From Boltzmann Machines to Born Machines

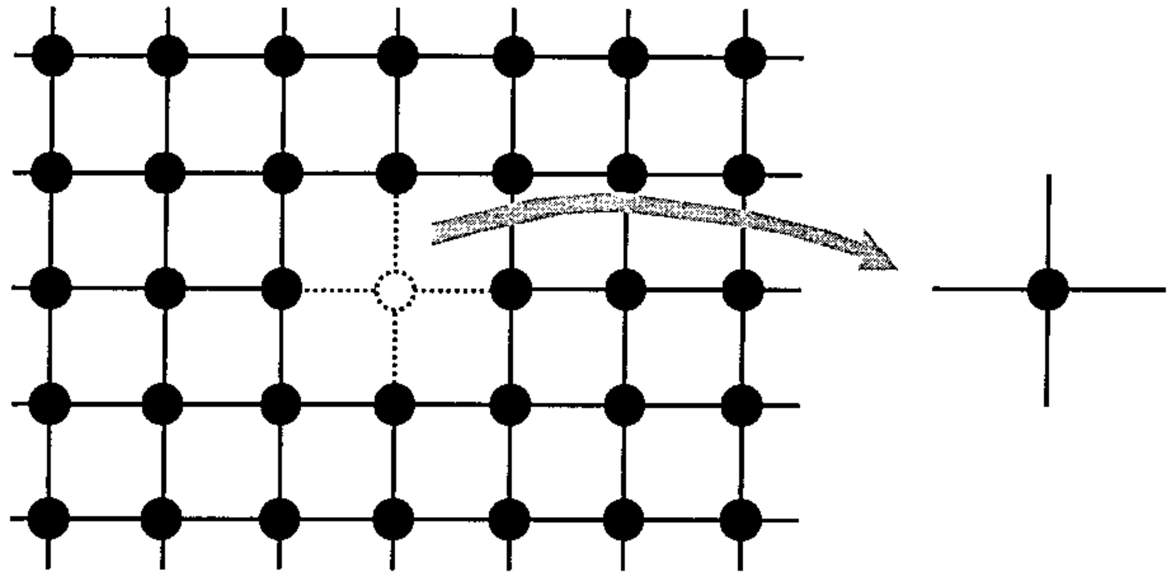
Lei Wang (王磊)
https://wangleiphy.github.io

Institute of Physics, Beijing
Chinese Academy of Sciences

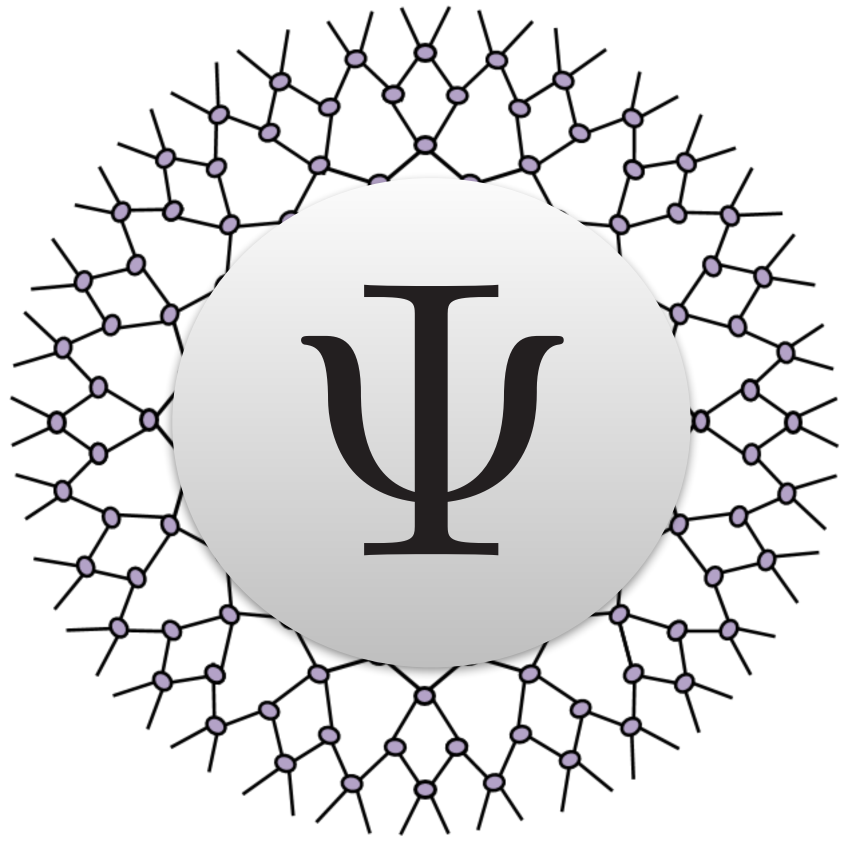
This talk is about
Physics for Machine Learning

Physicists' gifts to Machine Learning

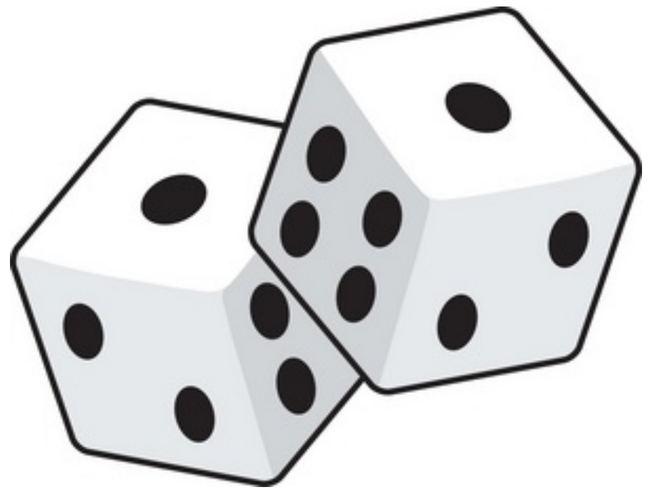
Mean Field Theory



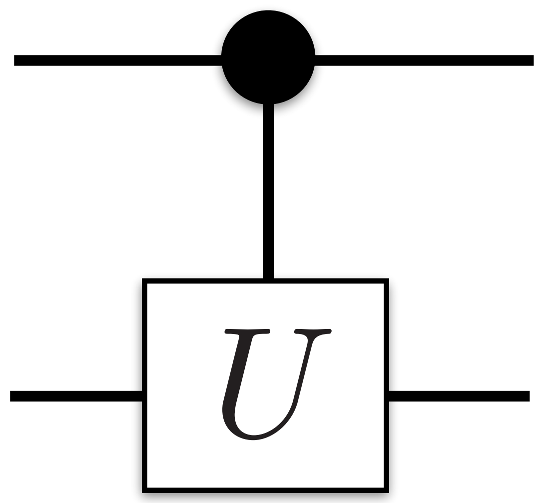
Tensor Networks



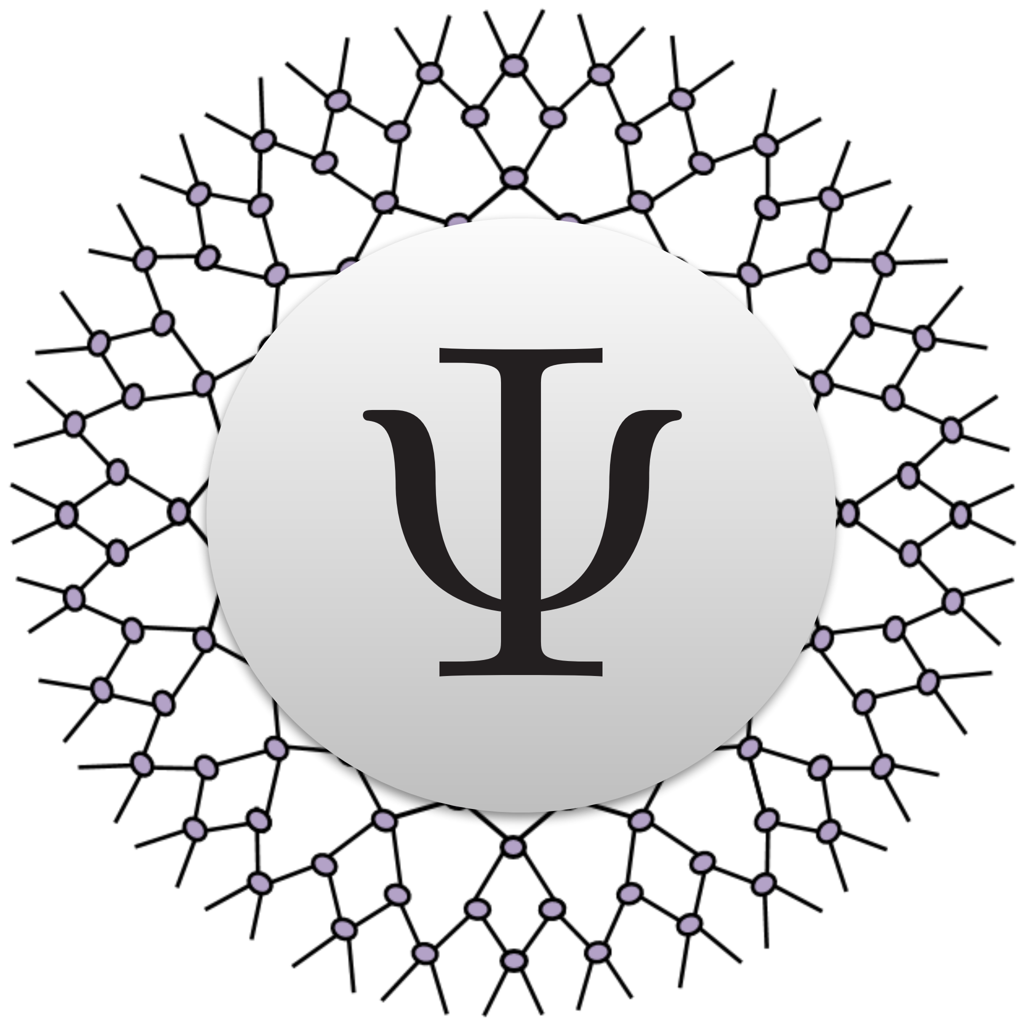
Monte Carlo Methods

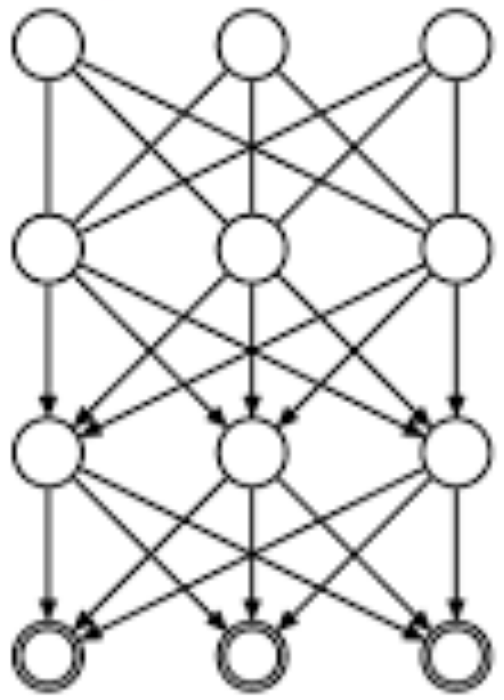


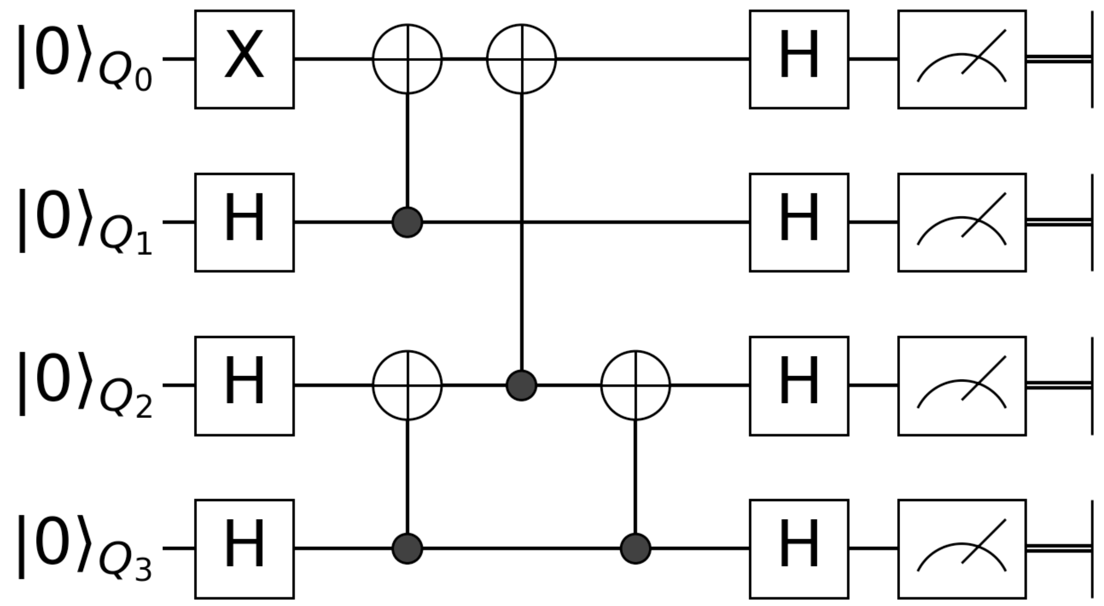
Quantum Computing



“Quantising” Machine Learning with Tensor Networks







Neural networks and Graphical probabilistic models

Glasser, Clark, Deng, Gao,
Chen, Cichocki, Levine …

Quantum circuits architecture and initialization

Kim, Swingle, Huggins,
Stoudenmire, …

Deep learning is more than function fitting

Discriminative

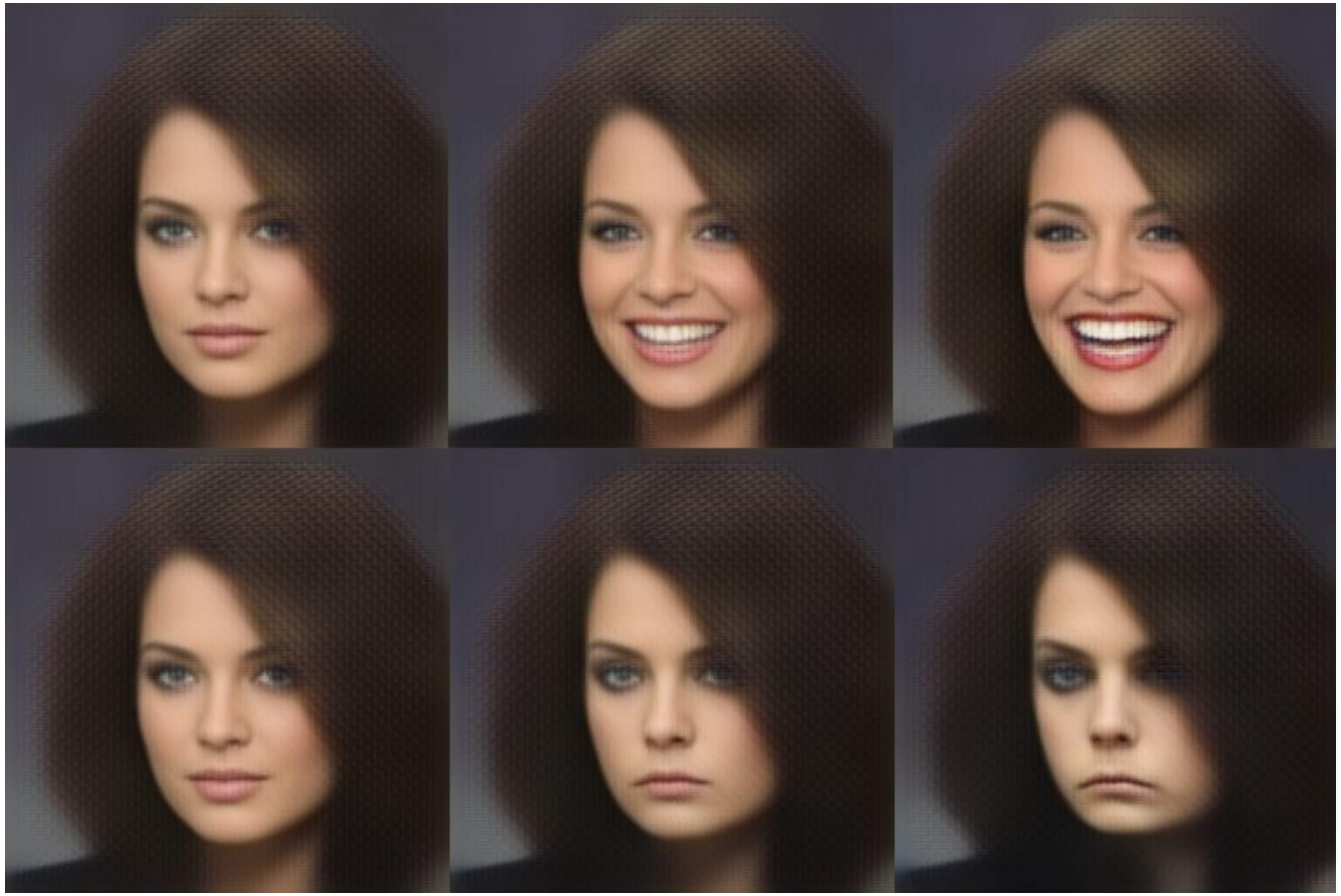
$$y = f(\boldsymbol{x})$$

or $p(y|\boldsymbol{x})$

Generative

$$p(\boldsymbol{x}|y) = \frac{p(\boldsymbol{x}, y)}{p(y)}$$

Interpolating the “smile vector”

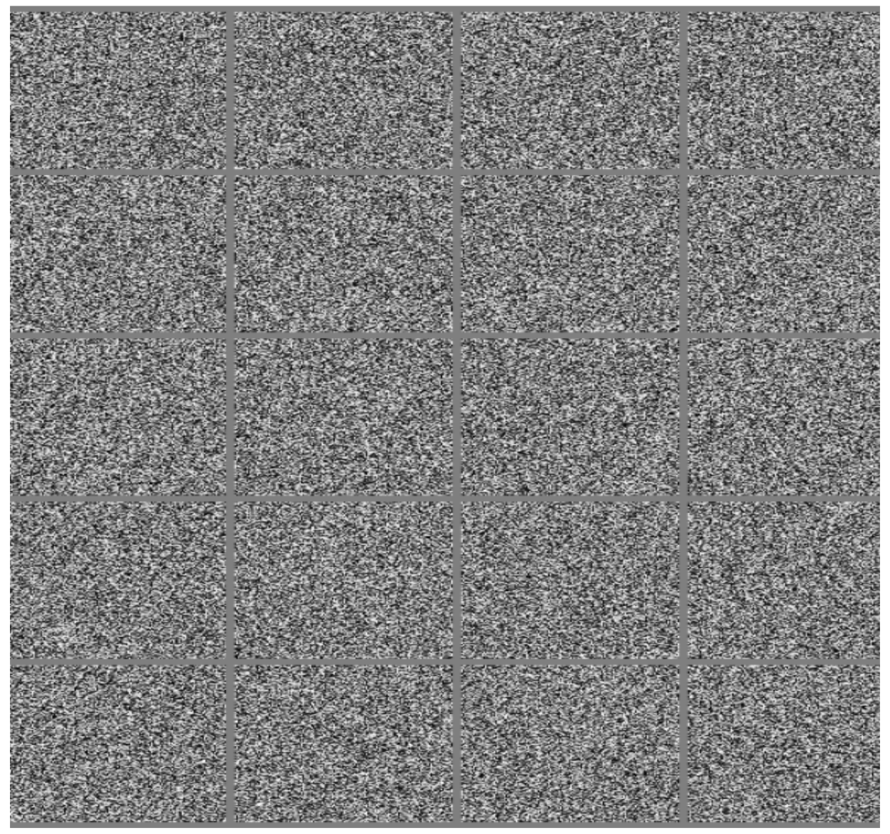


White, 1609.04468

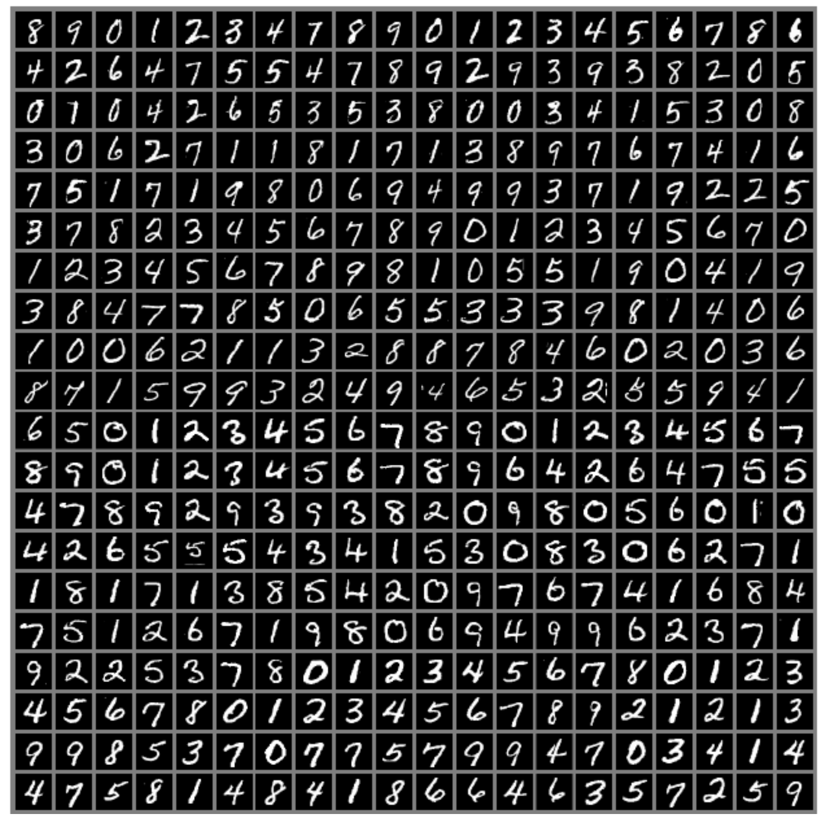
Probabilistic Generative Modeling

$$p(\boldsymbol{x})$$

How to express, learn, and sample from a high-dimensional probability distribution ?



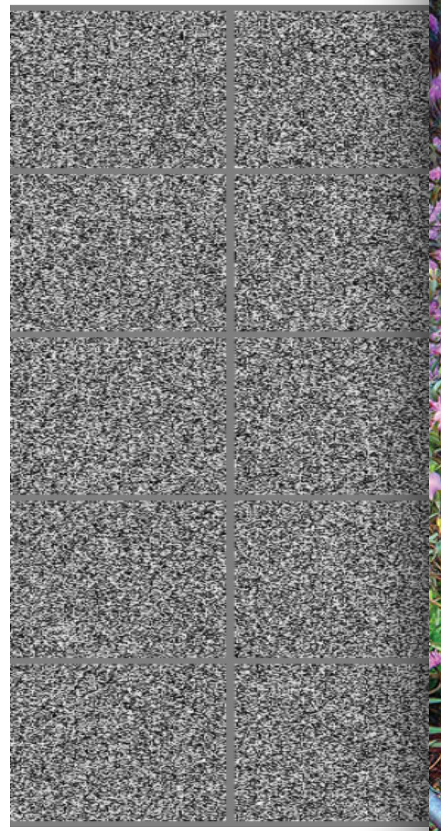
“random” images

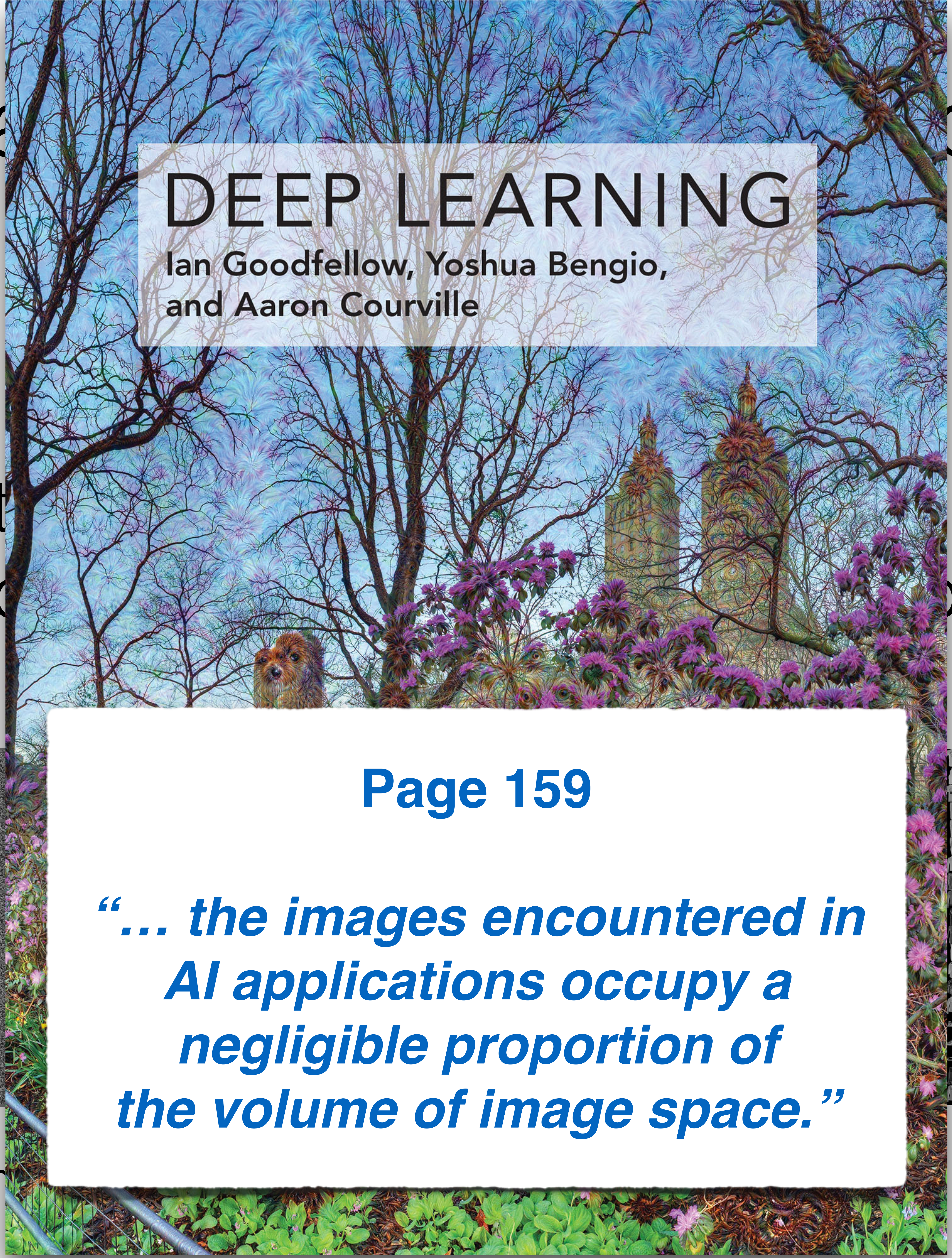


“natural” images

Proba… odeling

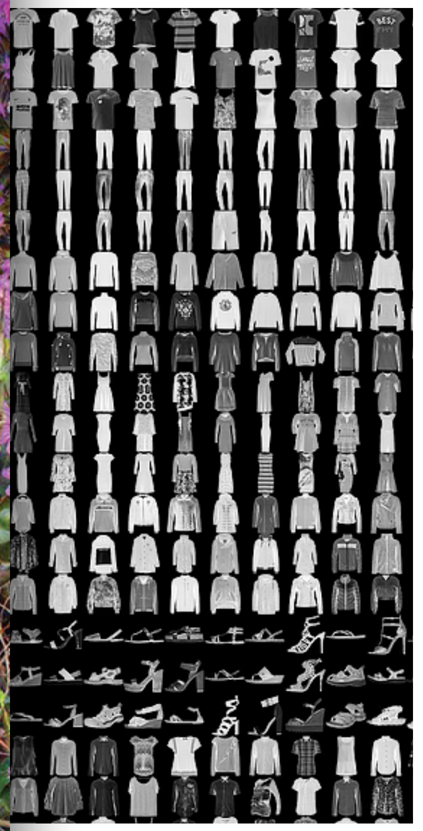
How t… from a
high-… ution ?





Page 159

"… the images encountered in AI applications occupy a negligible proportion of the volume of image space."

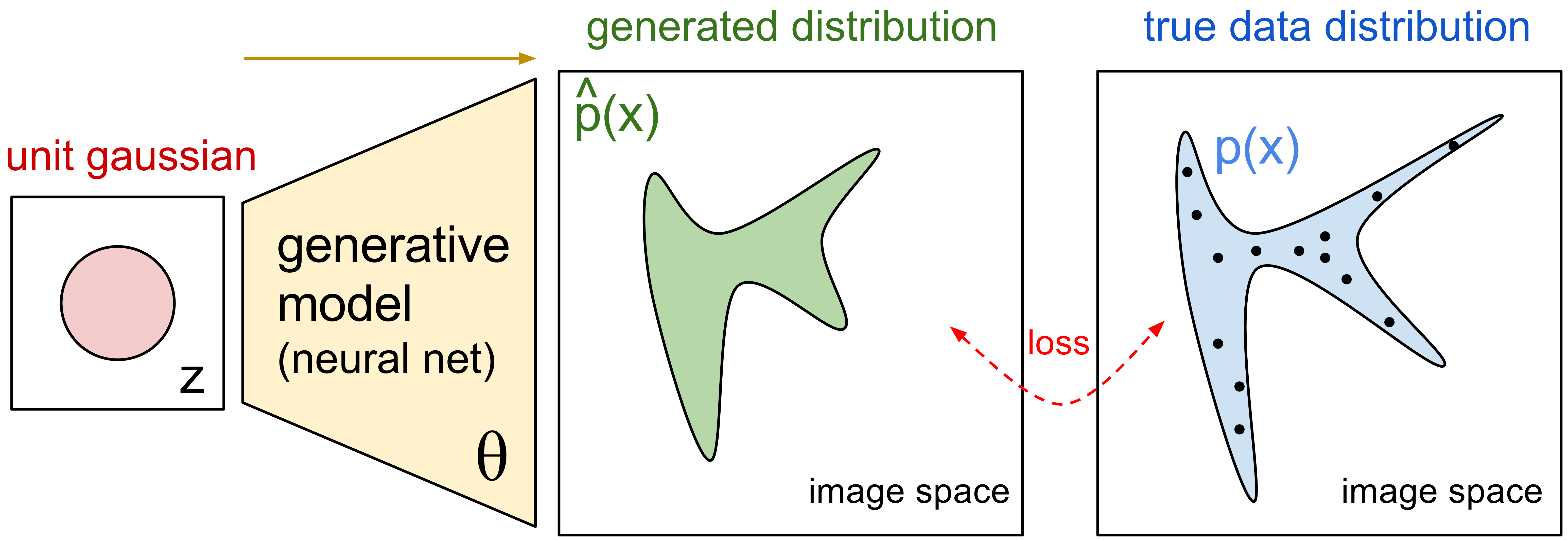


"random…

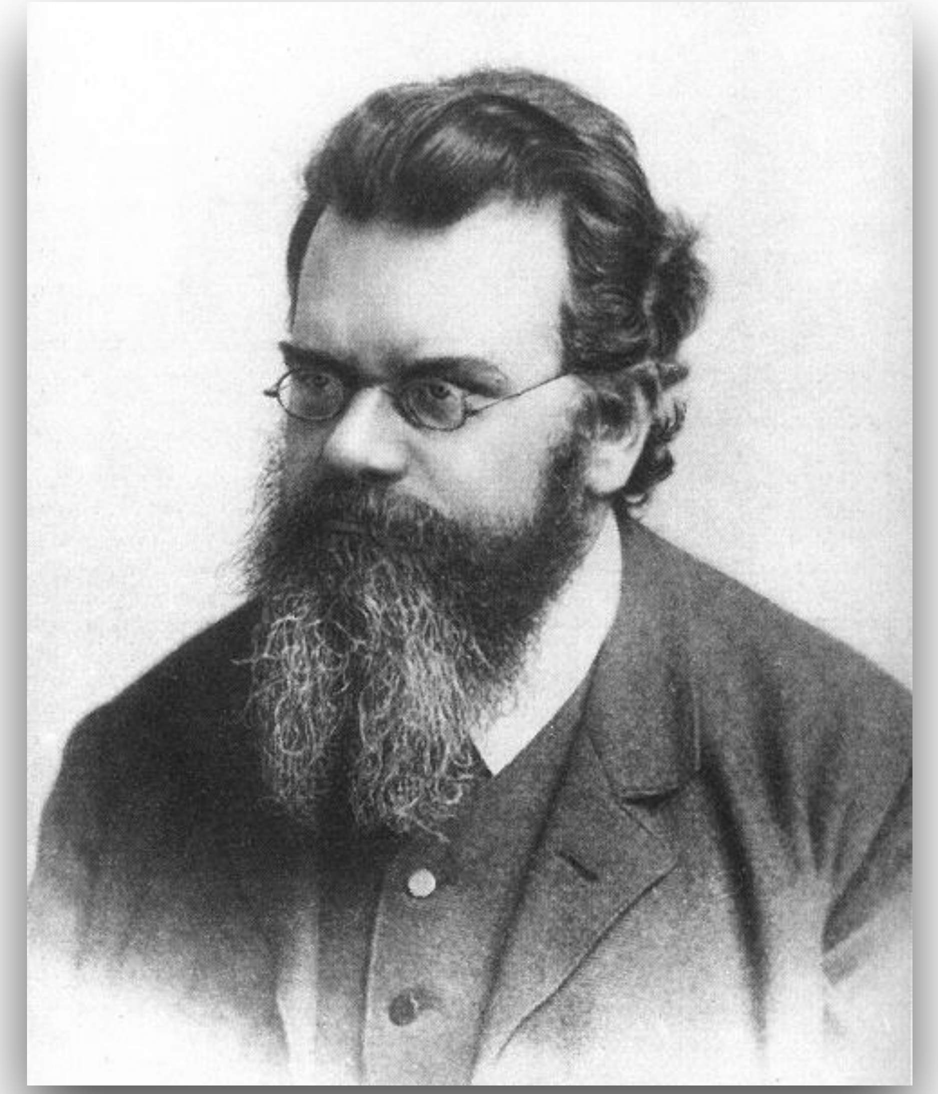
Probabilistic Generative Modeling

$$p(\boldsymbol{x})$$

How to express, learn, and sample from a high-dimensional probability distribution ?



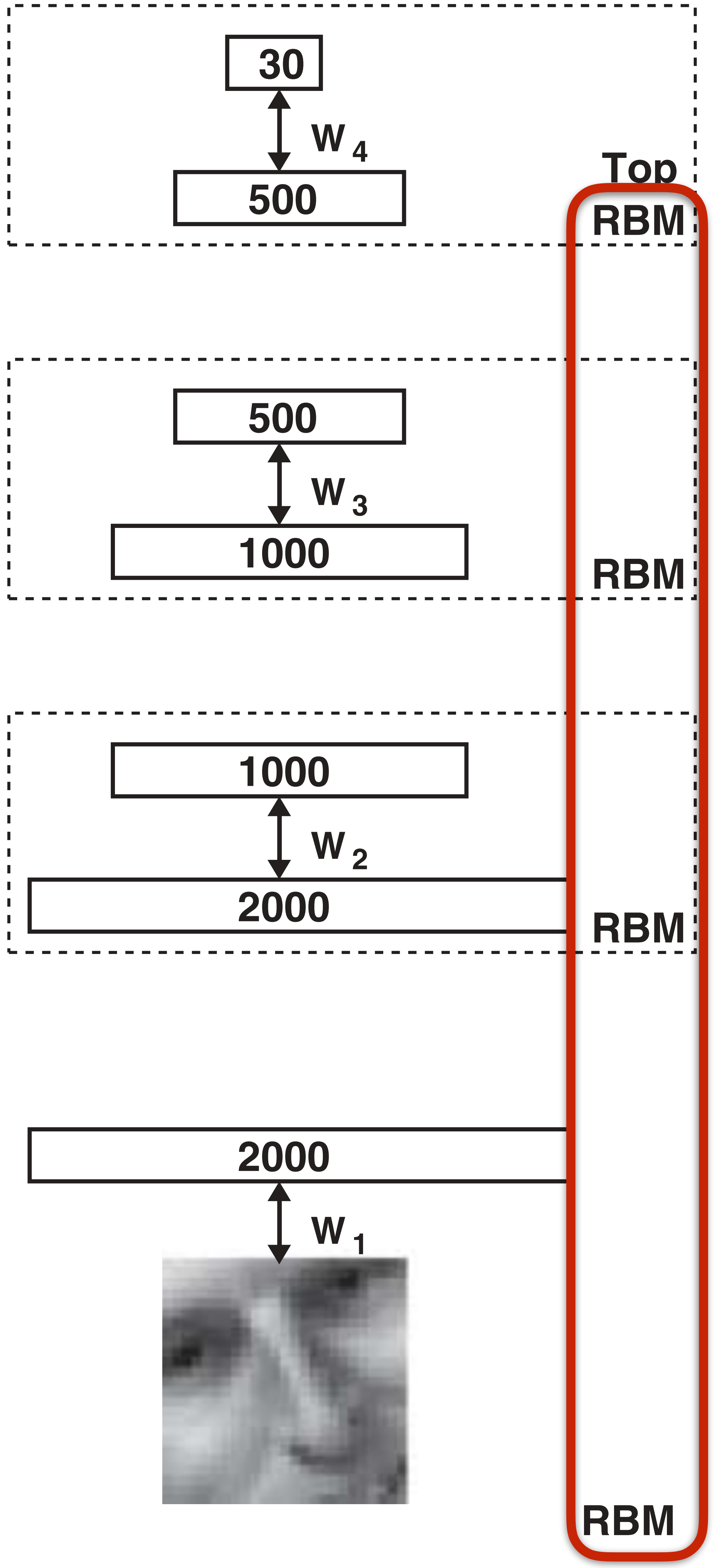
https://blog.openai.com/generative-models/



Boltzmann Machines

$$p(\boldsymbol{x}) = \frac{e^{-E(\boldsymbol{x})}}{Z}$$

statistical physics



Pretraining

Reducing the Dimensionality of Data with Neural Networks

G. E. Hinton* and R. R. Salakhutdinov

High-dimensional data can be converted to low-dimensional codes by training a multilayer neural network with a small central layer to reconstruct high-dimensional input vectors. Gradient descent can be used for fine-tuning the weights in such "autoencoder" networks, but this works well only if the initial weights are close to a good solution. We describe an effective way of initializing the weights that allows deep autoencoder networks to learn low-dimensional codes that work much better than principal components analysis as a tool to reduce the dimensionality of data.

Dimensionality reduction facilitates the classification, visualization, communication, and storage of high-dimensional data. A simple and widely used method is principal components analysis (PCA), which finds the directions of greatest variance in the data set and represents each data point by its coordinates along each of these directions. We describe a nonlinear generalization of PCA that uses an adaptive, multilayer "encoder" network

2006 VOL 313 SCIENCE www.sciencemag.org

Renaissance of deep learning

Feedback to physics

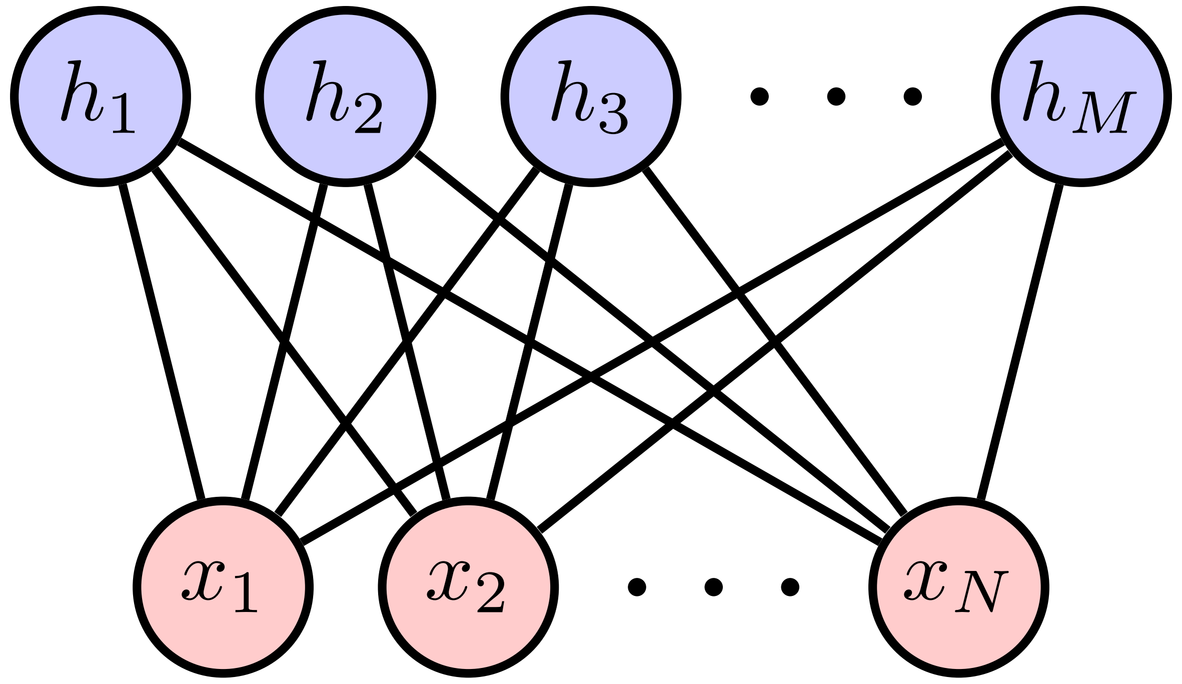
Wavefunctions ansatz
Quantum state tomography

Quantum error correction
Renormalization group…

Generative Modeling using Boltzmann Machines

$$\mathcal{L} = -\frac{1}{|\mathcal{T}|}\sum_{\boldsymbol{x}\in\mathcal{T}} \ln p\left(\boldsymbol{x}\right)$$ Negative log-likelihood loss

Learn

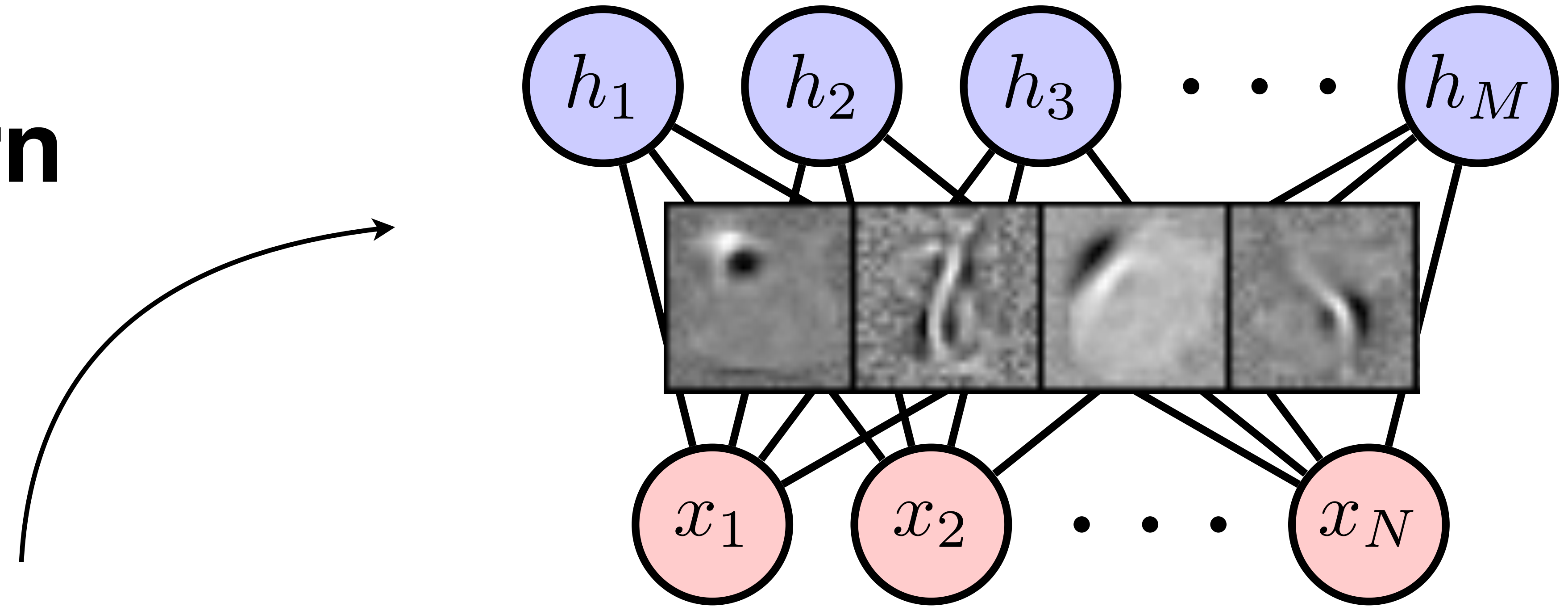


$$\nabla\mathcal{L} = \langle\nabla E\rangle_{\text{data}} - \langle\nabla E\rangle_{\text{model}}$$

Generative Modeling using Boltzmann Machines

$$\mathcal{L} = -\frac{1}{|\mathcal{T}|} \sum_{\boldsymbol{x} \in \mathcal{T}} \ln p(\boldsymbol{x})$$ Negative log-likelihood loss

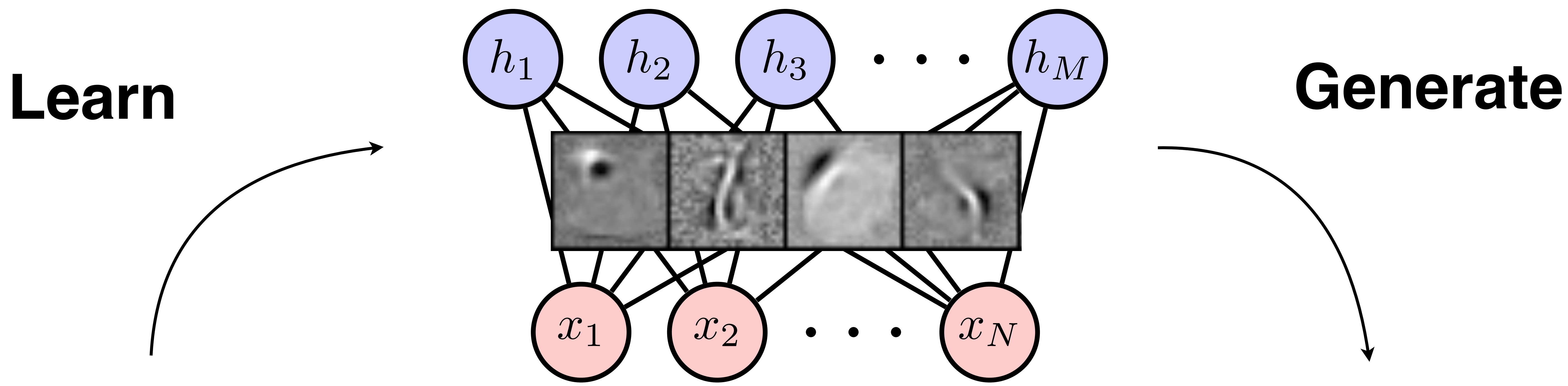
Learn



$$\nabla \mathcal{L} = \langle \nabla E \rangle_{\text{data}} - \langle \nabla E \rangle_{\text{model}}$$

Generative Modeling using Boltzmann Machines

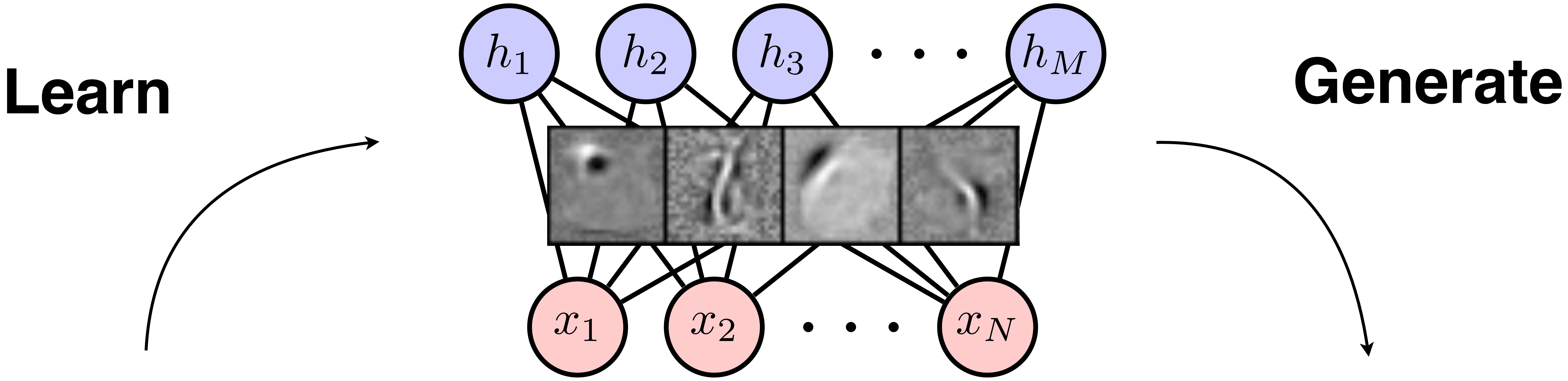
$$\mathcal{L} = -\frac{1}{|\mathcal{T}|}\sum_{\boldsymbol{x}\in\mathcal{T}} \ln p(\boldsymbol{x})$$ Negative log-likelihood loss



$$\nabla\mathcal{L} = \langle\nabla E\rangle_{\text{data}} - \langle\nabla E\rangle_{\text{model}}$$

Generative Modeling using Boltzmann Machines

$$\mathcal{L} = -\frac{1}{|\mathcal{T}|} \sum_{\boldsymbol{x} \in \mathcal{T}} \ln p(\boldsymbol{x})$$ Negative log-likelihood loss

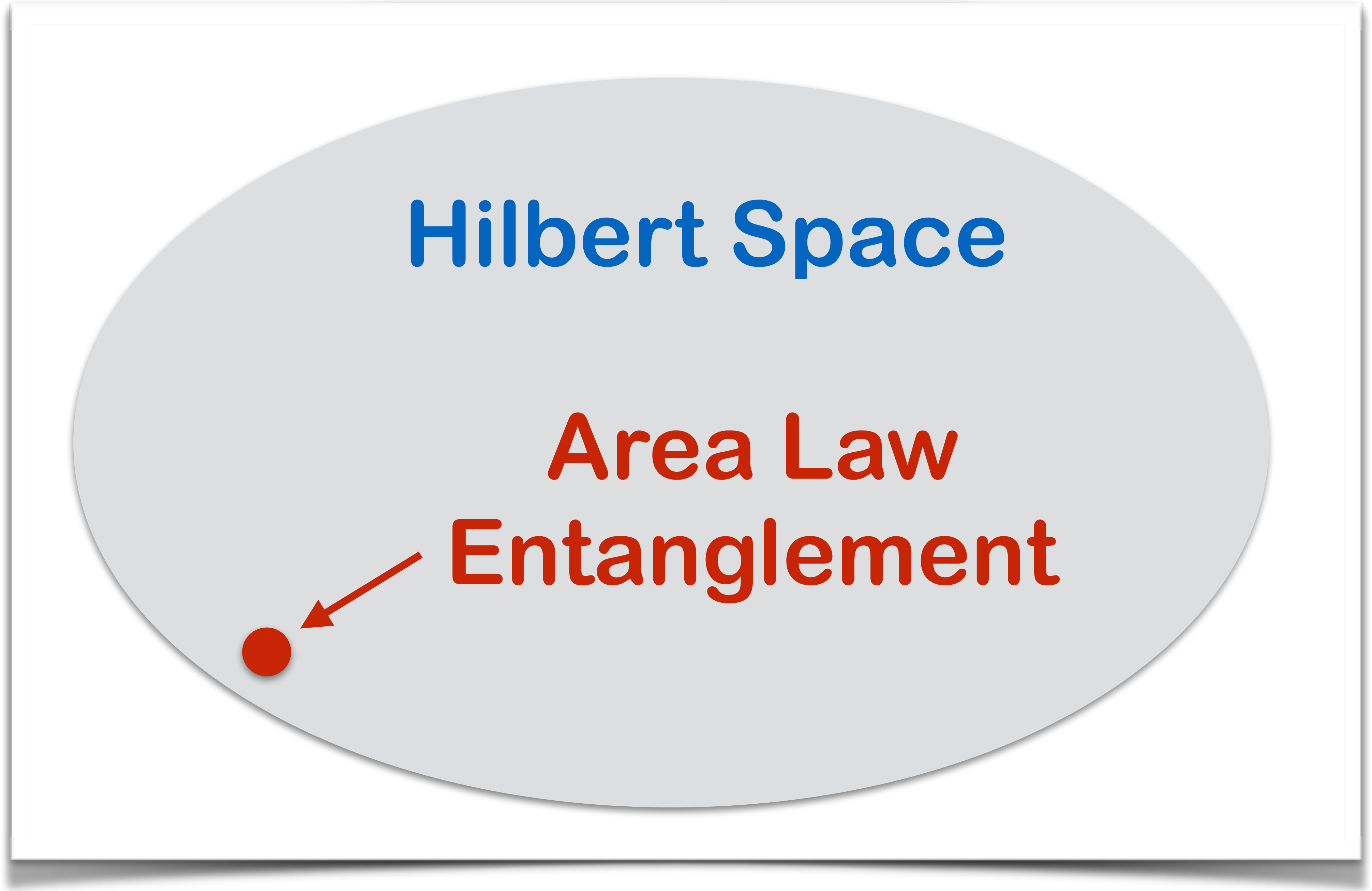


$$\nabla \mathcal{L} = \langle \nabla E \rangle_{\text{data}} - \langle \nabla E \rangle_{\text{model}}$$

Born Machines

$$p(\boldsymbol{x}) = \frac{|\Psi(\boldsymbol{x})|^2}{\mathcal{Z}}$$

quantum physics

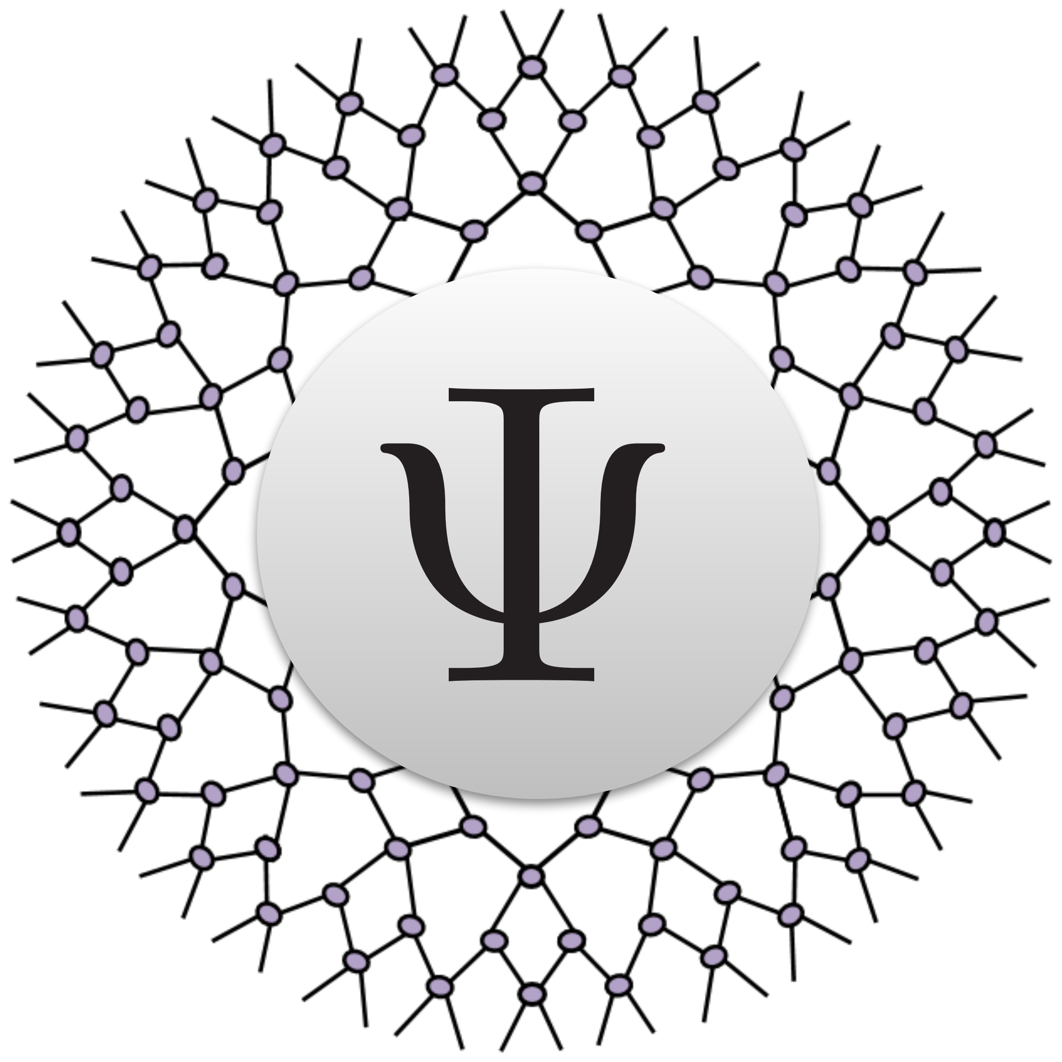


Born Machines

$$p(\boldsymbol{x}) = \frac{|\Psi(\boldsymbol{x})|^2}{\mathcal{Z}}$$

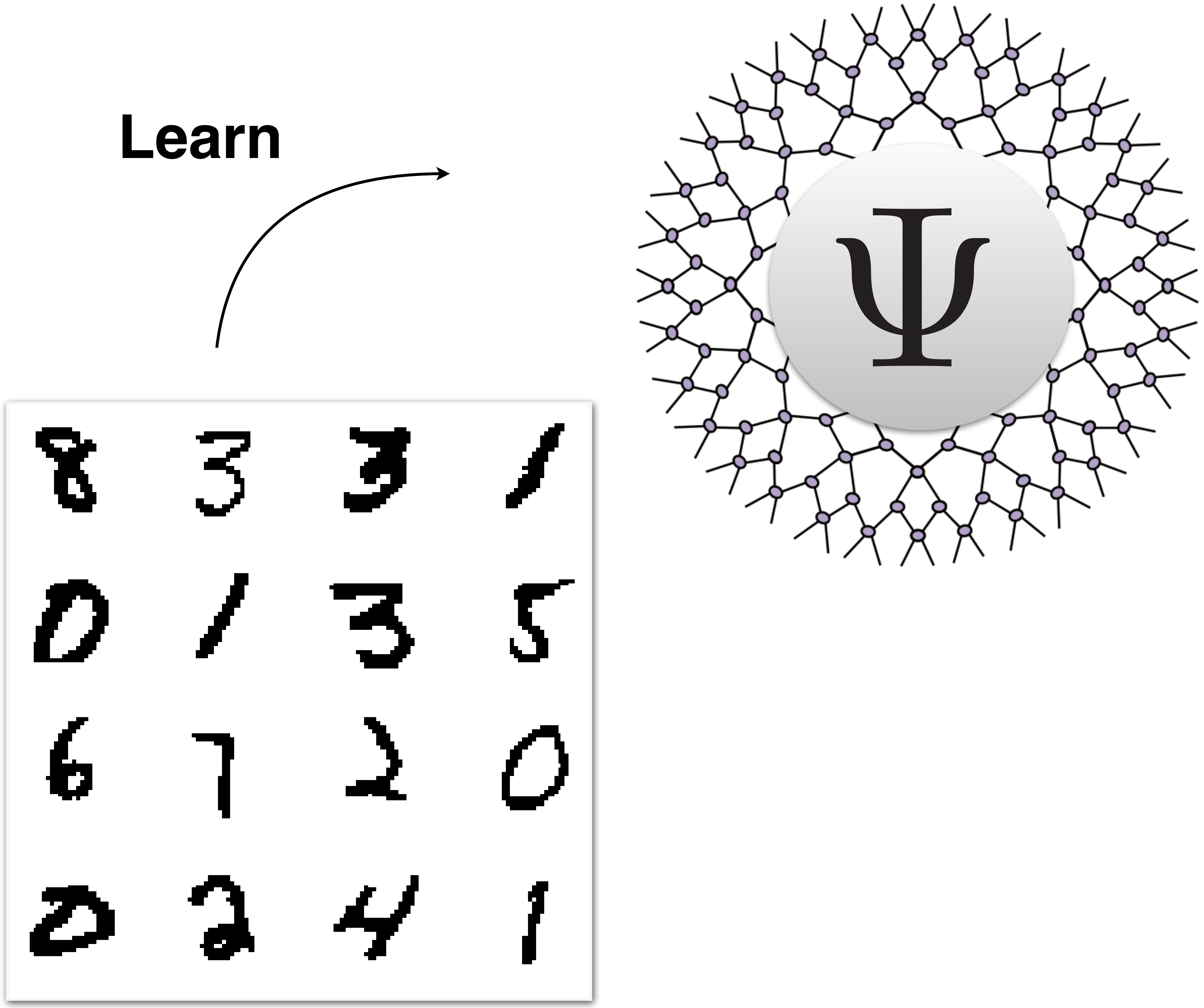
quantum physics

Quantum inspired generative modeling



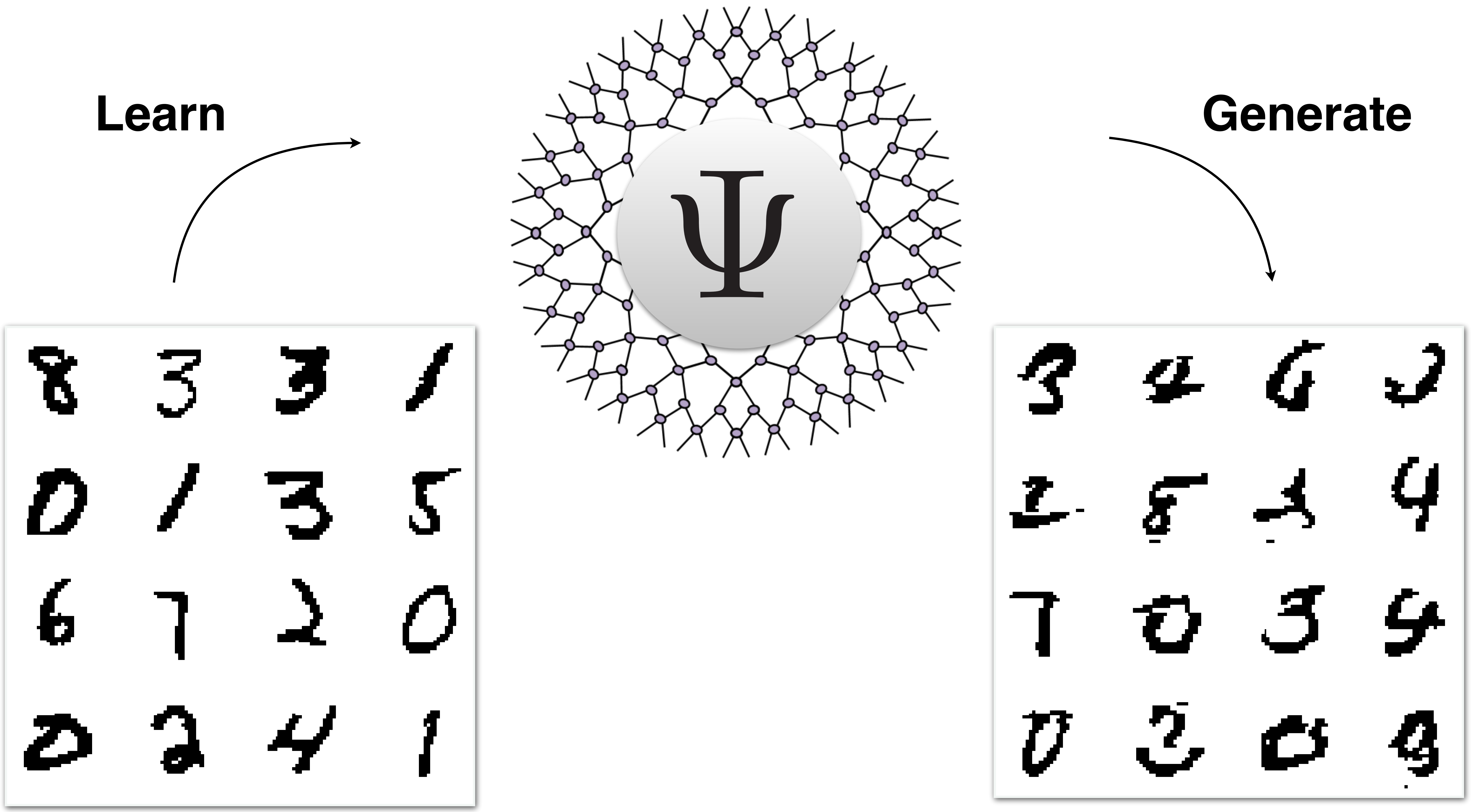
Han, Wang, Fan, LW, Zhang, 1709.01662, PRX in press

Quantum inspired generative modeling



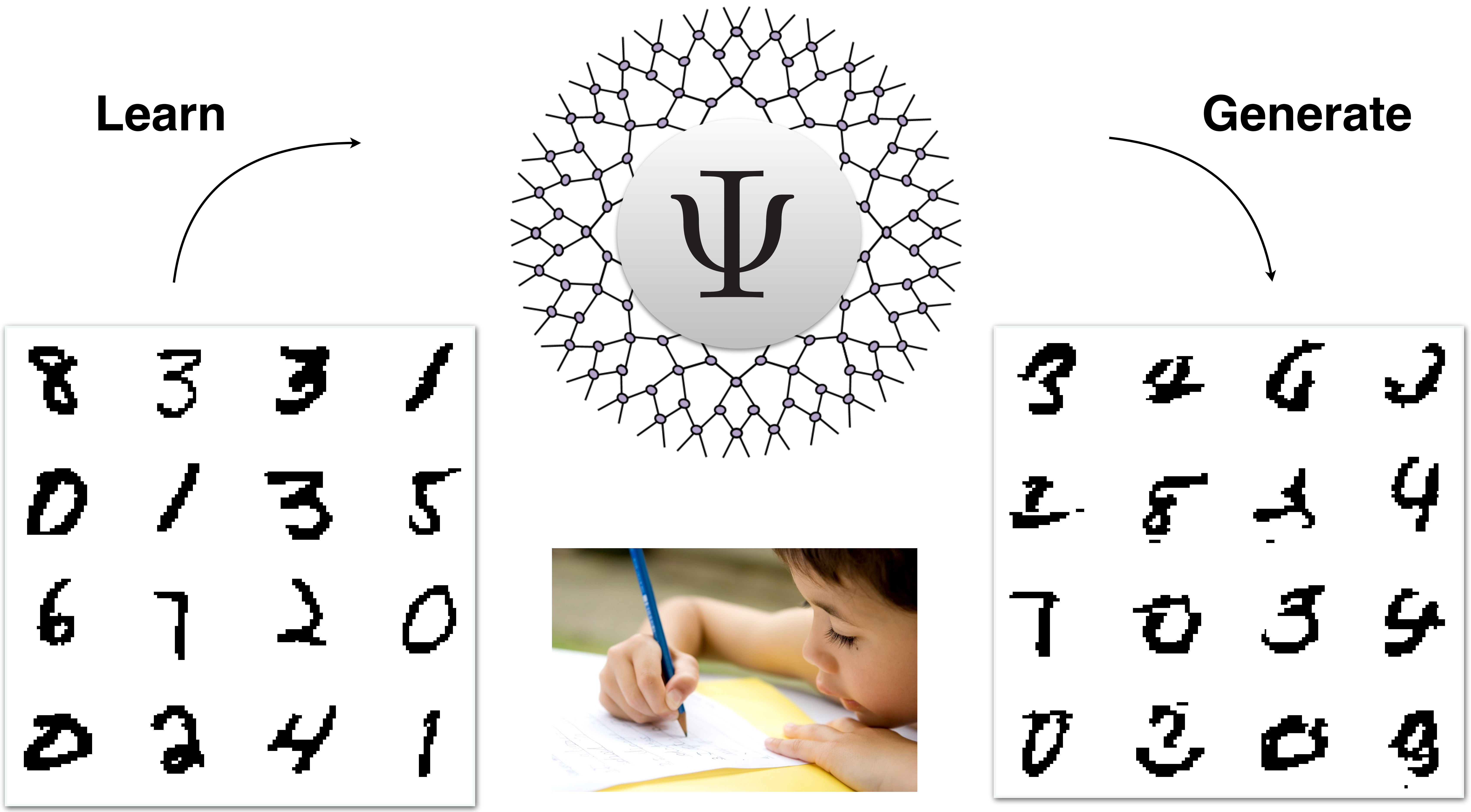
Han, Wang, Fan, LW, Zhang, 1709.01662, PRX in press

Quantum inspired generative modeling



Han, Wang, Fan, LW, Zhang, 1709.01662, PRX in press

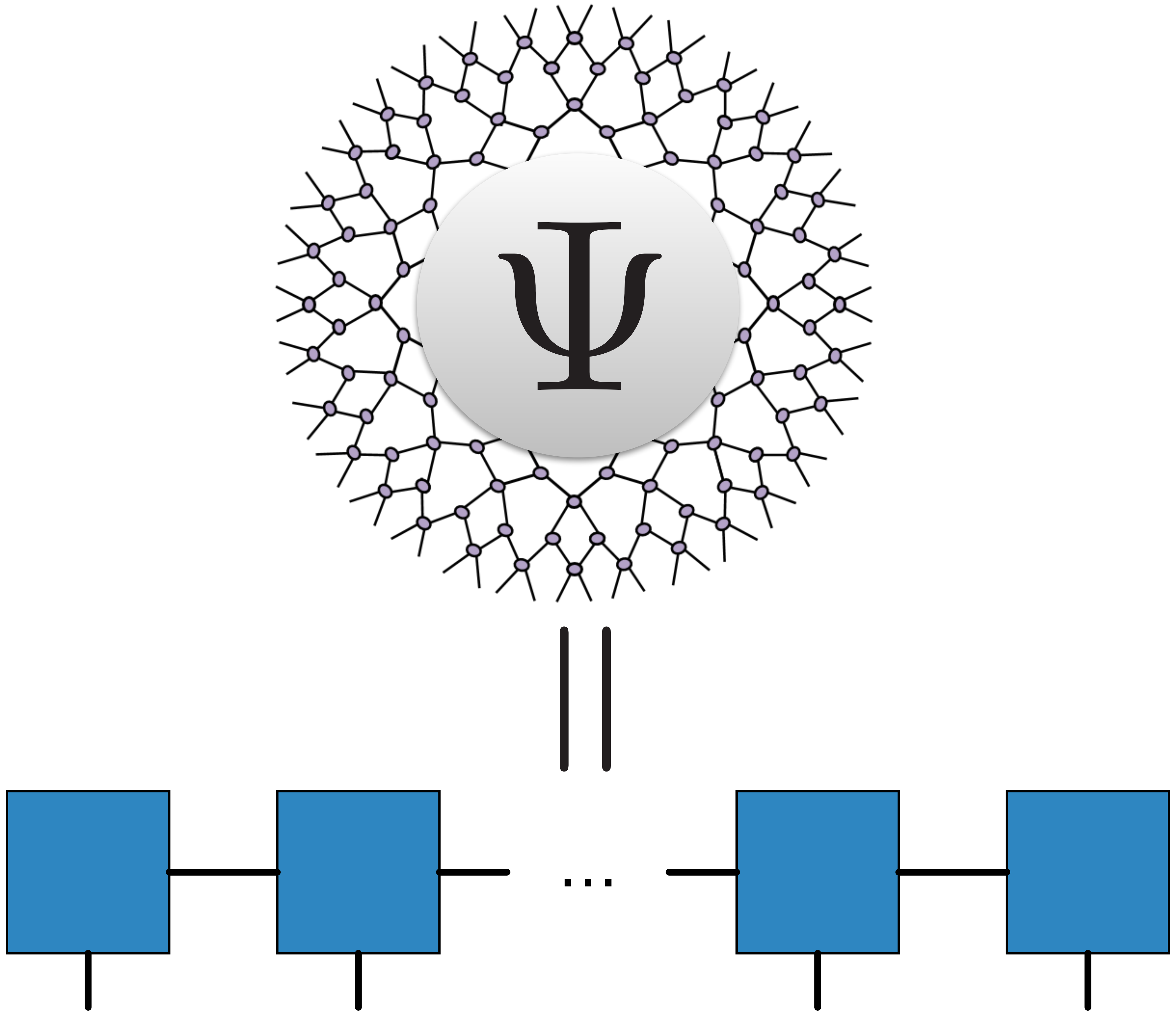
Quantum inspired generative modeling



"Teach a quantum state to write digits"

Han, Wang, Fan, LW, Zhang, 1709.01662, PRX in press

Generative modeling using Tensor Network States



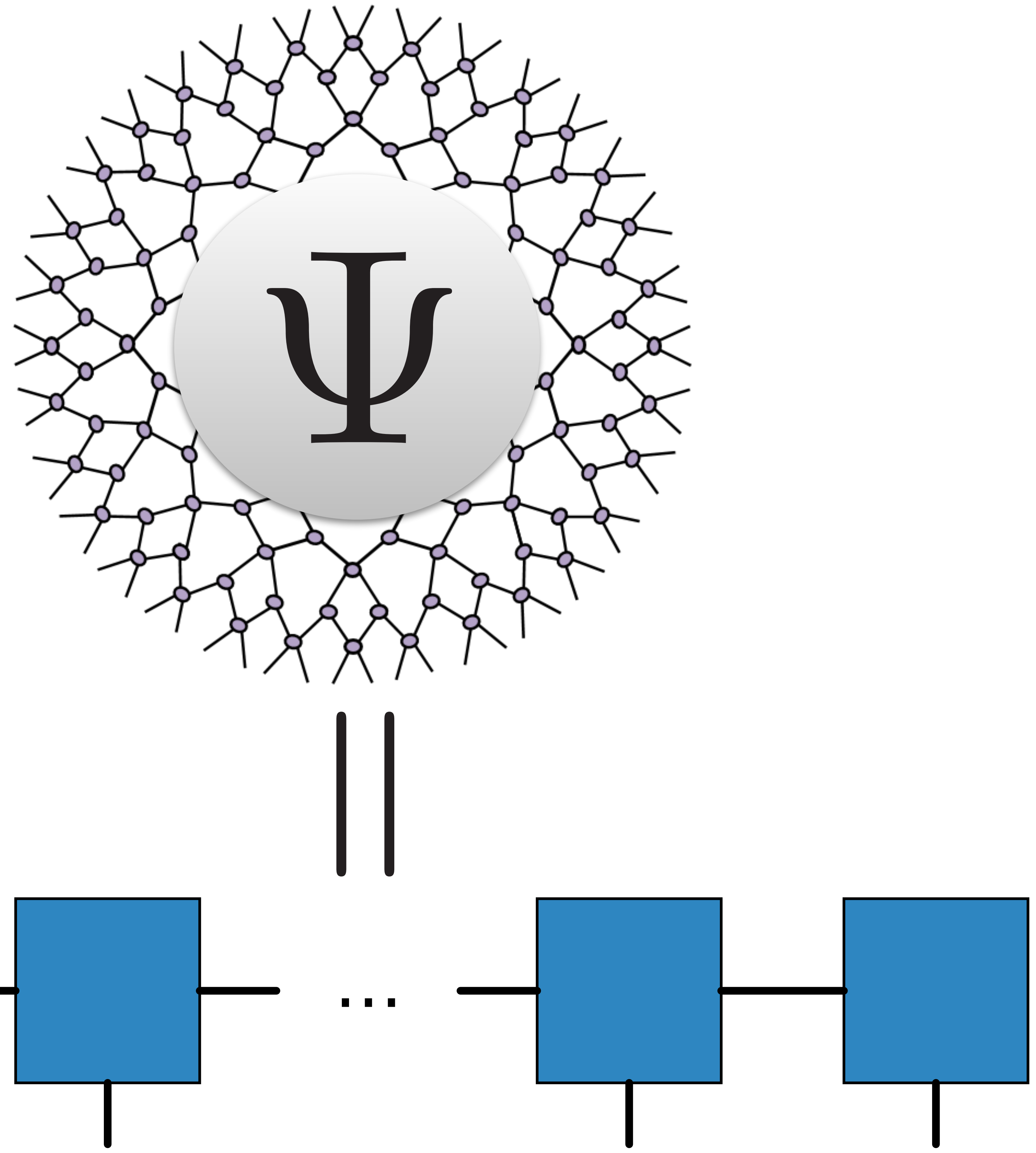
Stoudenmire, Schwab NIPS 2016
Stoudenmire Q. Sci. Tech. 2018

Liu et al 1710.04833
Liu et al 1803.09111

Hallam et al 1711.03357
Glasser et al 1806.05964

Gallego, Orus 1708.01525
Pestun et al 1711.01416

Generative modeling using Tensor Network States



Overview talk by Miles on 29th

Stoudenmire, Schwab N

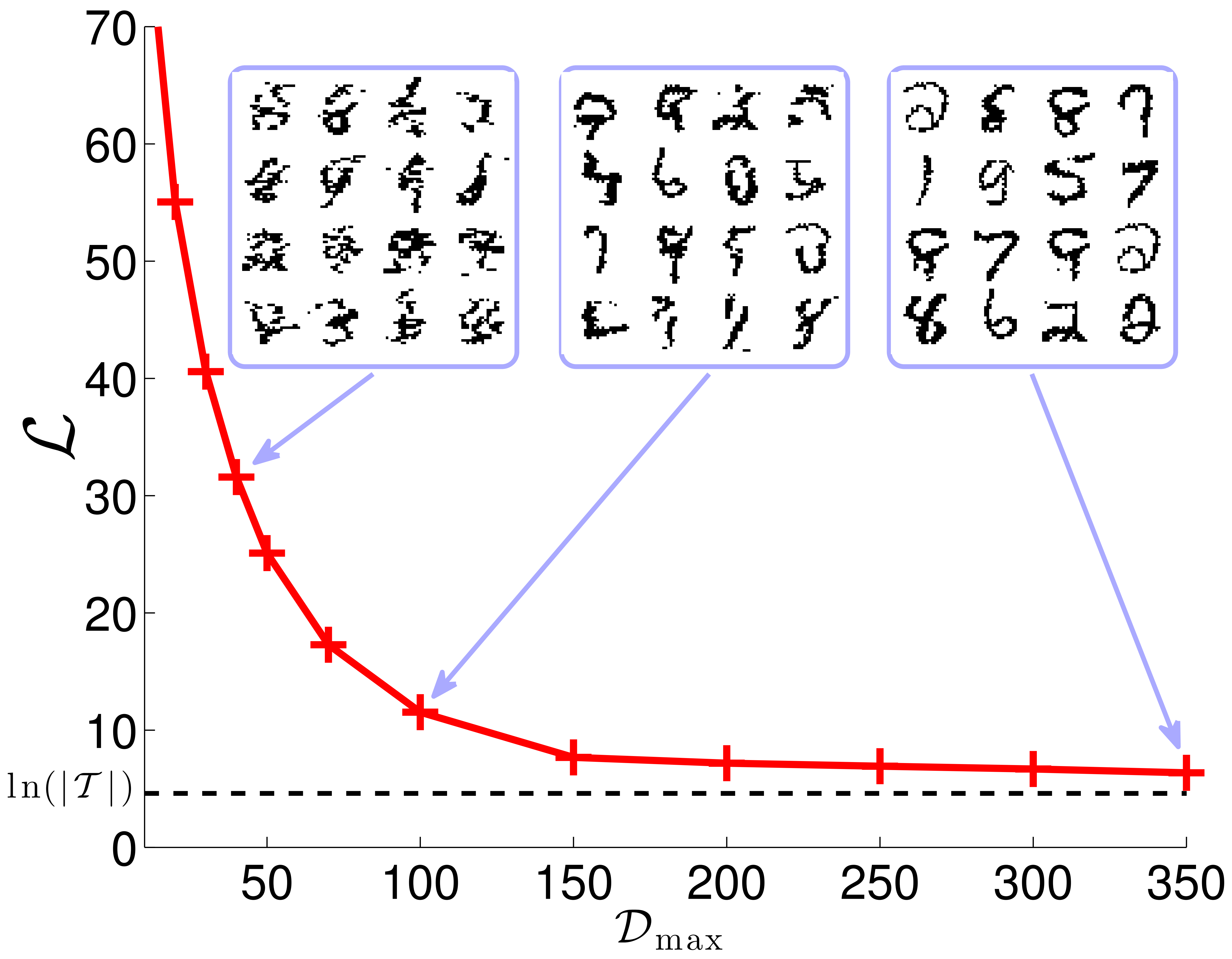
Stoudenmire Q. Sci. Tec

go, Orus 1708.01525

stun et al 1711.01416

What does it learn ?

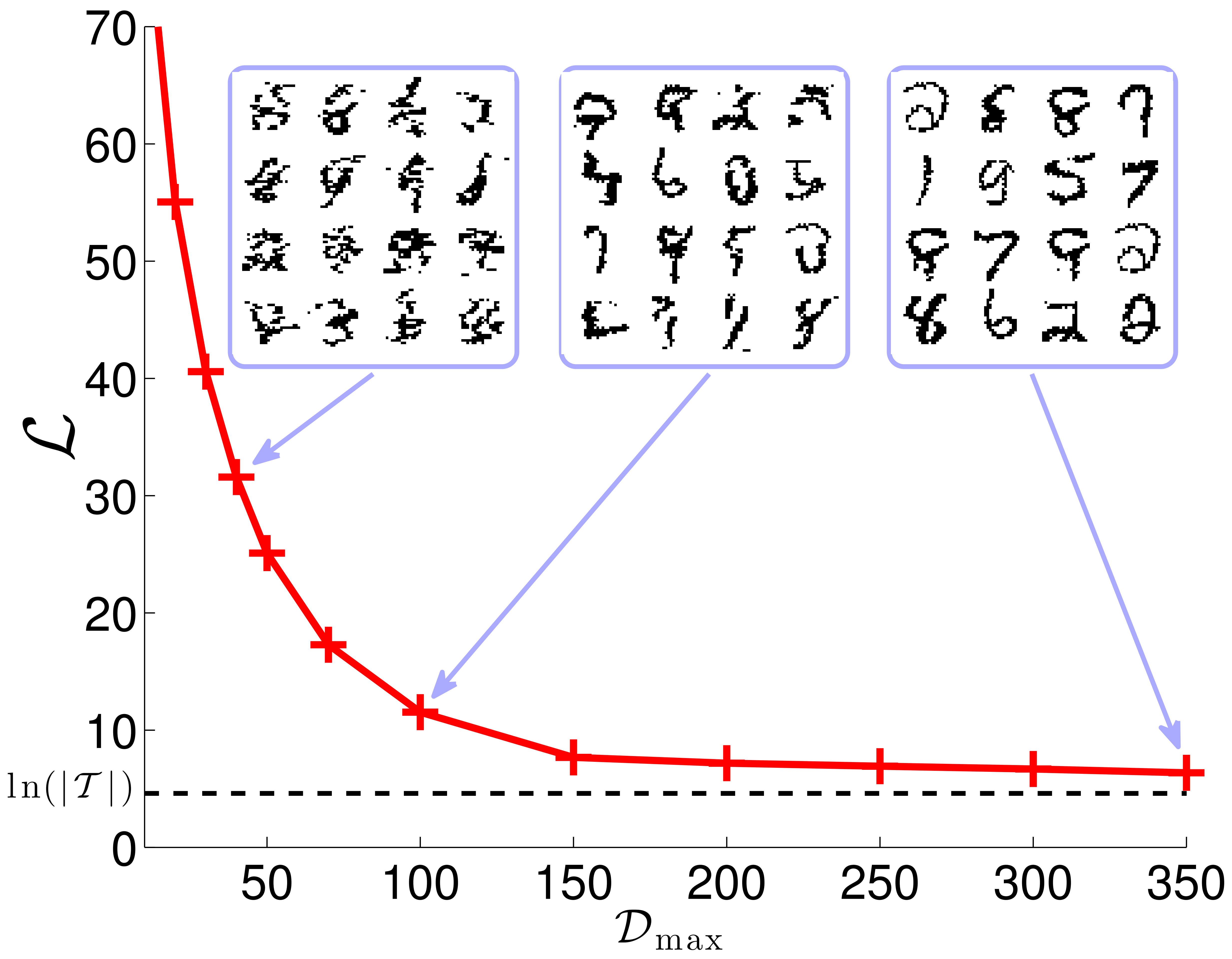
$$\mathcal{L} = -\frac{1}{|\mathcal{T}|}\sum_{\boldsymbol{x}\in\mathcal{T}} \ln p(\boldsymbol{x})$$

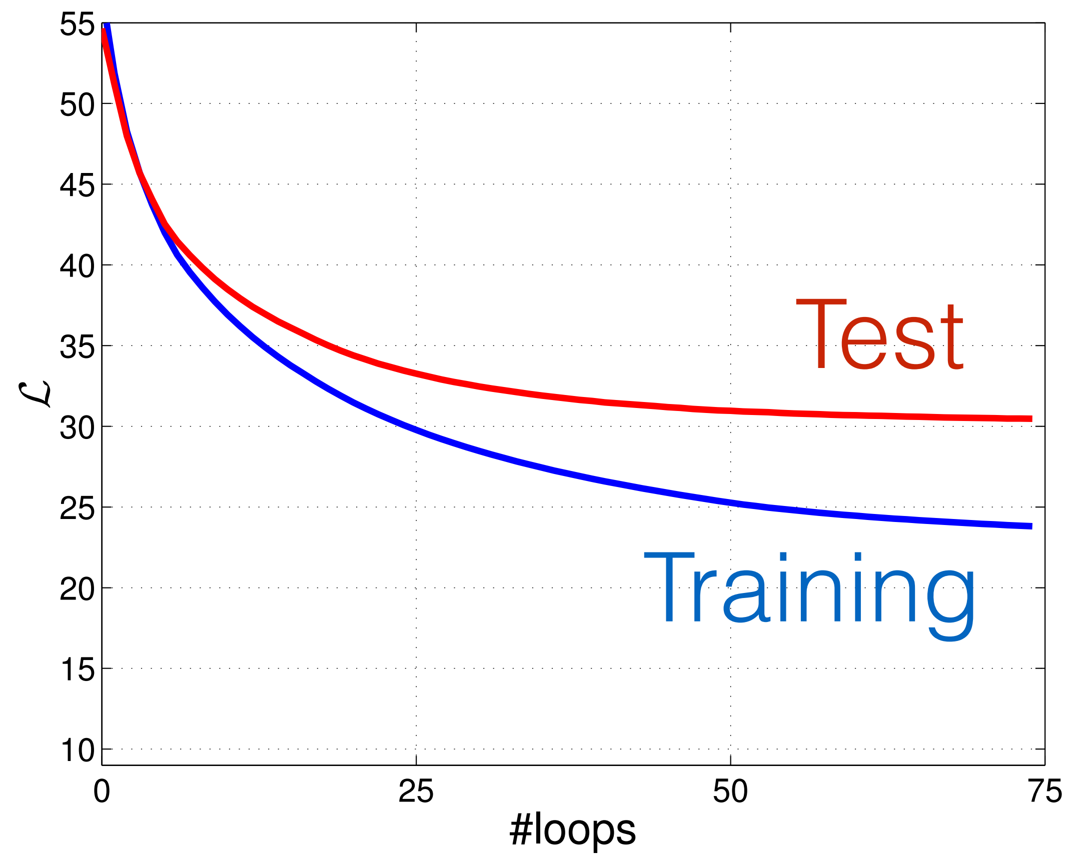


Captures longer range correlations with larger bond dimensions

What does it learn ?

$$\mathcal{L} = -\frac{1}{|\mathcal{T}|}\sum_{x \in \mathcal{T}} \ln p(x)$$





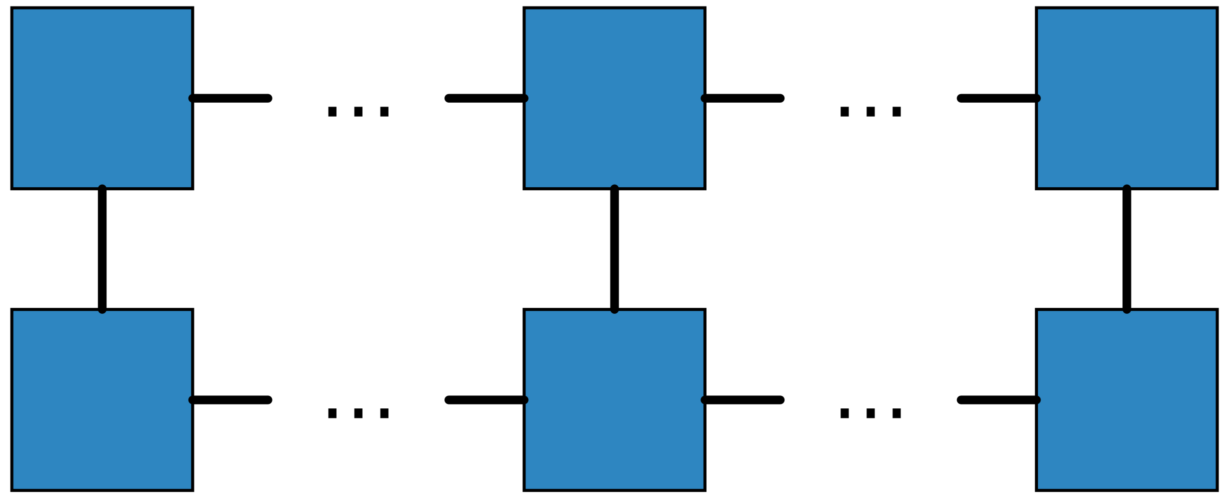
Captures longer range correlations with larger bond dimensions

Why bother ?

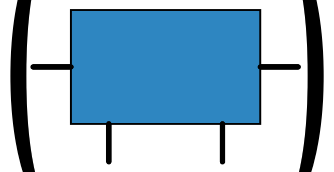
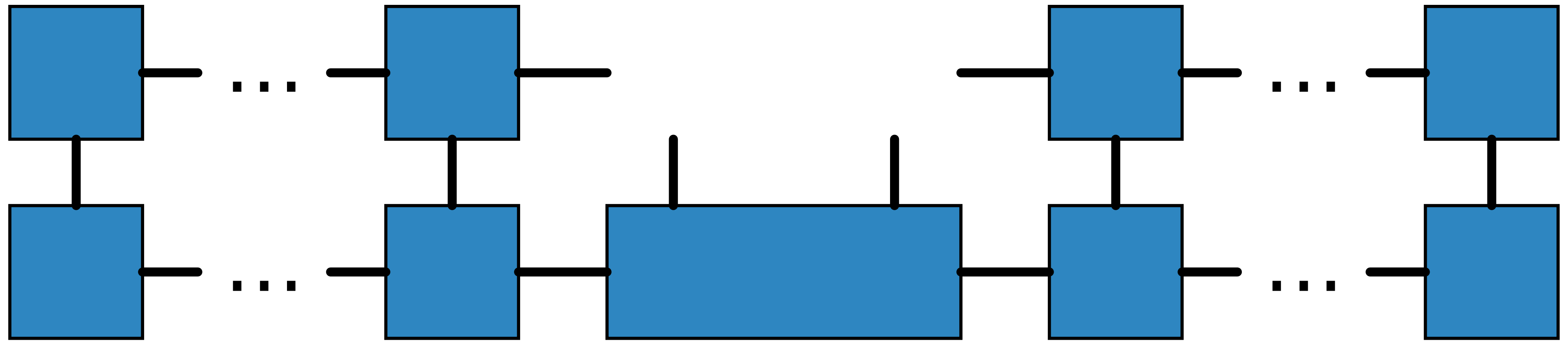
Representability | **Learning** | **Inference** | **Sampling**

Glasser, Clark, Deng,
Gao, Chen, Huang… 2017

Feature-I: Tractable Likelihood

$\mathcal{Z} =$ 

tractable via
efficient tensor contraction

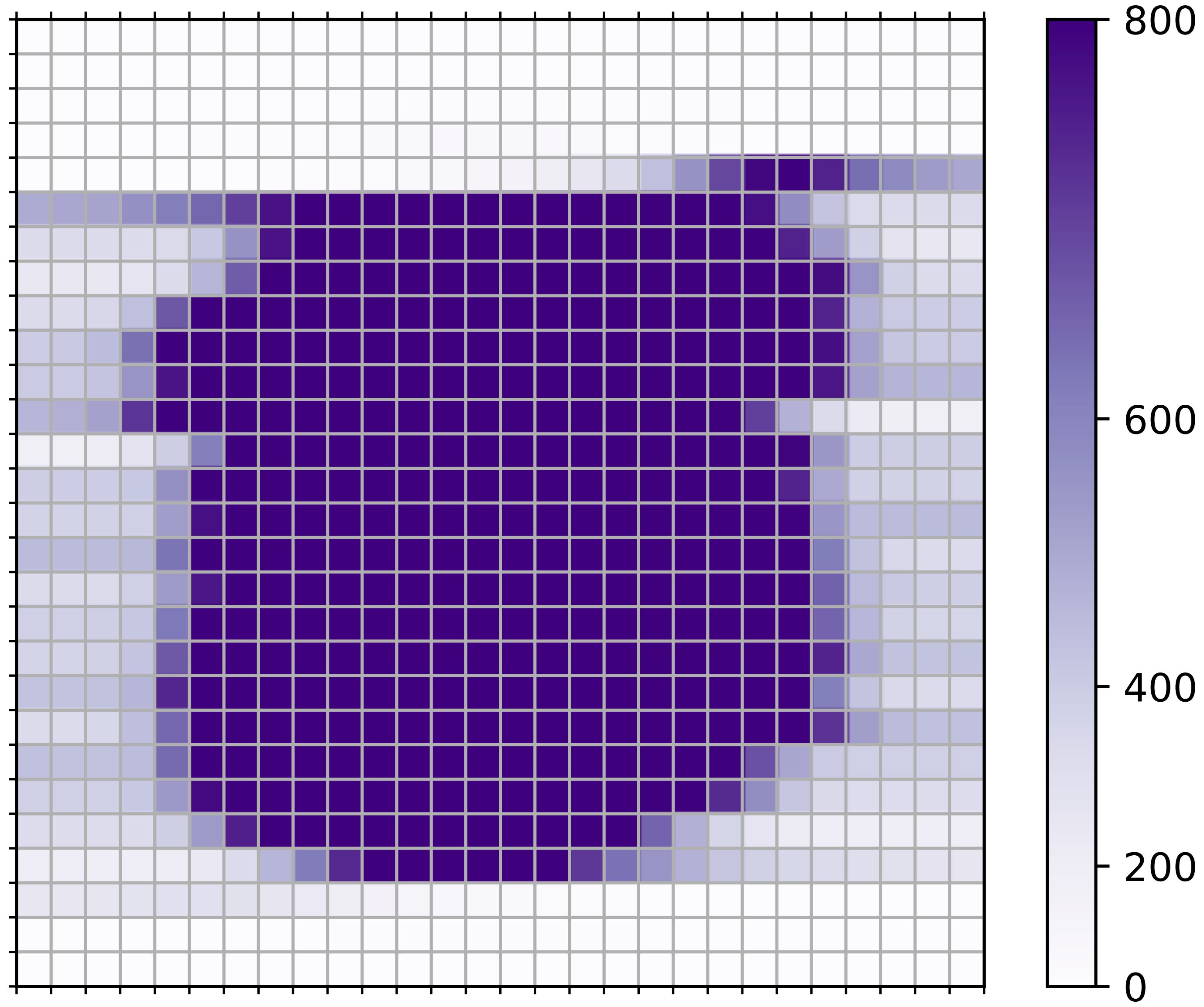
$\partial \mathcal{Z} / \partial \Big($  $\Big) = 2 \times$ 

Efficient & Unbiased learning compared to models with intractable partition functions

Feature-II: Adaptive Learning

Training images

Bond dimensions



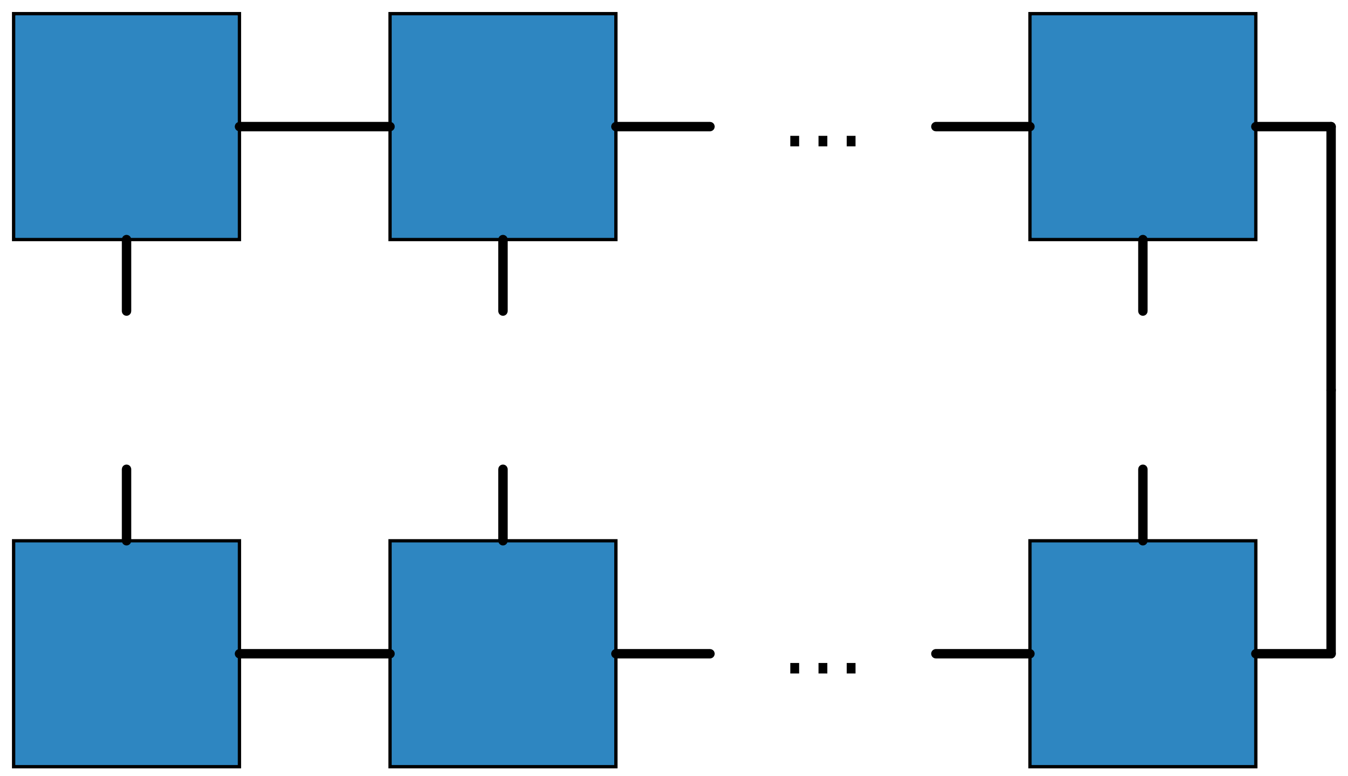
Adaptively grows the bond dimensions, thus
dynamically tuning the expressibility instead of fixed the # of params

Feature-III: Direct Generation

$$p(\boldsymbol{x}) = \prod_i \frac{p(\boldsymbol{x}_{<i+1})}{p(\boldsymbol{x}_{<i})} = \prod_i p(x_i|\boldsymbol{x}_{<i})$$

Ferris & Vidal 2012

$$p(\boldsymbol{x}_{<i}) =$$

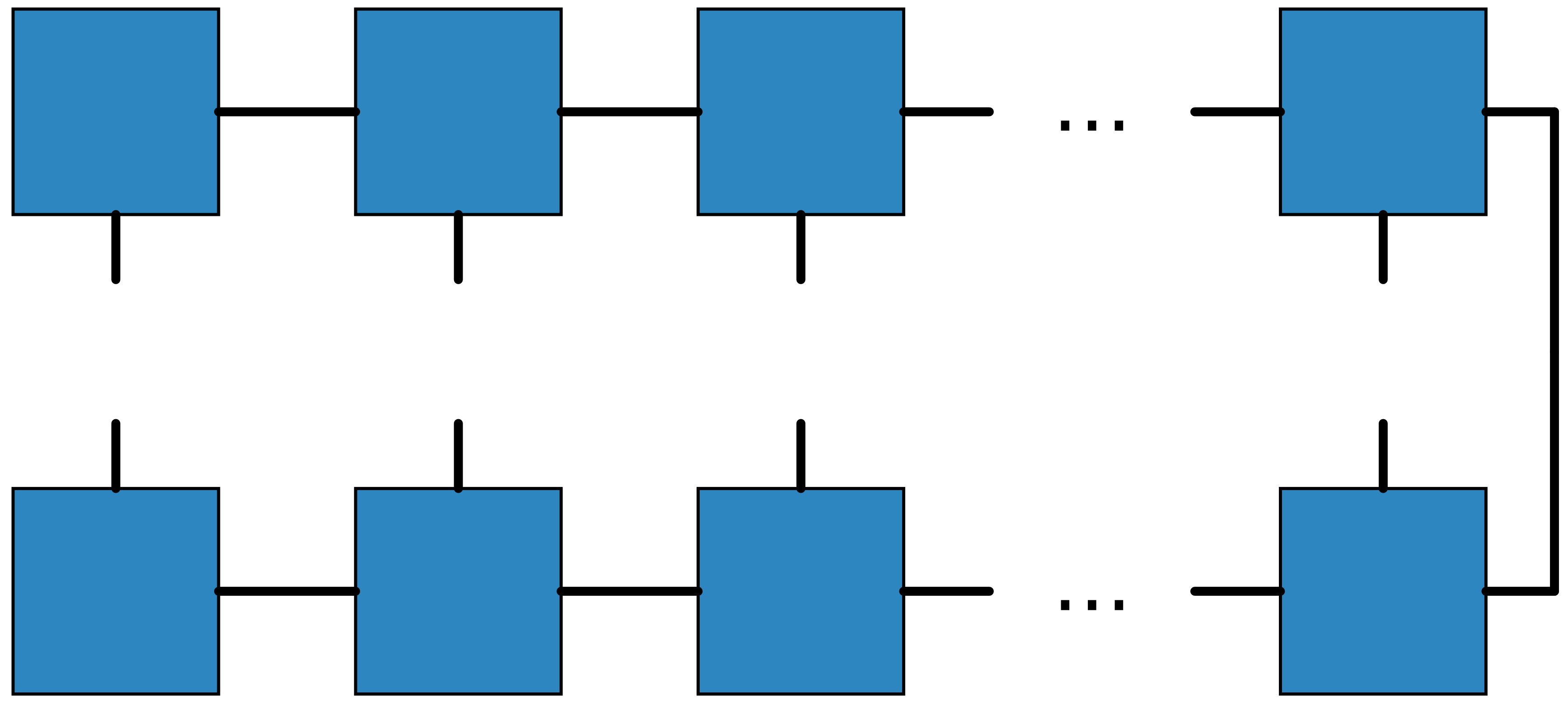


No thermalization issue compared to
slow mixing Gibbs sampling of Boltzmann Machines

Feature-III: Direct Generation

$$p(\boldsymbol{x}) = \prod_i \frac{p(\boldsymbol{x}_{<i+1})}{p(\boldsymbol{x}_{<i})} = \prod_i p(x_i|\boldsymbol{x}_{<i})$$

Ferris & Vidal 2012

$p(\boldsymbol{x}_{<i+1}) =$ 

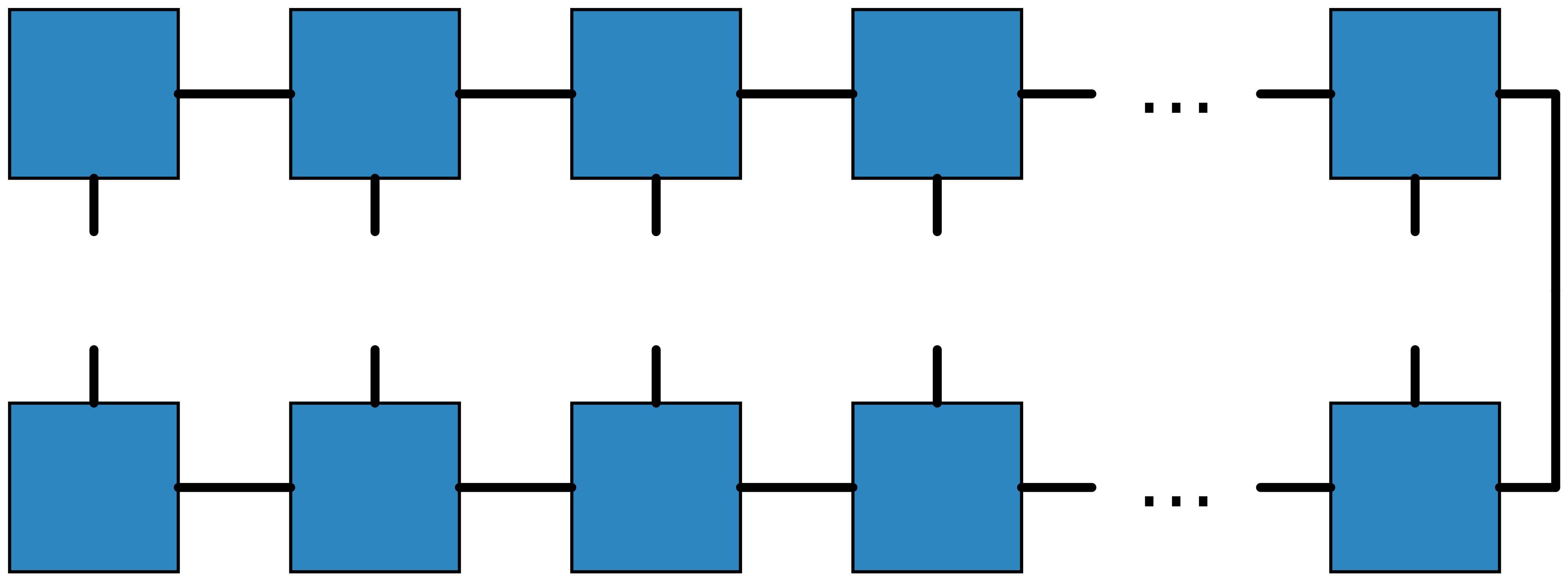
No thermalization issue compared to
slow mixing Gibbs sampling of Boltzmann Machines

Feature-III: Direct Generation

$$p(\boldsymbol{x}) = \prod_i \frac{p(\boldsymbol{x}_{<i+1})}{p(\boldsymbol{x}_{<i})} = \prod_i p(x_i|\boldsymbol{x}_{<i})$$

Ferris & Vidal 2012

$$p(\boldsymbol{x}_{<i+2}) =$$



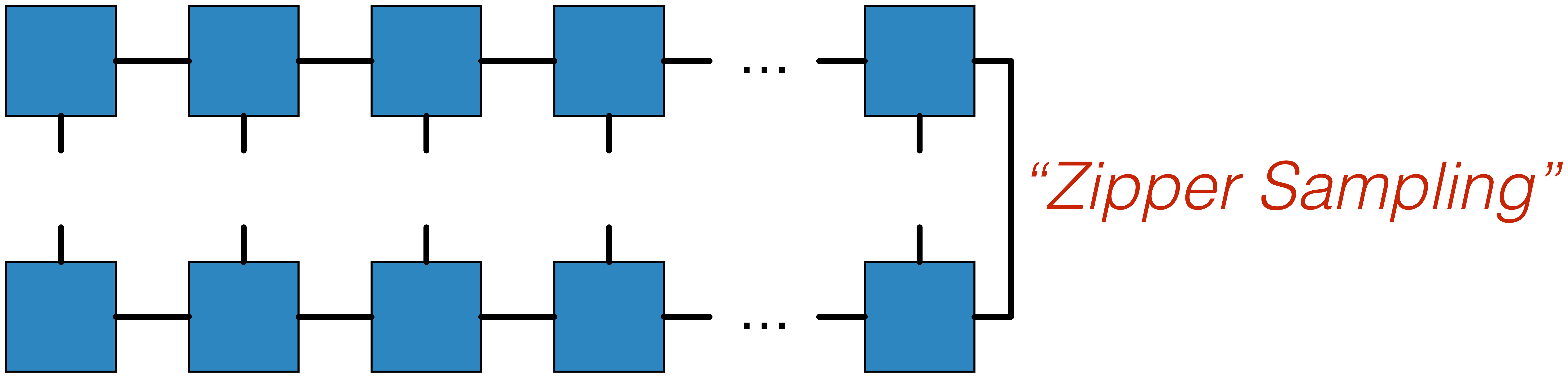
No thermalization issue compared to
slow mixing Gibbs sampling of Boltzmann Machines

Feature-III: Direct Generation

$$p(\boldsymbol{x}) = \prod_i \frac{p(\boldsymbol{x}_{<i+1})}{p(\boldsymbol{x}_{<i})} = \prod_i p(x_i|\boldsymbol{x}_{<i})$$

Ferris & Vidal 2012

$p(\boldsymbol{x}_{<i+2}) =$



No thermalization issue compared to
slow mixing Gibbs sampling of Boltzmann Machines

These advantages hold true for Tree tensor networks and MERA

Image Restoration

Han, Wang, Fan, LW, Zhang, 1709.01662, PRX in press

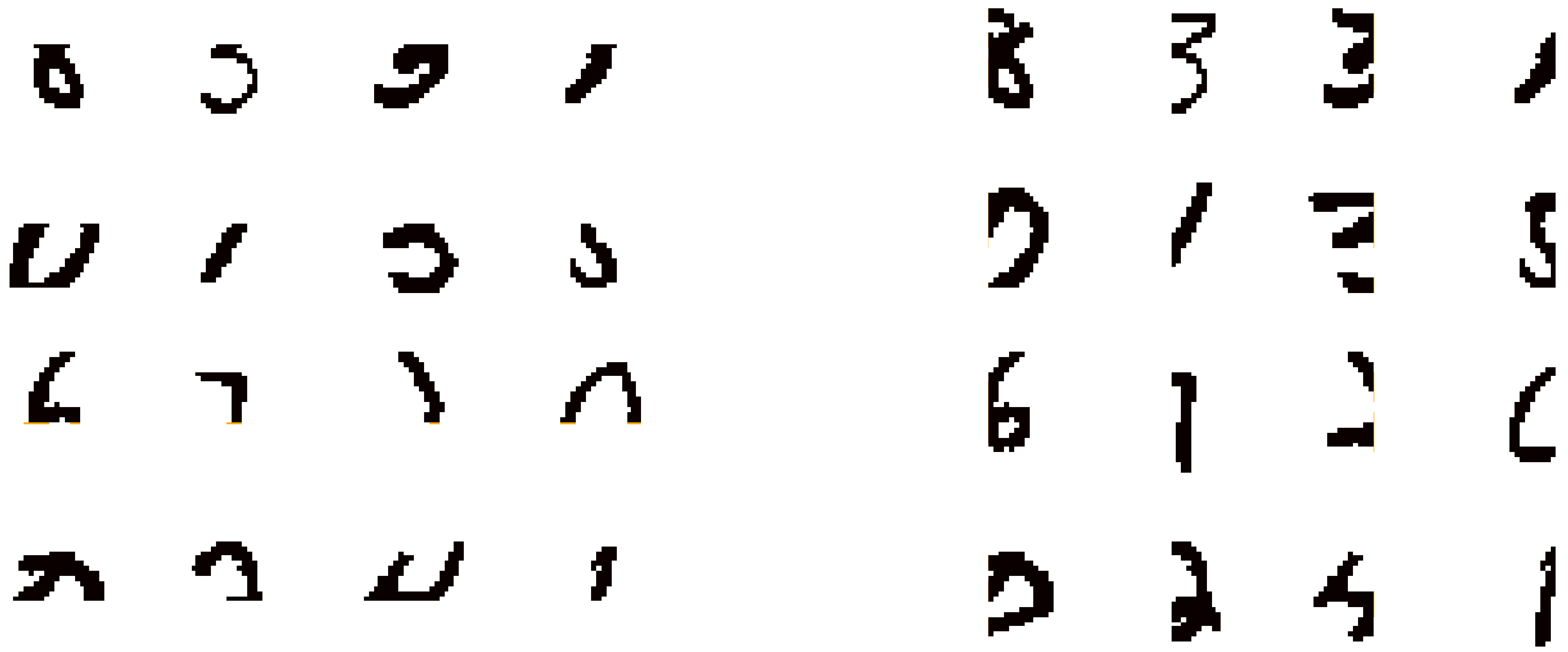
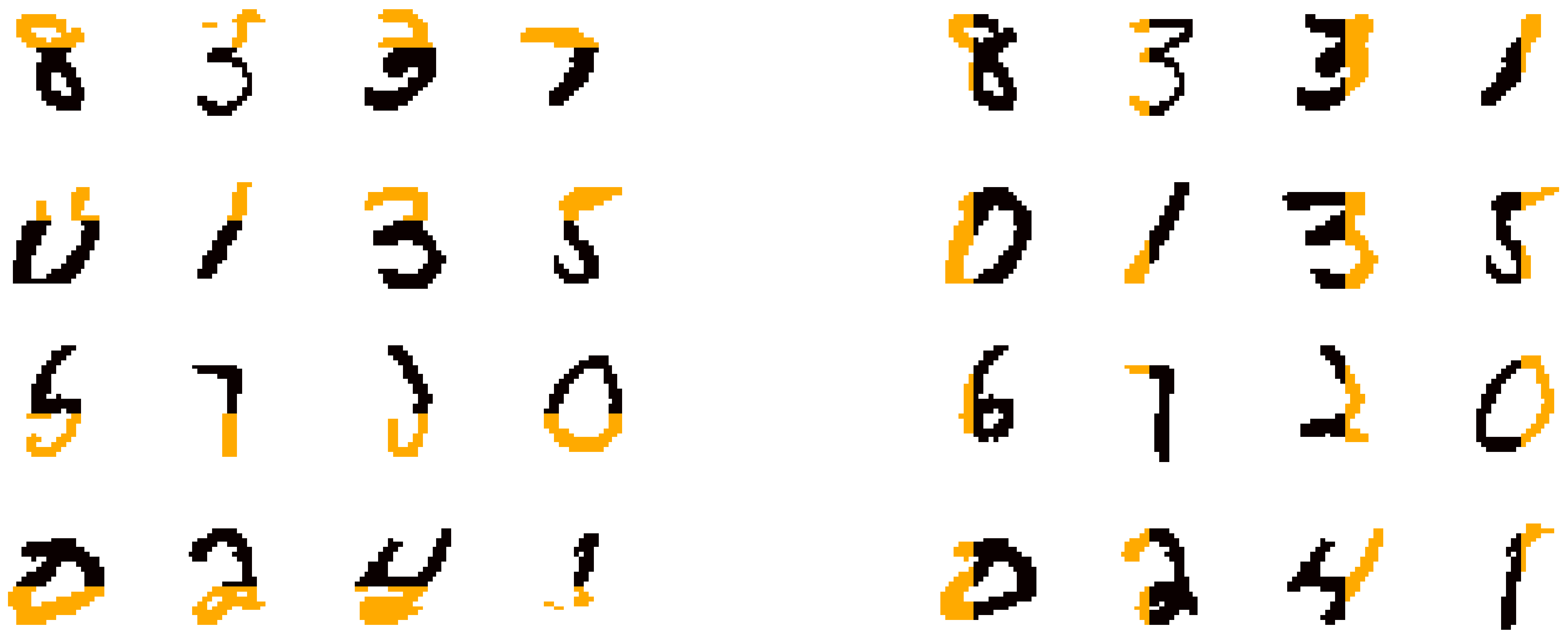


Image Restoration

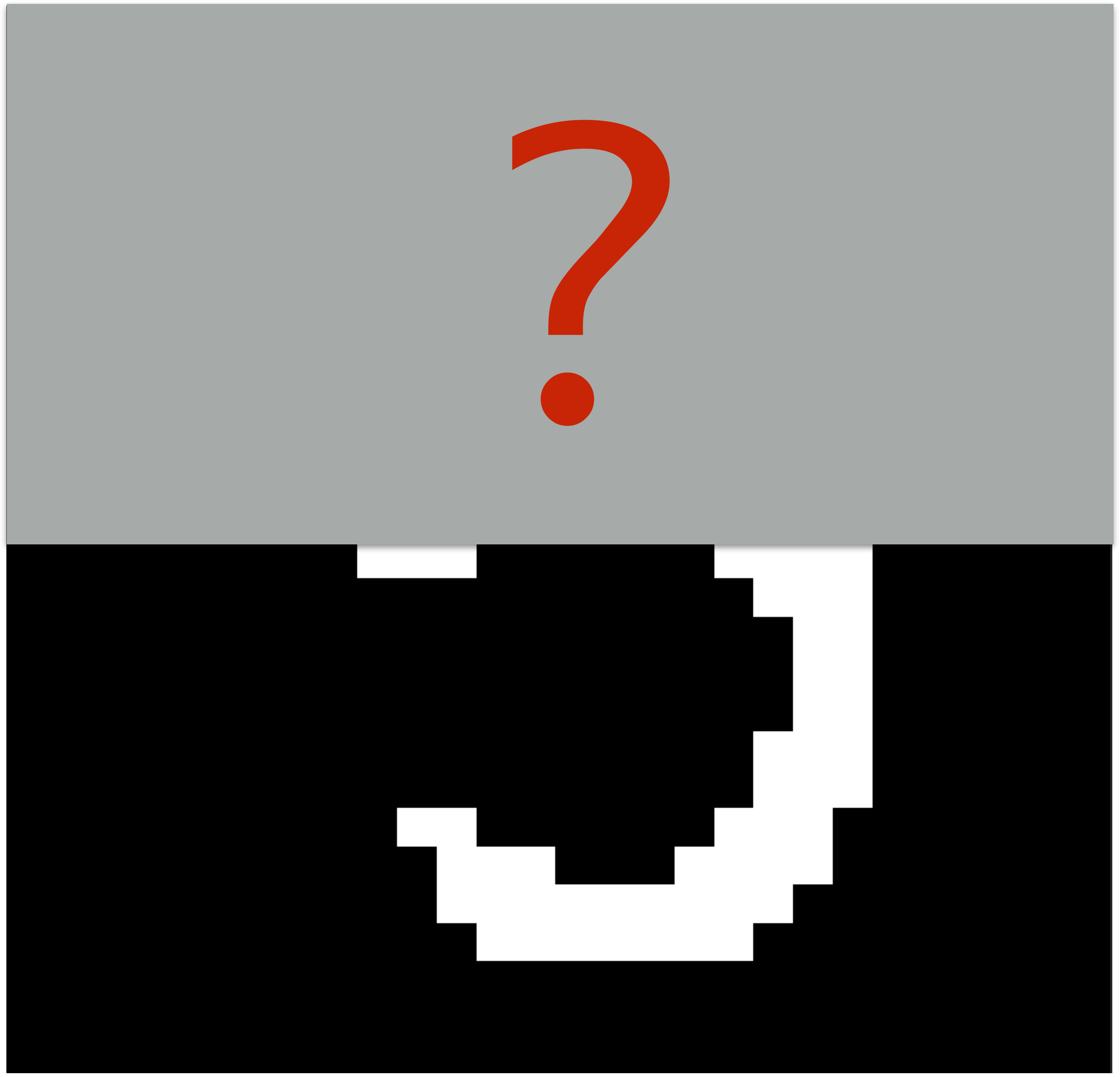
Han, Wang, Fan, LW, Zhang, 1709.01662, PRX in press



Arbitrary order compared to autoregressive models (state-of-the-art)

PixelCNN
PixelRNN

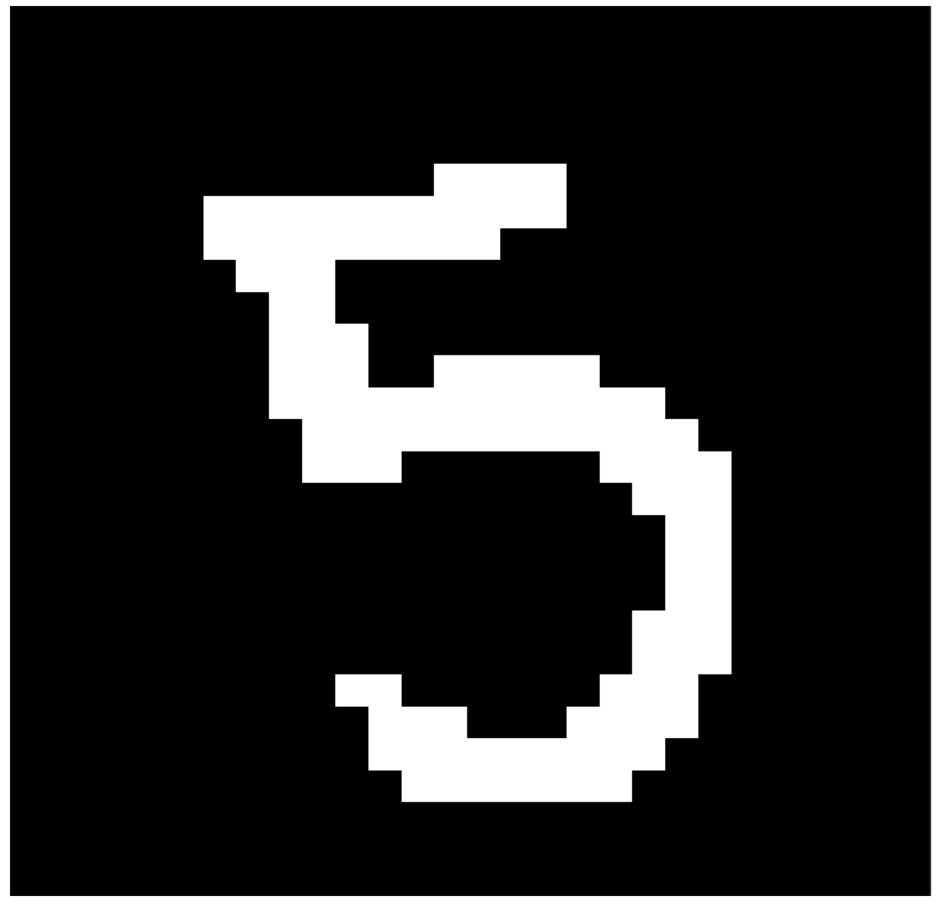
Quantum Perspective on Deep Learning

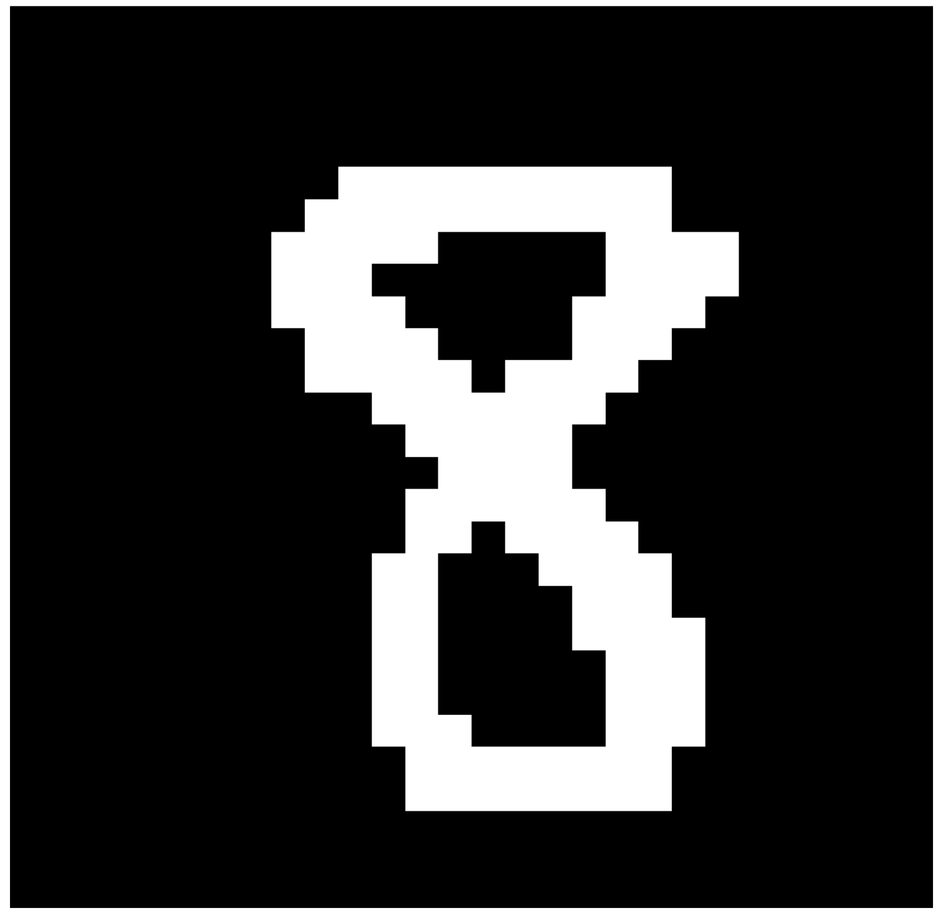


Quantum Perspective on Deep Learning

Q: How to quantify our prior knowledge on the data distribution?

A: Information pattern of the target probability functions





Quantum Perspective on Deep Learning

Q: How to quantify our prior knowledge on the data distribution?

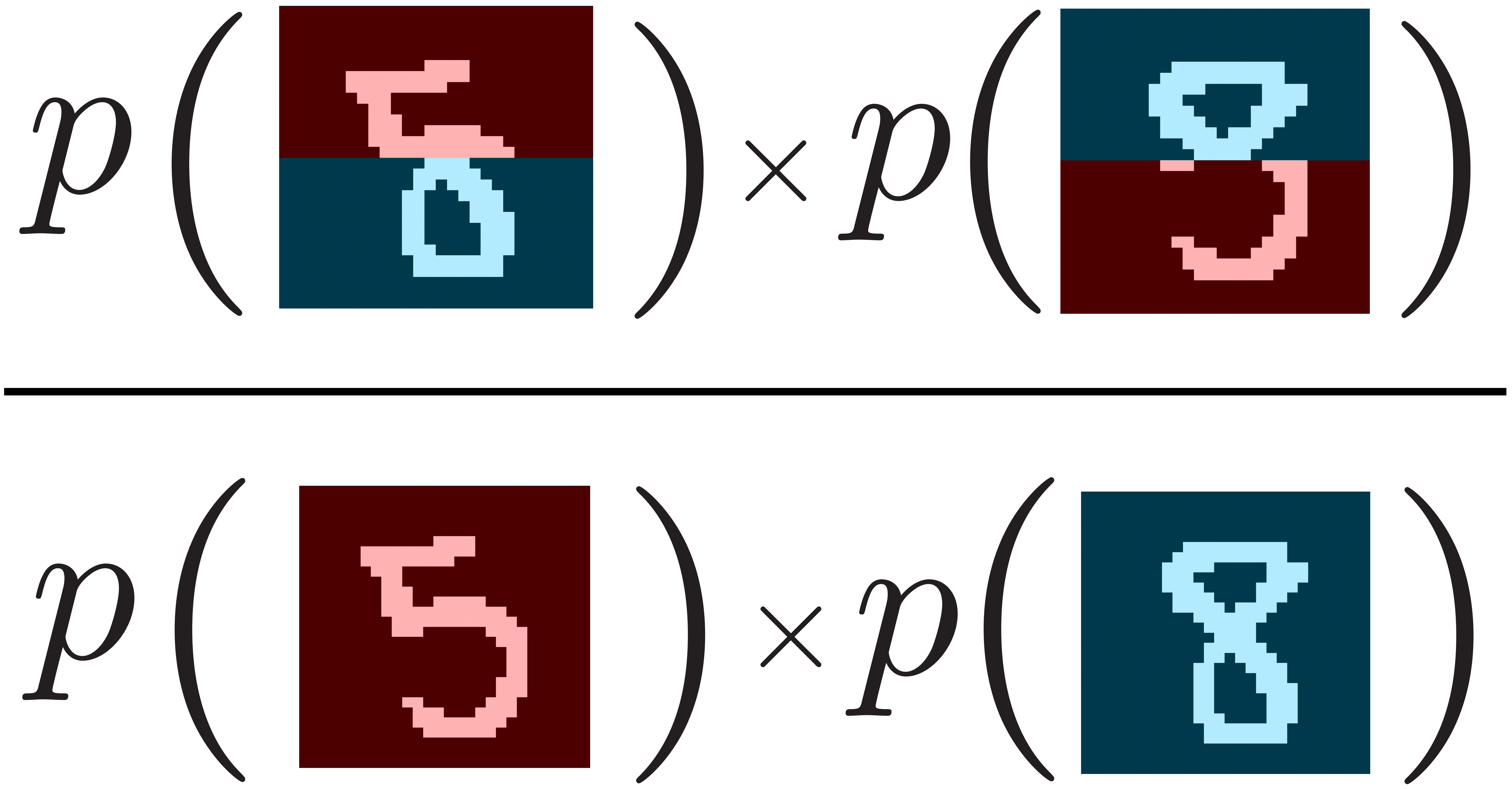
A: Information pattern of the target probability functions

Quantum Perspective on Deep Learning

Q: How to quantify our prior knowledge on the data distribution?

A: Information pattern of the target probability functions

Quantum Perspective on Deep Learning



Quantum Perspective on Deep Learning

Classical mutual information

$$I = -\left\langle \ln \left\langle \frac{p(\mathbf{x},\mathbf{y}')p(\mathbf{x}',\mathbf{y})}{p(\mathbf{x}',\mathbf{y}')p(\mathbf{x},\mathbf{y})} \right\rangle_{\mathbf{x}',\mathbf{y}'} \right\rangle_{\mathbf{x},\mathbf{y}}$$

Quantum Renyi entanglement entropy

$$S = -\ln \left\langle \left\langle \frac{\Psi(\mathbf{x},\mathbf{y}')\Psi(\mathbf{x}',\mathbf{y})}{\Psi(\mathbf{x}',\mathbf{y}')\Psi(\mathbf{x},\mathbf{y})} \right\rangle_{\mathbf{x}',\mathbf{y}'} \right\rangle_{\mathbf{x},\mathbf{y}}$$

Striking similarity implies common inductive bias

+Quantitative & interpretable approaches
+Principled structure design & learning

Cheng, Chen, LW,
1712.04144

arXiv:1704.01552v2 [cs.LG] 10 Apr 2017

Deep Learning and Quantum Entanglement: Fundamental Connections with Implications to Network Design

Yoav Levine YOAVLEVINE@CS.HUJI.AC.IL
David Yakira DAVIDYAKIRA@CS.HUJI.AC.IL
Nadav Cohen COHENNADAV@CS.HUJI.AC.IL
Amnon Shashua SHASHUA@CS.HUJI.AC.IL
The Hebrew University of Jerusalem

Abstract

Deep convolutional networks have witnessed unprecedented success in various machine learning applications. Formal understanding on what makes these networks so successful is gradually unfolding, but for the most part there are still significant mysteries to unravel. The inductive bias, which reflects prior knowledge embedded in the network architecture, is one of them. In this work, we establish a fundamental connection between the fields of quantum physics and deep learning. We use this connection for asserting novel theoretical observations regarding the role that the number of channels in each layer of the convolutional network fulfills in the overall inductive bias. Specifically, we show an equivalence between the function realized by a deep convolutional arithmetic circuit (ConvAC) and a quantum many-body wave function, which relies on their common underlying tensorial structure. This facilitates the use of quantum entanglement measures as well-defined quantifiers of a deep network's expressive ability to model intricate correlation structures of its inputs. Most importantly, the construction of a deep convolutional arithmetic circuit in terms of a Tensor Network is made available. This description enables us to carry a graph-theoretic analysis of a convolutional network, tying its expressiveness to a min-cut in the graph which characterizes it. Thus, we demonstrate a direct control over the inductive bias of the designed deep convolutional network via its channel numbers, which we show to be related to the min-cut in the underlying graph. This result is relevant to any practitioner designing a convolutional network for a specific task. We theoretically analyze convolutional arithmetic circuits, and empirically validate our findings on more common convolutional networks which involve ReLU activations and max pooling. Beyond the results described above, the description of a deep convolutional network in well-defined graph-theoretic tools and the formal structural connection to quantum entanglement, are two interdisciplinary bridges that are brought forth by this work.

Deep Learning and Quantum Entanglement: Fundamental Connections with Implications to Network Design

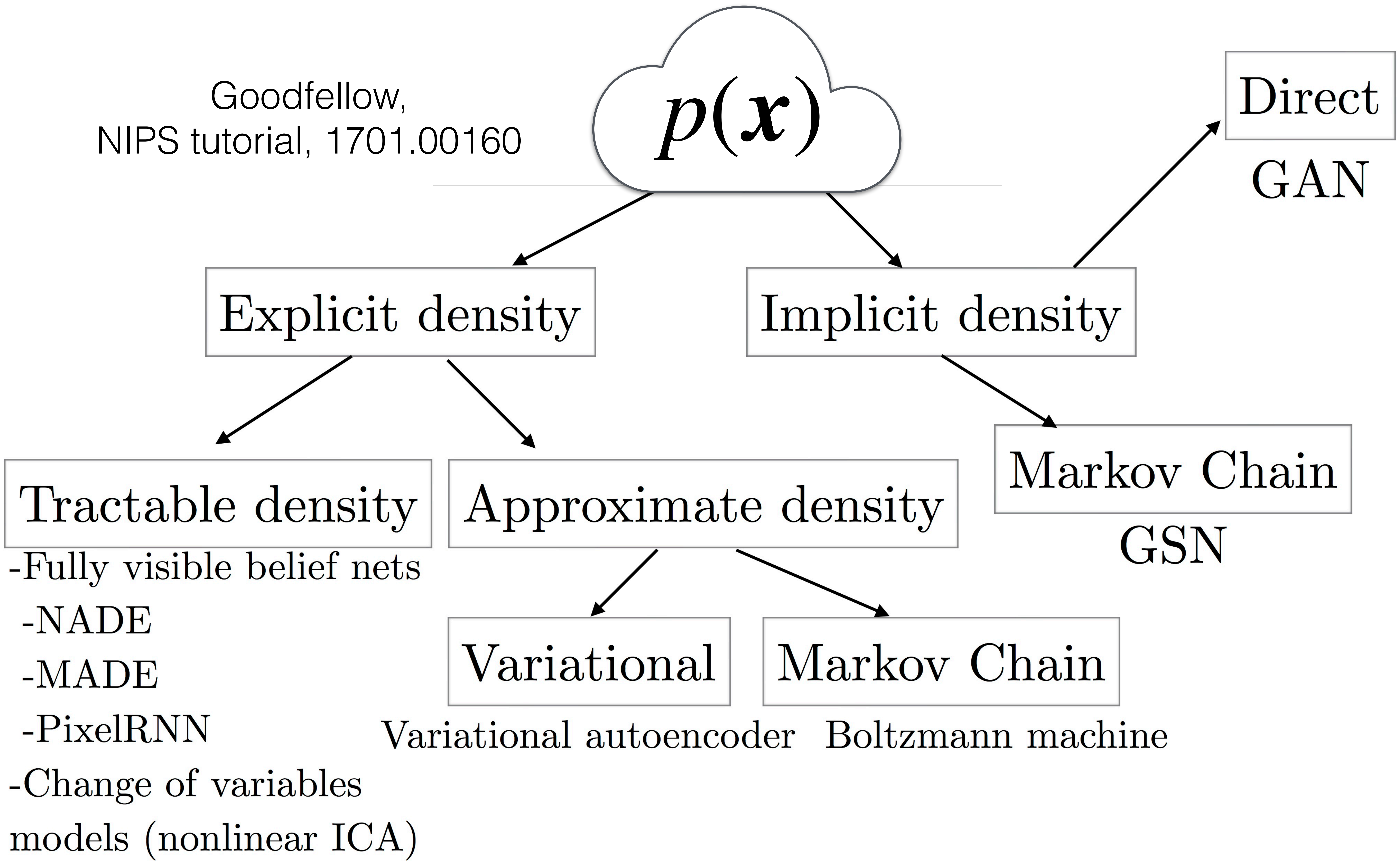
Yoav Levine YOAVLEVINE@CS.HUJI.AC.IL
David Yakira DAVIDYAKIRA@CS.HUJI.AC.IL
Nadav Cohen COHENNADAV@CS.HUJI.AC.IL
Amnon Shashua SHASHUA@CS.HUJI.AC.IL

arXiv:1

our findings on more common convolutional networks which involve ReLU activations and max pooling. Beyond the results described above, the description of a deep convolutional network in well-defined graph-theoretic tools and the formal structural connection to quantum entanglement, are two interdisciplinary bridges that are brought forth by this work.

Physics genes of generative models

Goodfellow,
NIPS tutorial, 1701.00160

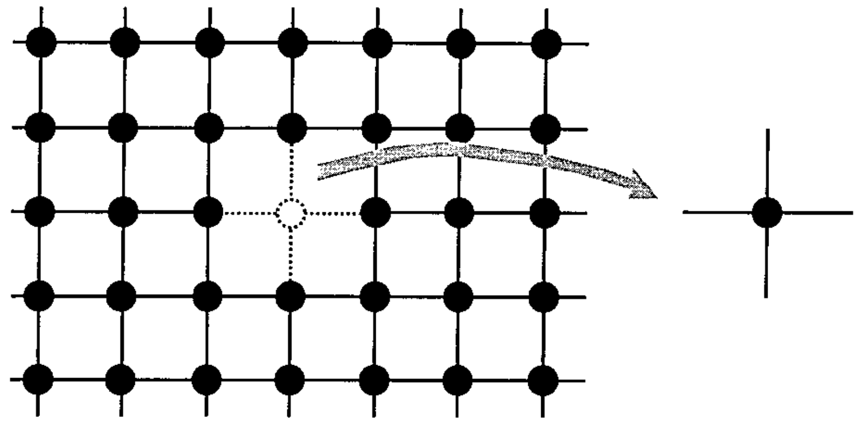


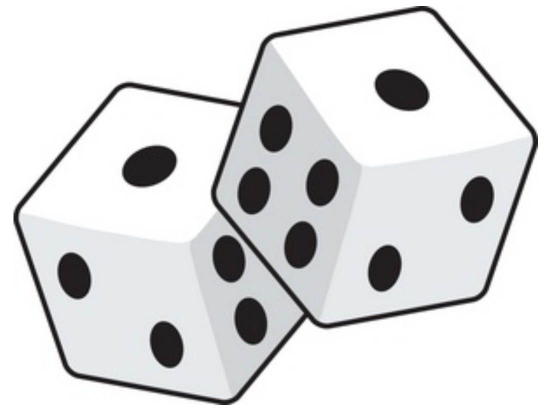
Physics genes of generative models

Goodfellow,
NIPS tutorial, 1701.00160

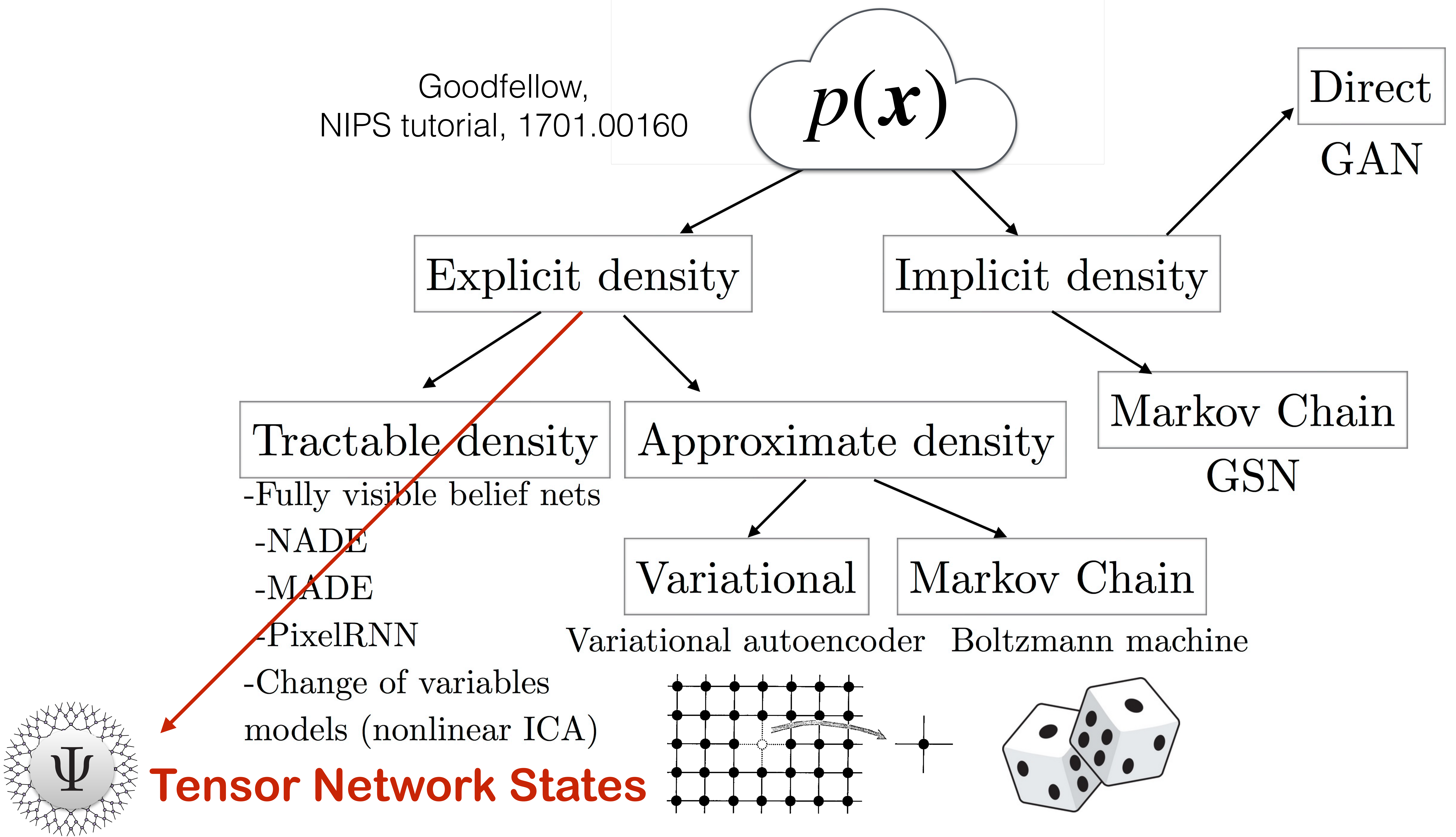
$p(\boldsymbol{x})$

- Explicit density
  - Tractable density
    - -Fully visible belief nets
      - -NADE
      - -MADE
      - -PixelRNN
    - -Change of variables models (nonlinear ICA)
  - Approximate density
    - Variational
      - Variational autoencoder
    - Markov Chain
      - Boltzmann machine
- Implicit density
  - Direct
    - GAN
  - Markov Chain
    - GSN

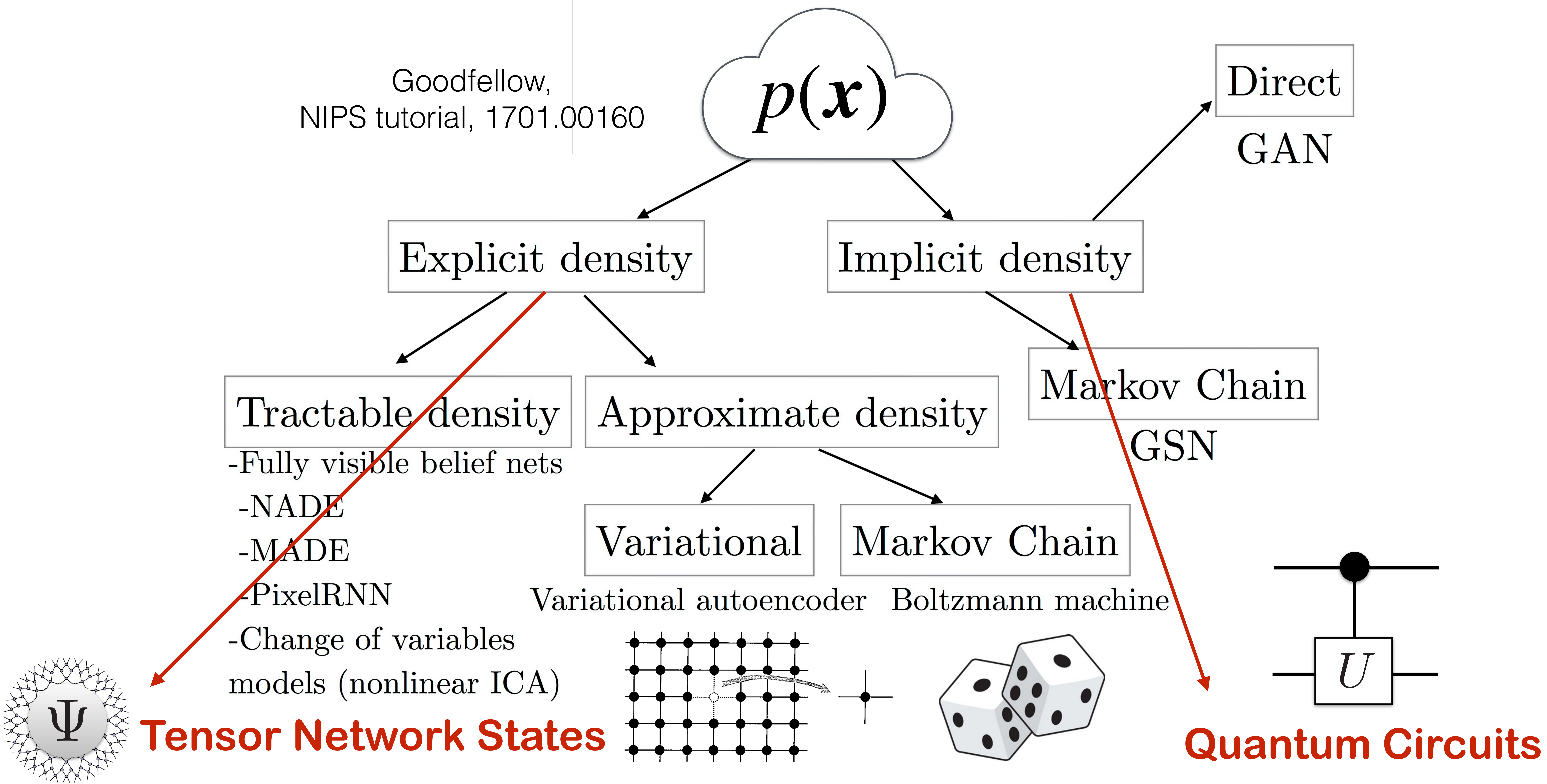




Physics genes of generative models

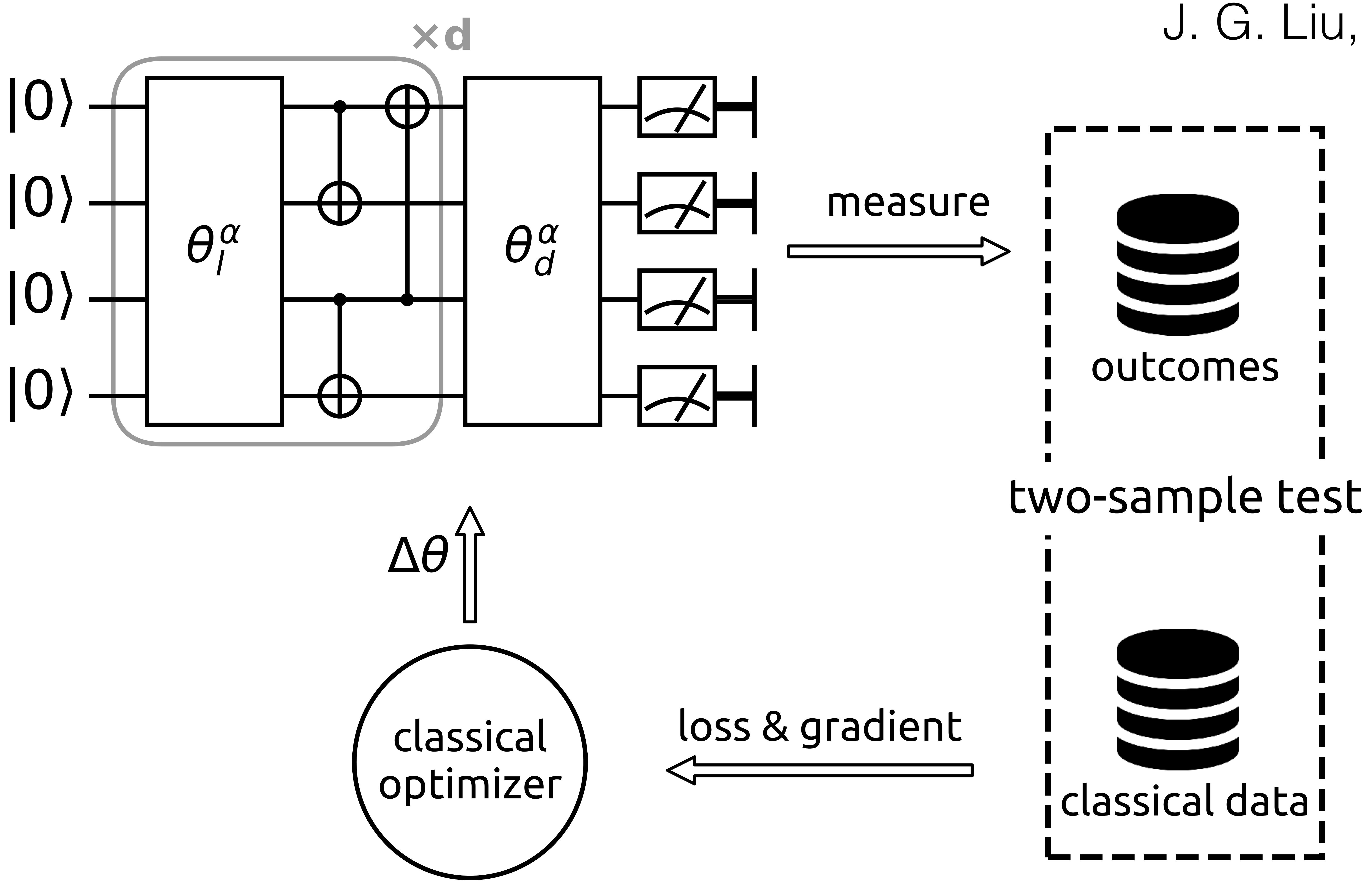


Physics genes of generative models



Quantum Circuit Born Machine

J. G. Liu, LW, 1804.04168



Train the quantum circuit as a probabilistic generative model

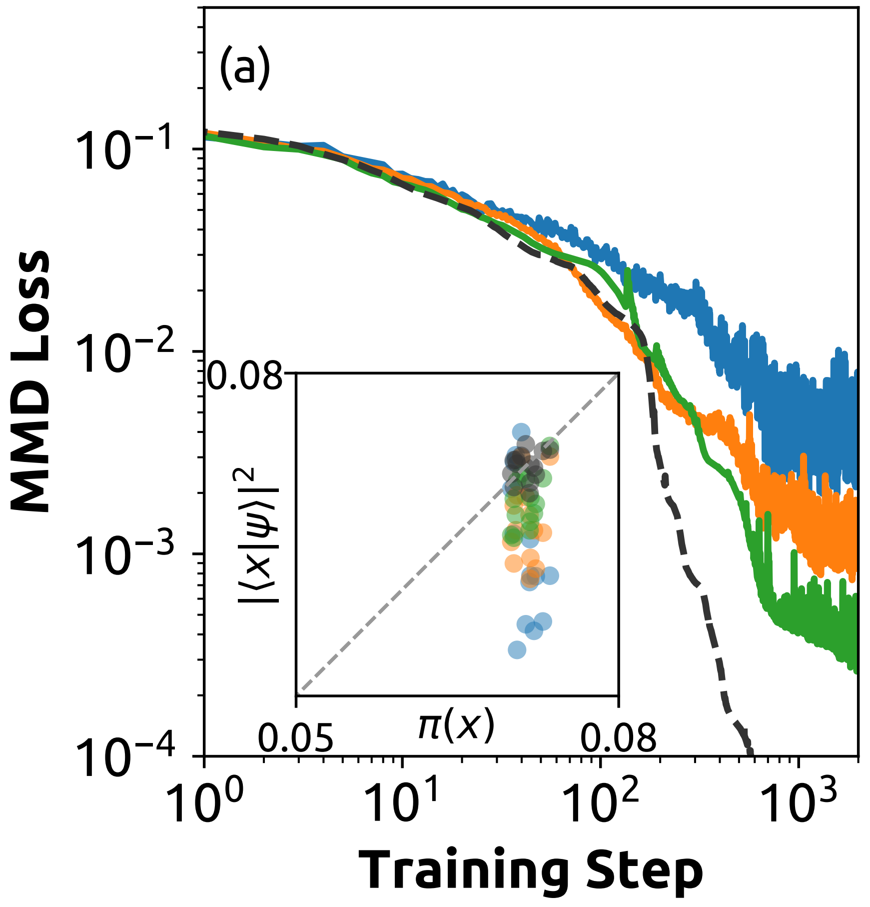
Quantum sampling complexity underlines the “quantum supremacy”

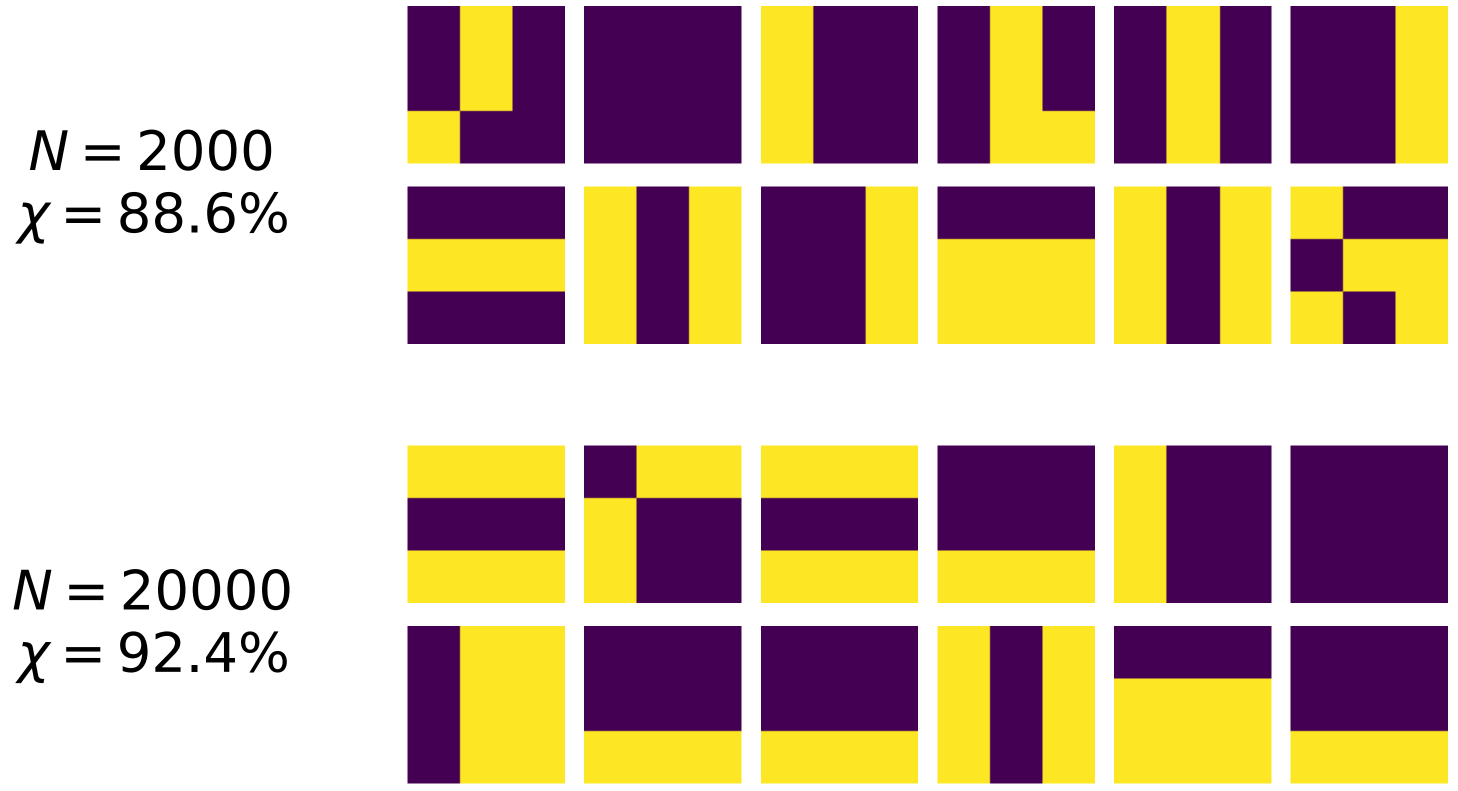
Learning the Quantum Circuit

J. G. Liu, LW, 1804.04168

- Objectivity function for the quantum implicit model: maximum mean discrepancy

- Differentiable learning of the circuit parameters: unbiased gradient estimator

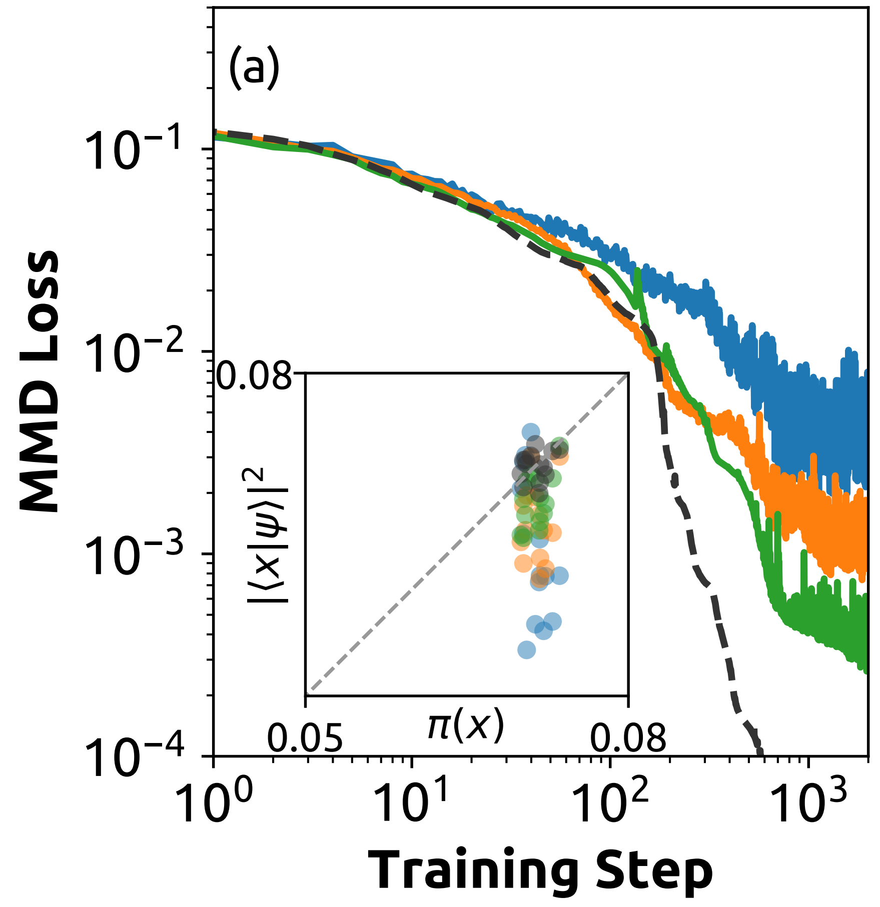


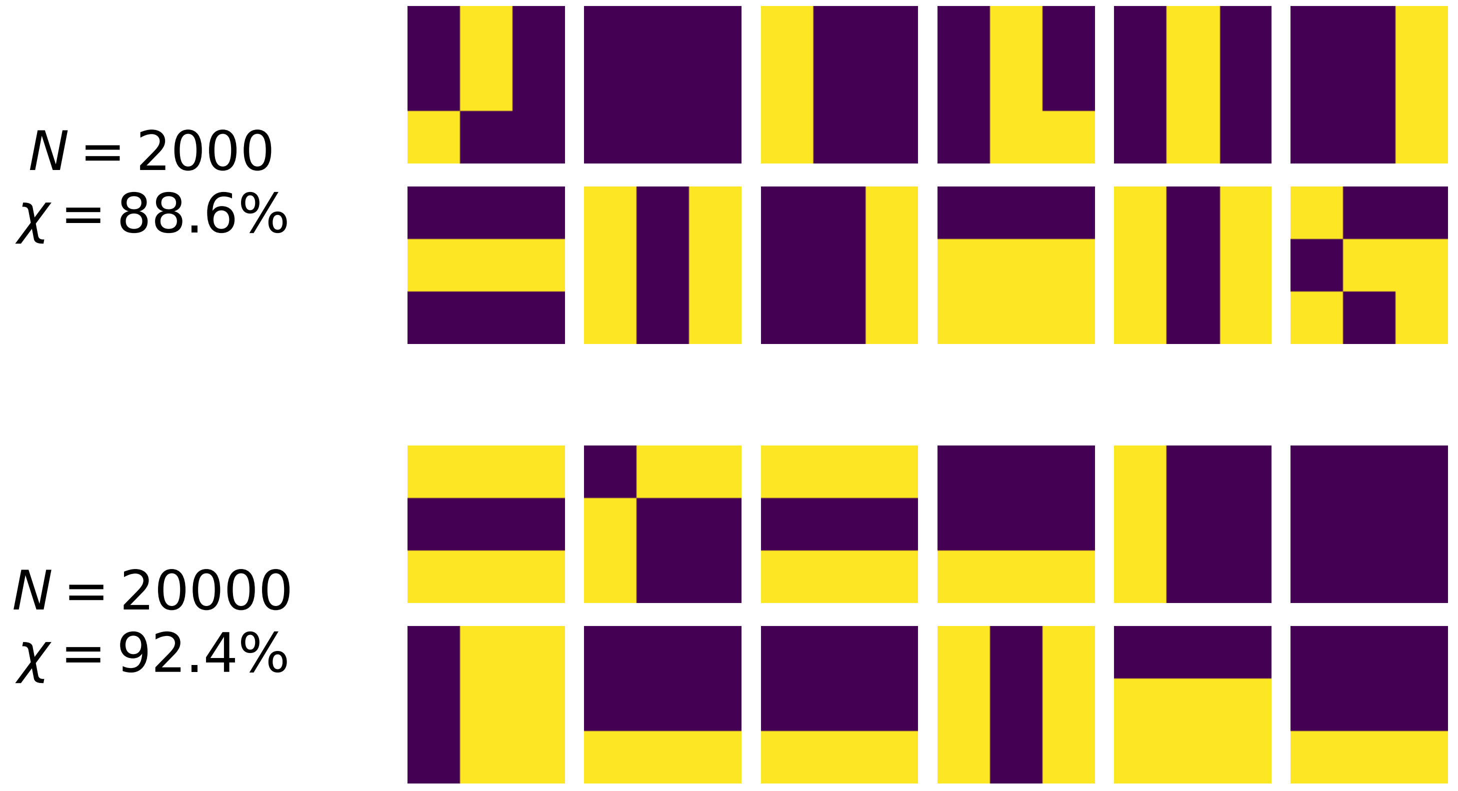


Learning the Quantum Circuit

J. G. Liu, LW, 1804.04168

- Objectivity function for the quantum implicit model: maximum mean discrepancy
- Differentiable learning of the circuit parameters: unbiased gradient estimator





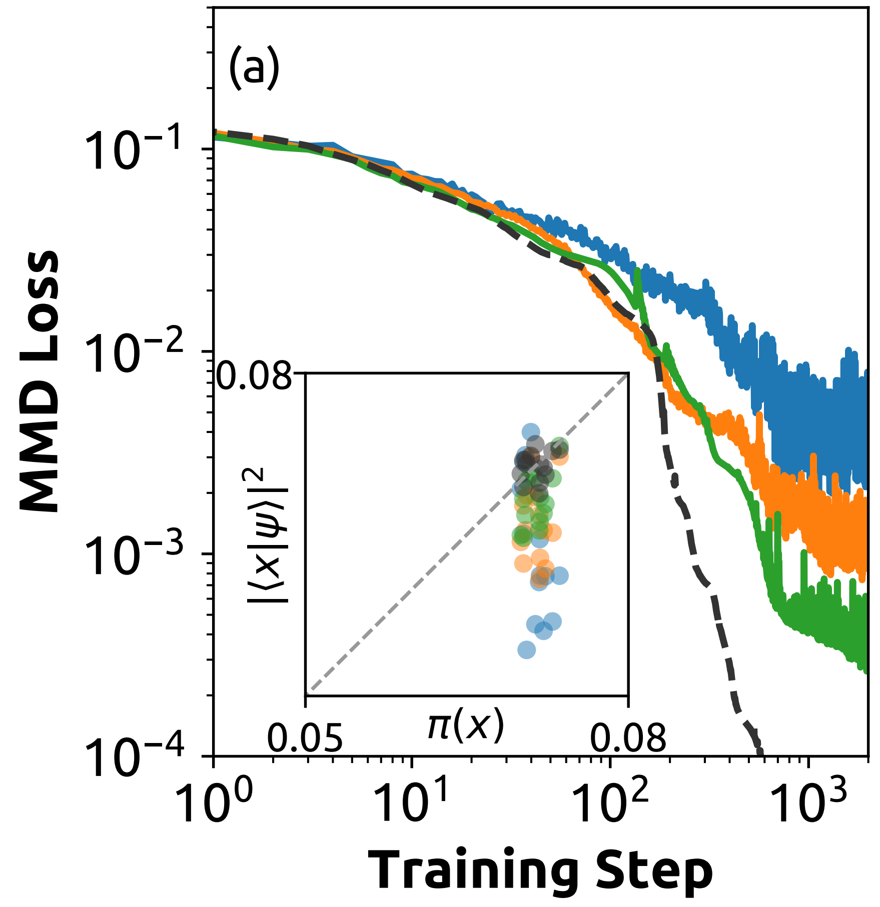
Born Machine experiment	Benedetti, Garcia-Pintos, Nam, Perdomo-Ortiz, 1801.07686
TNS inspired circuit architecture	Huggins, Patel, Whaley, Stoudenmire, 1803.11537
Quantum generative model	Gao, Zhang, Duan, 1711.02038
Quantum adversarial training	Dallaire-Demers, Lloyd, Benedetti 1804.08641,1804.09139, 1806.00463

Learning the Quantum Circuit

J. G. Liu, LW, 1804.04168

- Objectivity function for the quantum implicit model: maximum mean discrepancy

- Differentiable learning of the circuit parameters: unbiased gradient estimator



Born Machine ... 686

TNS inspired ...

Quantum gen...

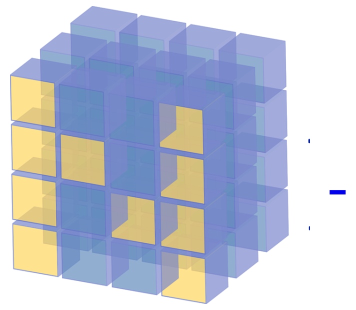
Quantum adv... 139, 1806.00463

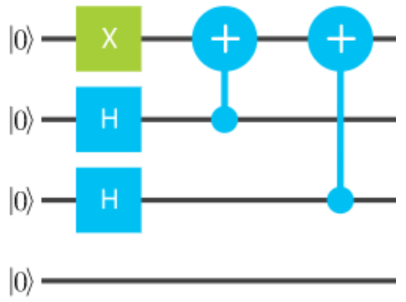
Quantum Software 2.0

Karpathy, Medium 2017

Try it yourself!

http://lib.itp.ac.cn/html/panzhang/mps/tutorial/





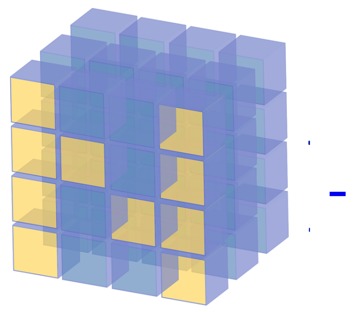
https://github.com/GiggleLiu/QuantumCircuitBornMachine

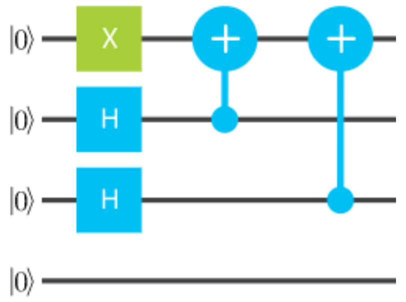
https://github.com/QuantumBFS/Yao.jl/blob/master/examples/QCBM.ipynb

Thank You!

Try it yourself!

http://lib.itp.ac.cn/html/panzhang/mps/tutorial/





https://github.com/GiggleLiu/QuantumCircuitBornMachine

https://github.com/QuantumBFS/Yao.jl/blob/master/examples/QCBM.ipynb

Thank You!

Pan Zhang **Zhao-Yu Han** **Jun Wang** **Jin-Guo Liu**
Jing Chen **Song Cheng** **Roger Luo** **Tao Xiang**