c_{\vec{x}}, \quad \mu \in \mathbf{R}.$$

$$\psi = \int \left[\mathrm{d}\xi_{\vec{k}} \right] e^{\int_{\mathrm{BZ}} \mathrm{d}\vec{k} \left(v_{\vec{k}} \xi_{-\vec{k}} c^\dagger_{\vec{k}} - \frac{1}{2} \xi_{-\vec{k}} v_{-\vec{k}} u_{\vec{k}} \xi_{\vec{k}} \right)} |0\rangle$$

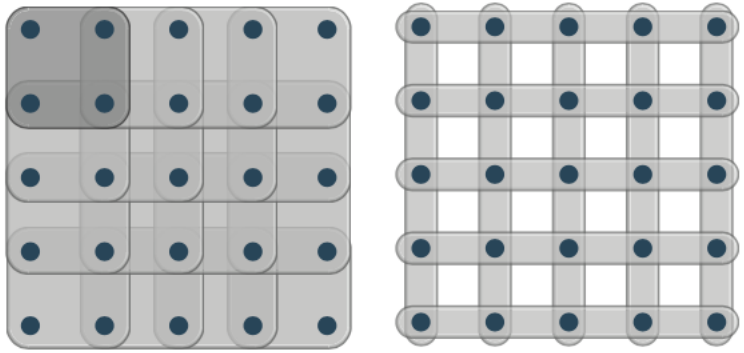
Huang, Y., & Moore, J. E. (2017). Neural network representation of tensor network and chiral states. *arXiv:1701.06246 [Cond-Mat]*. Retrieved from http://arxiv.org/abs/1701.06246



(a) Jastrow (b) MPS



(a) Local RBM as an EPS (b) RBM as a non-local SBS

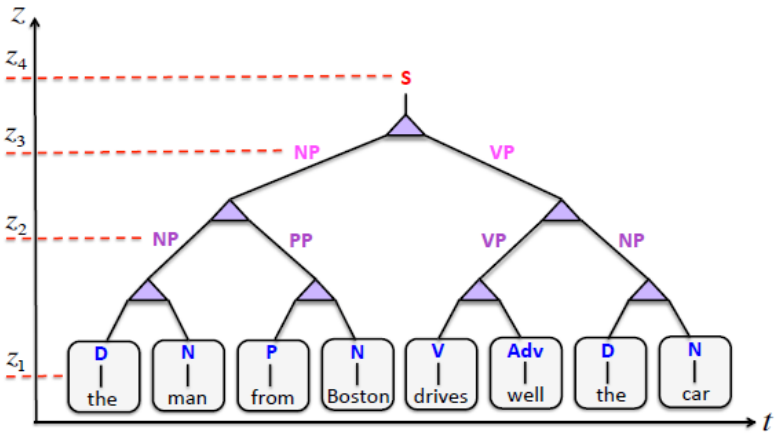


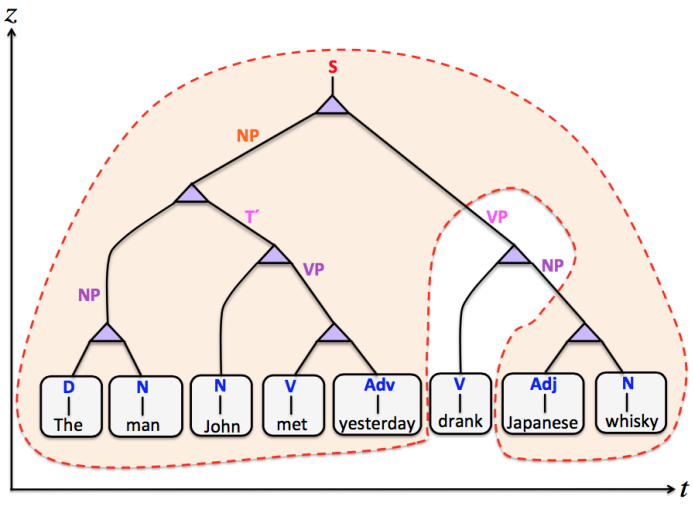
(c) EPS (d) SBS

Ansatz	$(E_w - E_0)/N$	$\lvert\langle\psi_w\rvert\psi_{\text{Laughlin}}\rangle\rvert$
EPS 2×2	4.3×10^{-2}	46.10%
EPS 3×3	2.2×10^{-2}	75.79%
sRBM $M' = 1$	8.3×10^{-2}	0.01%
sRBM $M' = 2$	3.1×10^{-2}	46.32%
sRBM $M' = 4$	2.5×10^{-2}	59.07%
RBM $M = N$	5.8×10^{-4}	99.7%
RBM $M = 2N$	1.1×10^{-5}	99.99%

Glasser, I., Pancotti, N., August, M., Rodriguez, I. D., & Cirac, J. I. (2017). Neural Networks Quantum States, String-Bond States and chiral topological states. *arXiv:1710.04045 [Cond-Mat, Physics:quant-Ph, Stat]*. Retrieved from http://arxiv.org/abs/1710.04045

Tensor concepts in language model





Gallego, A. J., & Orus, R. (2017). The physical structure of grammatical correlations: equivalences, formalizations and consequences. *arXiv:1708.01525 [Cond-Mat, Physics:physics, Physics:quant-Ph]*. Retrieved from http://arxiv.org/abs/1708.01525

Benefit

- Entanglement spectrum
- Gauge invariance
- Exact tensor decomposition math
- Expression power evaluation
- Tensor Algorithm
- ……

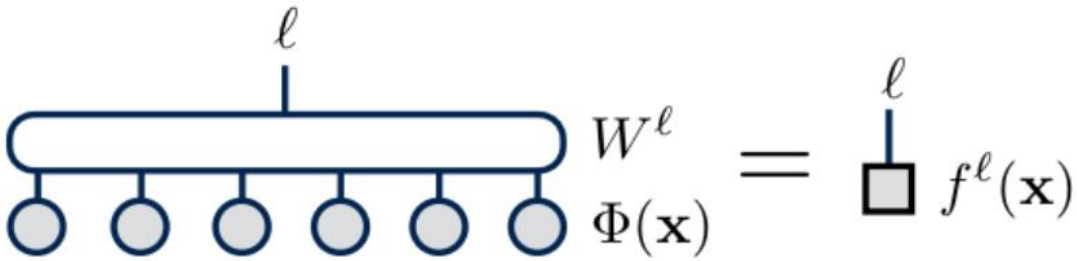
Disadvantage

- High cost
- Non-local terms are excluded
- ……

Outline

- Tensor network in a nutshell
- TN concepts in machine learning
- TN method in machine learning

representation



$$f^\ell(\mathbf{x}) = W^\ell \cdot \Phi(\mathbf{x})$$

$$W^\ell_{s_1 s_2 \cdots s_N} = \sum_{\{\alpha\}} A^{\alpha_1}_{s_1} A^{\alpha_1 \alpha_2}_{s_2} \cdots A^{\ell;\alpha_j \alpha_{j+1}}_{s_j} \cdots A^{\alpha_{N-1}}_{s_N}$$

Stoudenmire, E. M., & Schwab, D. J. (2016). Supervised Learning with Quantum-Inspired Tensor Networks. *arXiv:1605.05775 [Cond-Mat, Stat]*. Retrieved from http://arxiv.org/abs/1605.05775

target

Used quadratic cost function

$$C = \frac{1}{2} \sum_{n=1}^{N_T} \sum_{\ell} (f^{\ell}(\mathbf{x}_n) - \delta^{\ell}_{L_n})^2$$

Wants $f^{L_n}(x_n) = 1$, other $f^{\ell}(x_n) = 0$

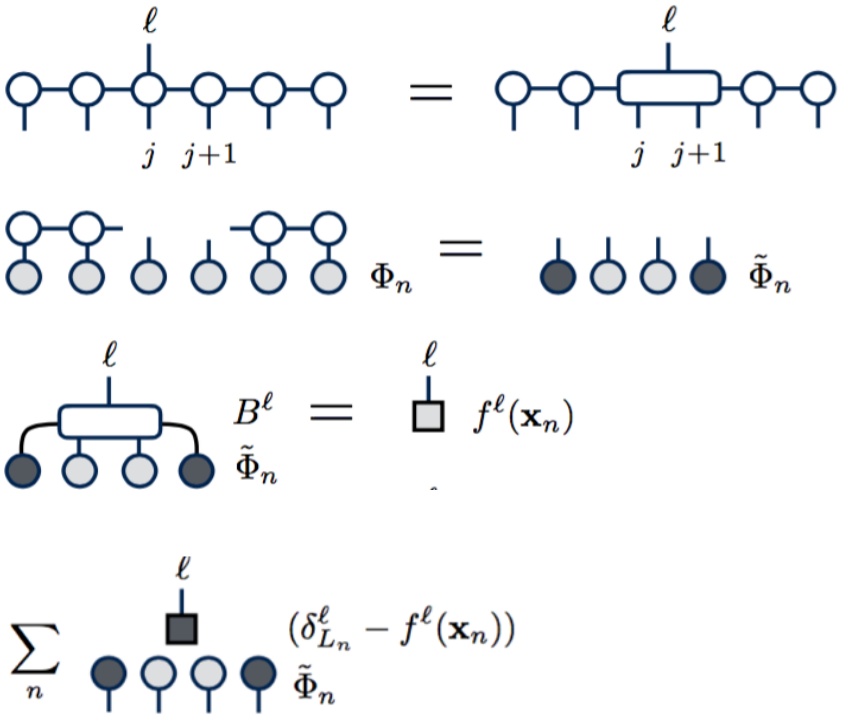
(tensor) gradient descent

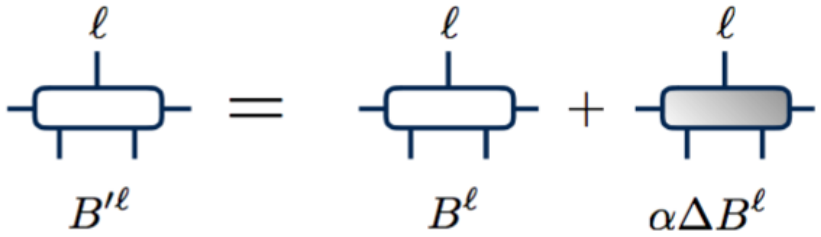
$$C = \frac{1}{2} \sum_{n=1}^{N_T} \sum_{\ell} (f^{\ell}(\mathbf{x}_n) - \delta^{\ell}_{L_n})^2$$

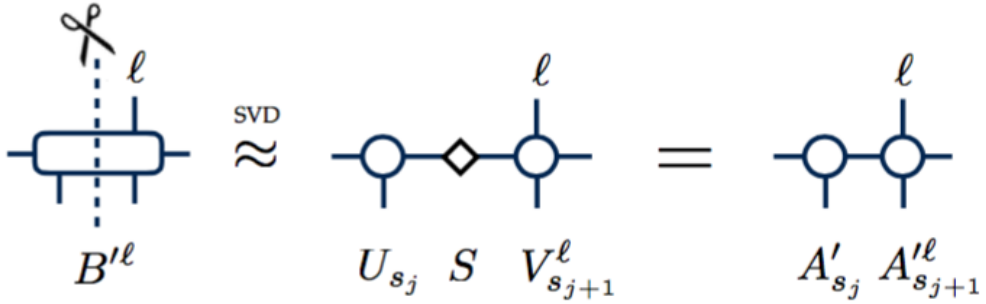
$$\Delta B^{\ell} \stackrel{\text{def}}{=} -\frac{\partial C}{\partial B^{\ell}}$$

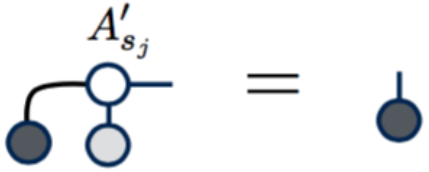
$$= \sum_{n=1}^{N_T} \sum_{\ell'} (\delta^{\ell'}_{L_n} - f^{\ell'}(\mathbf{x}_n)) \frac{\partial f^{\ell'}(\mathbf{x}_n)}{\partial B^{\ell}}$$

$$= \sum_{n=1}^{N_T} (\delta^{\ell}_{L_n} - f^{\ell}(\mathbf{x}_n)) \tilde{\Phi}_n \ .$$









$$B^{\alpha_{j-1}\ell\alpha_{j+1}}_{s_j s_{j+1}} = \sum_{\alpha'_j \alpha_j} U^{\alpha_{j-1}}_{s_j \alpha'_j} S^{\alpha'_j}{}_{\alpha_j} V^{\alpha_j \ell \alpha_{j+1}}_{s_{j+1}} \ ,$$

$$A'_{s_j} = U_{s_j} \qquad\qquad A'^{\ell}_{s_{j+1}} = S V^{\ell}_{s_{j+1}}$$

The scaling of the above algorithm is $d^3\, m^3\, N\, N_L\, N_T$

MNIST Experiment

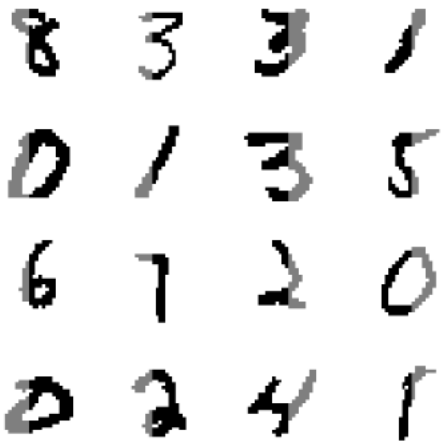
Results

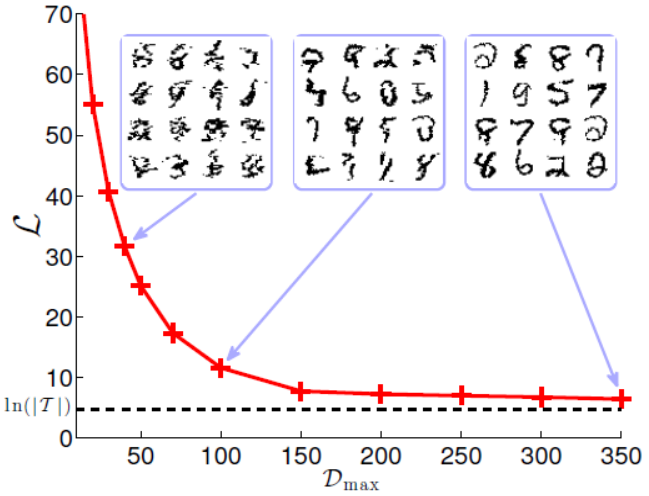
Only ~3 sweeps needed to converge

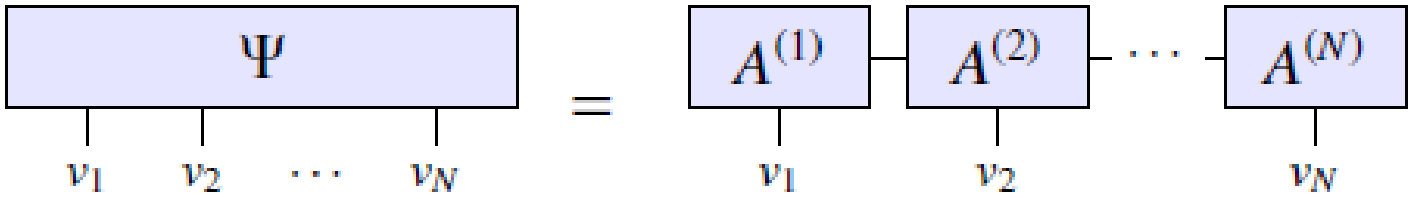
Bond dimension	Test Set Error
m = 10	~5% (500/10,000 incorrect)
m = 20	~2% (200/10,000 incorrect)
m = 120	0.97% (97/10,000 incorrect)

Bond dimension	Test Set Error
m = 10	~5% (500/10,000 incorrect)
m = 20	~2% (200/10,000 incorrect)
m = 120	0.97% (97/10,000 incorrect)

1.7	LeCun et al. 1998
1.1	LeCun et al. 1998
1.1	LeCun et al. 1998
1.1	LeCun et al. 1998
0.95	LeCun et al. 1998
0.85	LeCun et al. 1998
0.8	LeCun et al. 1998
0.7	LeCun et al. 1998
0.83	Lauer et al., Pattern Recognition 40-6, 2007
0.56	Lauer et al., Pattern Recognition 40-6, 2007
0.54	Lauer et al., Pattern Recognition 40-6, 2007
0.59	Labusch et al., IEEE TNN 2008
0.6	Simard et al., ICDAR 2003
0.4	Simard et al., ICDAR 2003
0.89	Ranzato et al., CVPR 2007
0.62	Ranzato et al., CVPR 2007
0.60	Ranzato et al., NIPS 2006
0.39	Ranzato et al., NIPS 2006
0.53	Jarrett et al., ICCV 2009
0.35	Ciresan et al. IJCAI 2011
0.27 +-0.02	Ciresan et al. ICDAR 2011
0.23	Ciresan et al. CVPR 2012







Zhao-Yu Han, Jun Wang, Heng Fan, Lei Wang, Pan Zhang, *Unsupervised Generative Modeling Using Matrix Product States*, 1709.01662

Training Algorithm: $\mathcal{O}(|\mathcal{T}|N\mathcal{D}_{\max}^3)$

$$\mathcal{L} = -\frac{1}{|\mathcal{T}|}\sum_{v\in\mathcal{T}} \ln\left[\mathbb{P}(v)\right]$$

$$\frac{\partial\mathcal{L}}{\partial A_{i_{k-1}i_{k+1}}^{(k,k+1)w_k w_{k+1}}} = \frac{Z'}{Z} - \frac{2}{|\mathcal{T}|}\sum_{v\in\mathcal{T}}\left[\frac{\Psi'(v)}{\Psi(v)}\right]$$

$$A_{i_{k-1}i_{k+1}}^{(k,k+1)w_k w_{k+1}} = A_{i_{k-1}i_{k+1}}^{(k,k+1)w_k w_{k+1}} - \eta\frac{\partial\mathcal{L}}{\partial A_{i_{k-1}i_{k+1}}^{(k,k+1)w_k w_{k+1}}}$$

Generative Algorithm: $\mathcal{O}(N\mathcal{D}_{\max}^3)$

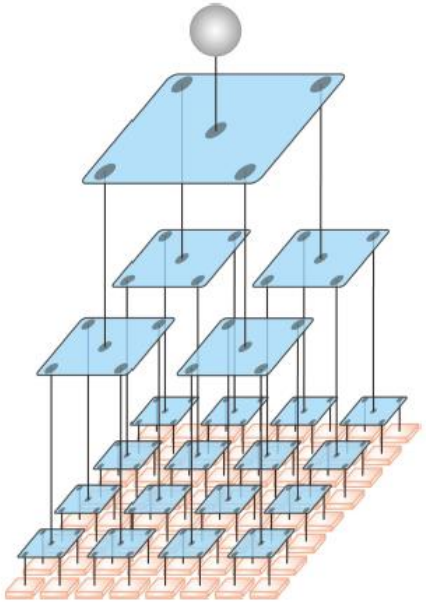
$$\mathbb{P}(v_{k-1}|v_k, v_{k+1}, \ldots, v_N) = \frac{\mathbb{P}(v_{k-1}, v_k, \ldots, v_N)}{\mathbb{P}(v_k, \ldots, v_N)}$$

$$\tilde{\mathbf{L}}^{[j]} = \mathbf{\Psi} \prod_{n=1}^{N} \mathbf{v}^{[j,n]},$$

$$f = \sum_{j=1}^{J} (\prod_{nn'} \mathbf{v}^{[j,n']\dagger} \mathbf{\Psi}^{\dagger} \mathbf{\Psi} \mathbf{v}^{[j,n]} - 2 \prod_{n} \mathbf{L}^{[j]\dagger} \mathbf{\Psi} \mathbf{v}^{[j,n]} + 1).$$

$$f = \sum_{j=1}^{J} |\tilde{\mathbf{L}}^{[j]} - \mathbf{L}^{[j]}|^2,$$

$$f = -\sum_{j=1}^{J} \prod_{n} \mathbf{L}^{[j]\dagger} \mathbf{\Psi} \mathbf{v}^{[j,n]}.$$



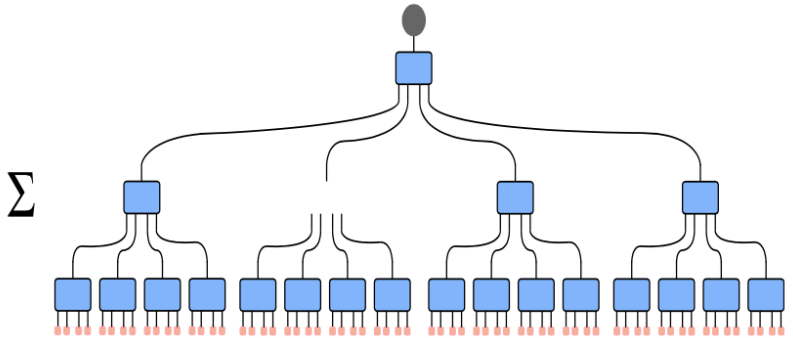


FIG. 3. Illustration of the environment tensor.

Liu, D., Ran, S.-J., Wittek, P., Peng, C., García, R. B., Su, G., & Lewenstein, M. (2017). Machine Learning by Two-Dimensional Hierarchical Tensor Networks: A Quantum Information Theoretic Perspective on Deep Architectures. *arXiv:1710.04833 [Cond-Mat, Physics:physics, Physics:quant-Ph, Stat]*. Retrieved from http://arxiv.org/abs/1710.04833

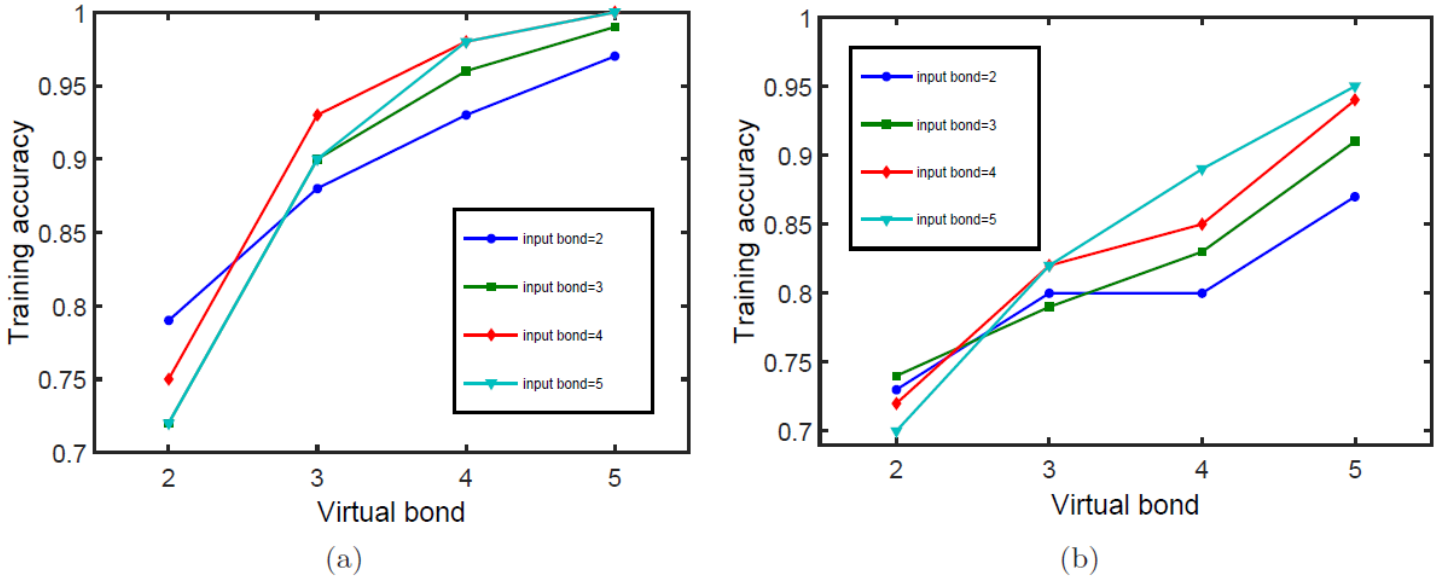
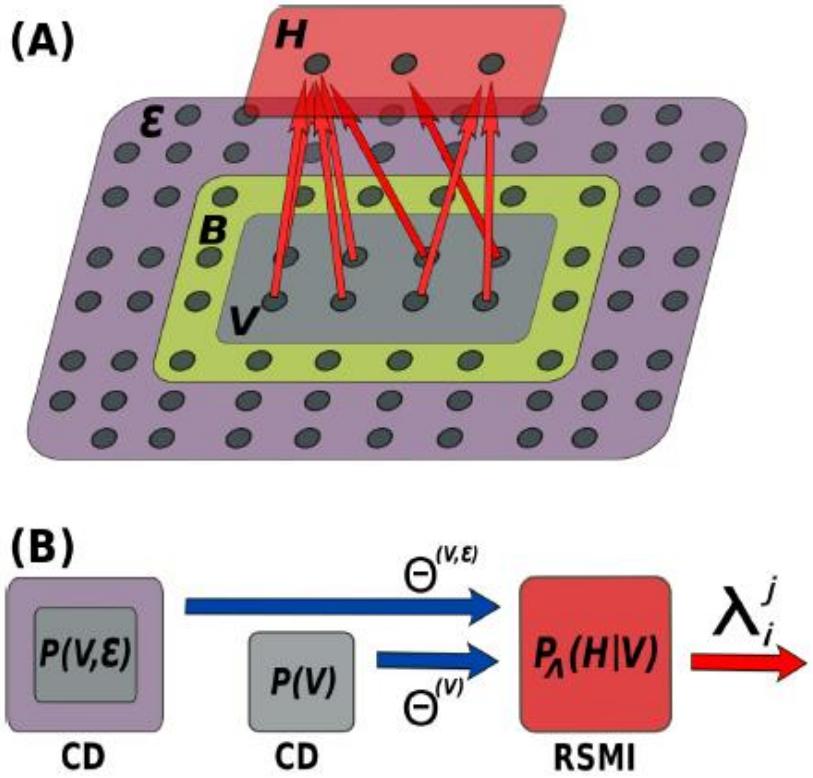
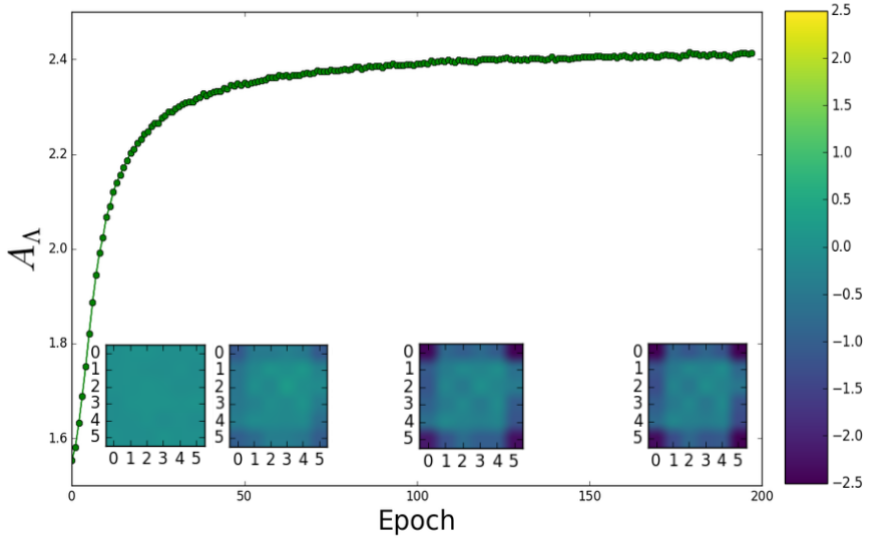
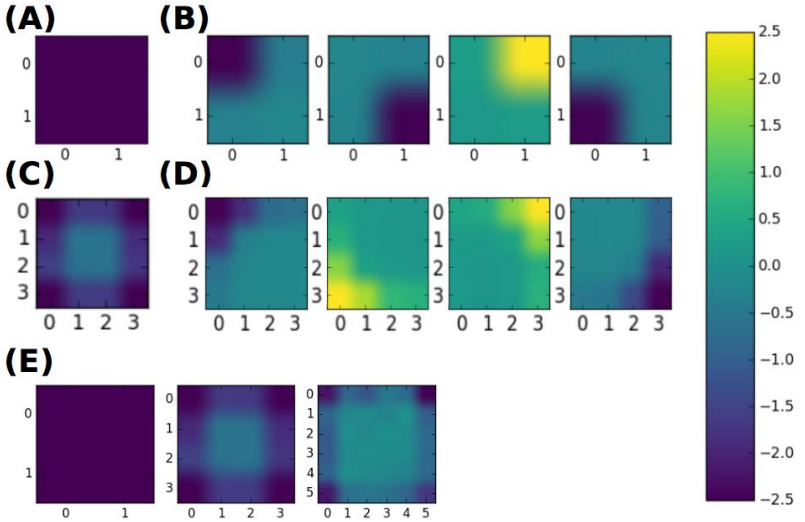


FIG. 4. Binary classification accuracy on CIFAR-10. (a) Number of training samples=200; (b) Number of training samples=600.

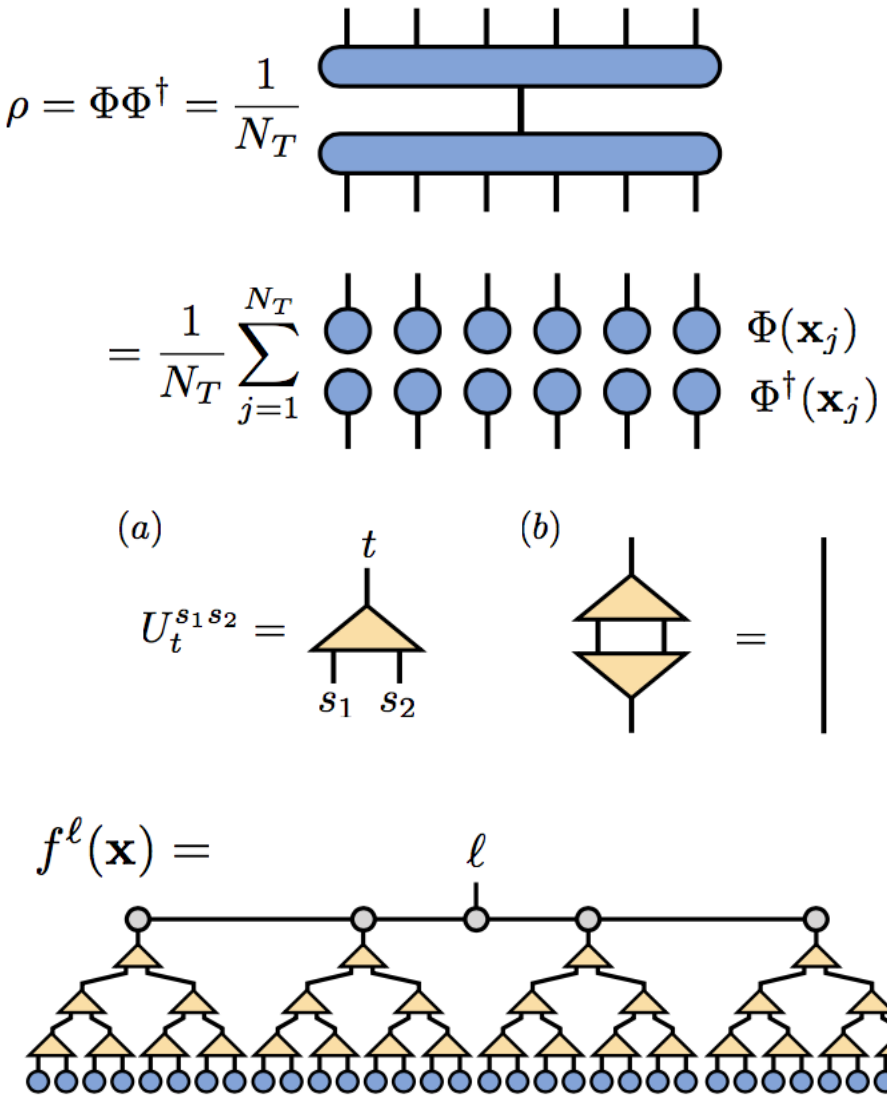
Liu, D., Ran, S.-J., Wittek, P., Peng, C., García, R. B., Su, G., & Lewenstein, M. (2017). Machine Learning by Two-Dimensional Hierarchical Tensor Networks: A Quantum Information Theoretic Perspective on Deep Architectures. *arXiv:1710.04833 [Cond-Mat, Physics:physics, Physics:quant-Ph, Stat]*. Retrieved from http://arxiv.org/abs/1710.04833

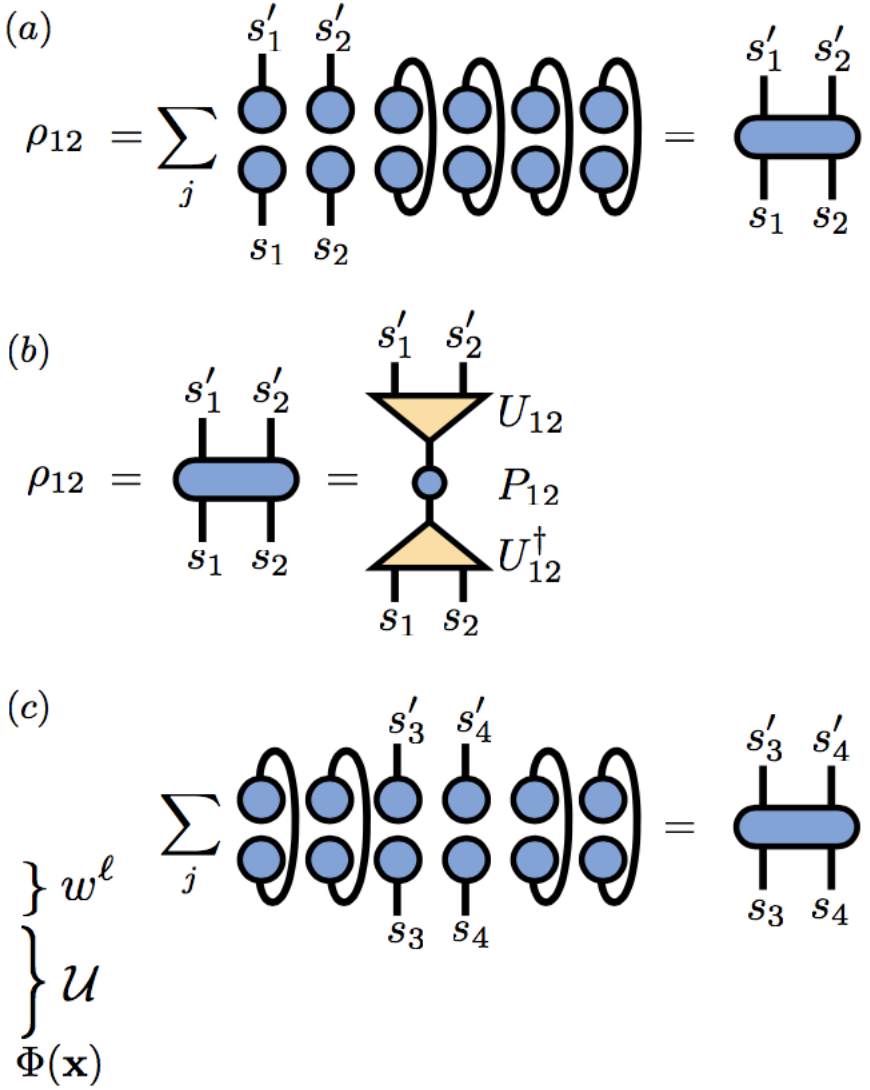






Koch-Janusz, M., & Ringel, Z. (2017). Mutual Information, Neural Networks and the Renormalization Group. *arXiv:1704.06279 [Cond-Mat]*. Retrieved from http://arxiv.org/abs/1704.06279

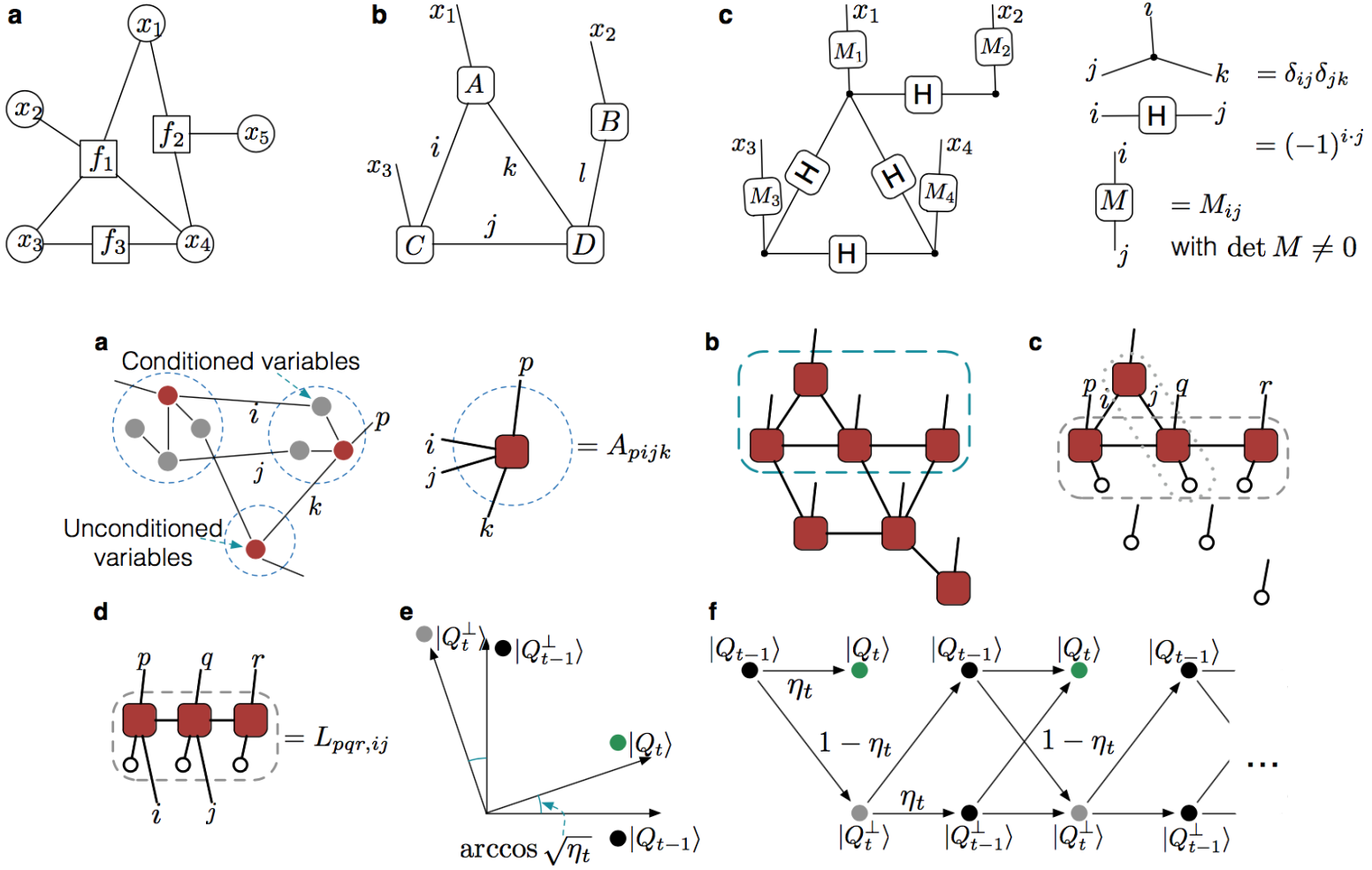




1.

Stoudenmire, E. M. Learning Relevant Features of Data with Multi-scale Tensor Networks. *arXiv:1801.00315 [cond-mat, stat]* (2017).

The optimized model reaches 95.38% accuracy on the training set, and 88.97% accuracy on the testing set. While $\sim 89\%$ test accuracy is significantly less than achieved on the much easier MNIST handwriting dataset, many of the available benchmarks using state-of-the-art approaches for fashion MNIST without preprocessing are in fact comparable to the results here, for example XGBoost (89.8%), AlexNet (89.9%), and a two-layer convolutional neural network trained with Keras (87.6%). Better results are attainable; the best we are aware of is a GoogLeNet reaching 93.7% test accuracy. But the



learning based on a quantum generative model. We prove that our proposed model is exponentially more powerful to represent probability distributions compared with classical generative models and has exponential speedup in training and inference at least for some instances under a reasonable assumption in computational complexity theory. Our result opens a new direction for quantum machine learning and offers a remarkable example in which a quantum algorithm shows exponential

Gao, X., Zhang, Z. & Duan, L. An efficient quantum algorithm for generative machine learning. *arXiv:1711.02038 [quant-ph, stat]* (2017).

Thanks

Take home message

- Tensor networks correlated to many ML architecture
- Datasets in ML are not totally unfamiliar to physicists
- Tensor viewpoints/techniques could be transferred to ML
- Lots of work to be done.