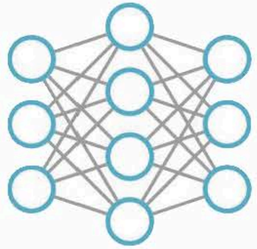
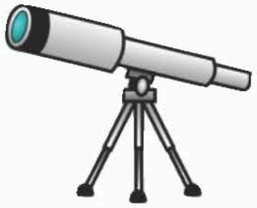
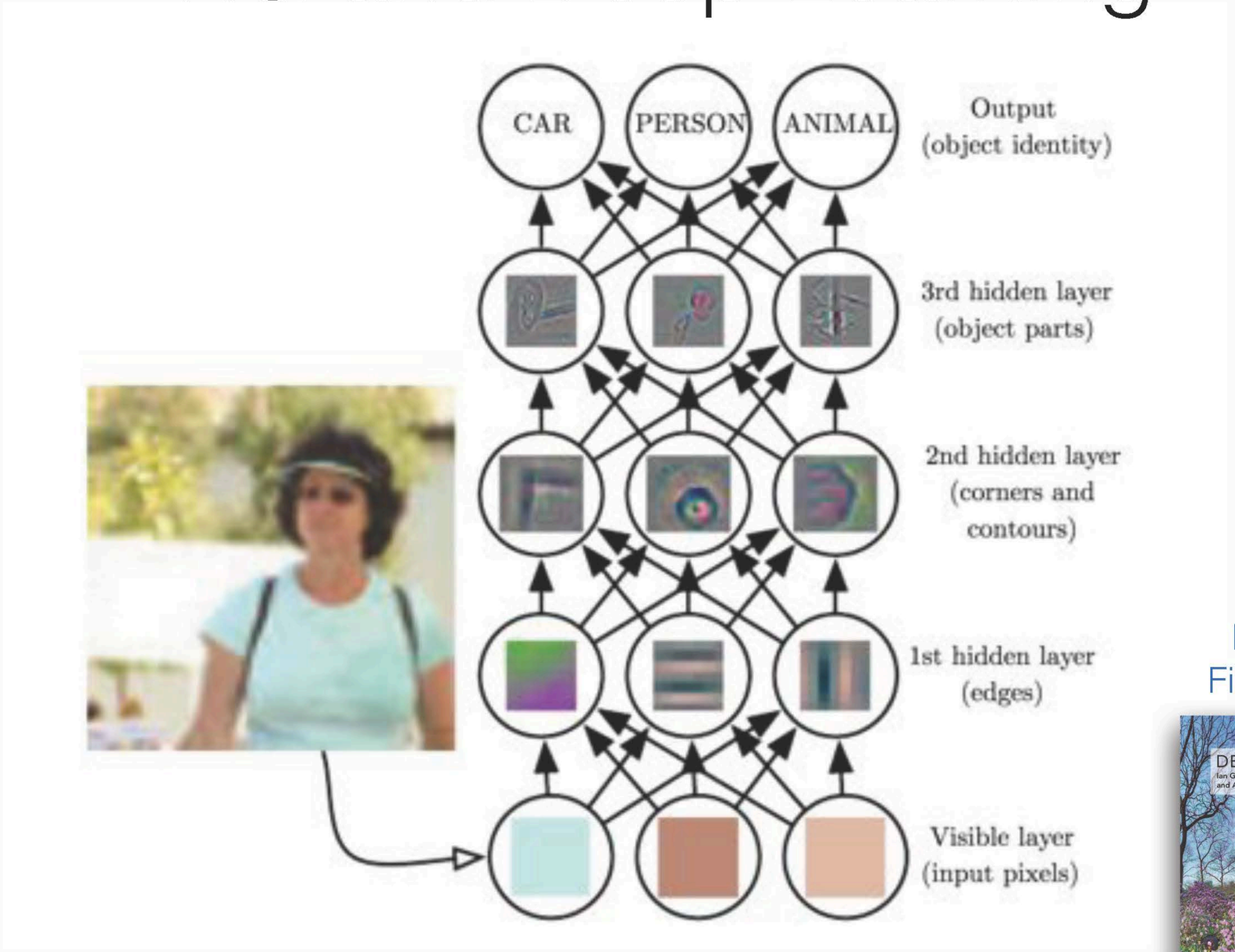


Neural Network Renormalization Group

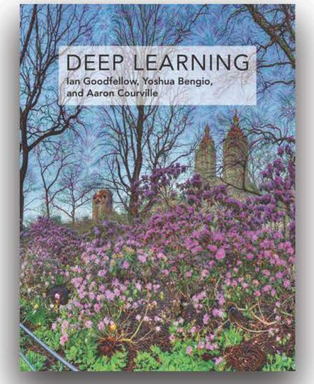


Lei Wang (王磊)
Institute of Physics, CAS
https://wangleiphy.github.io

RG and Deep Learning

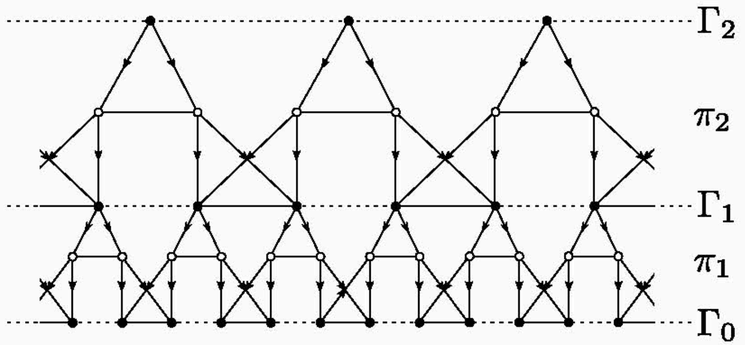


Page 6
Figure 1.2

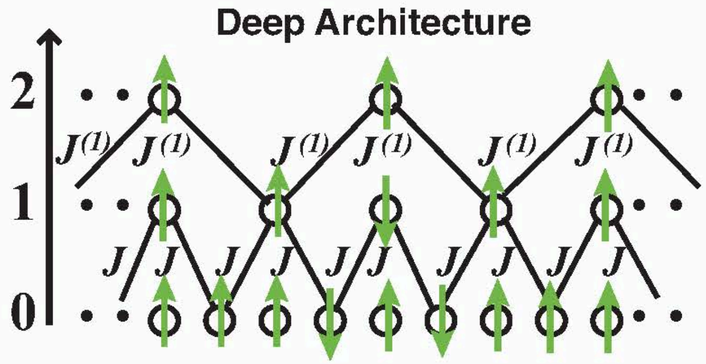


Goodfellow, Bengio, Courville, http://www.deeplearningbook.org/

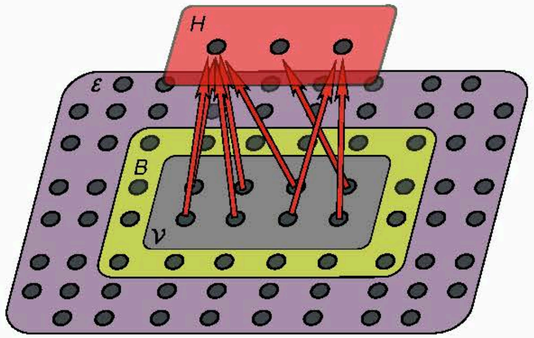
RG and Deep Learning



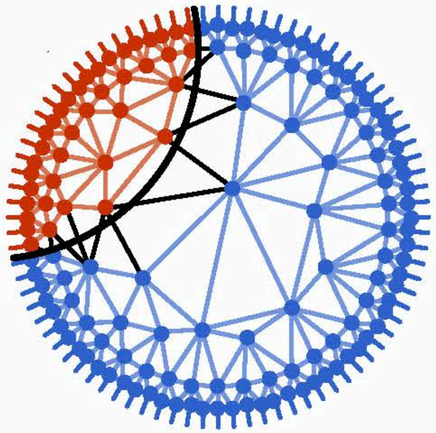
Bény, 1301.3124



Mehta and Schwab, 1410.3831



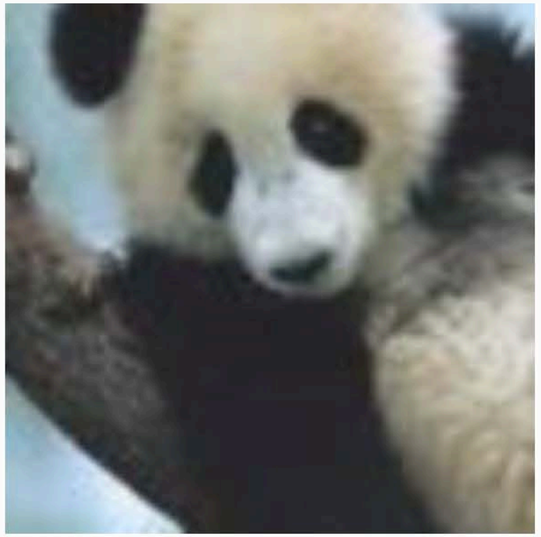
Koch-Janusz and Ringel, 1704.06279



You, Yang, Qi, 1709.01223

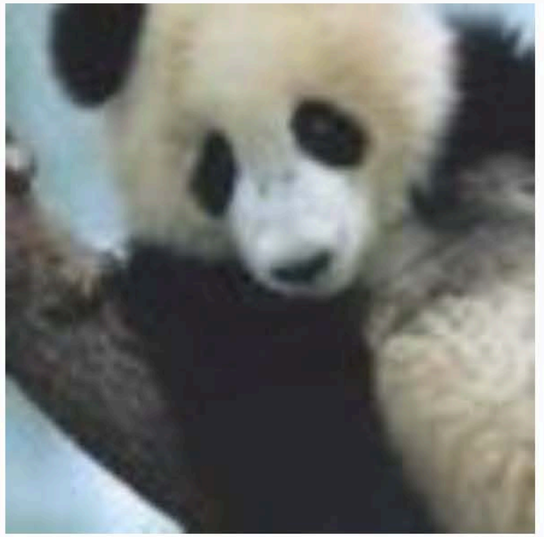
and more…

RG and Deep Learning



$+ .007 \times$

$=$



Panda
58% confidence

Goodfellow et al, 2014

Gibbon
99% confidence

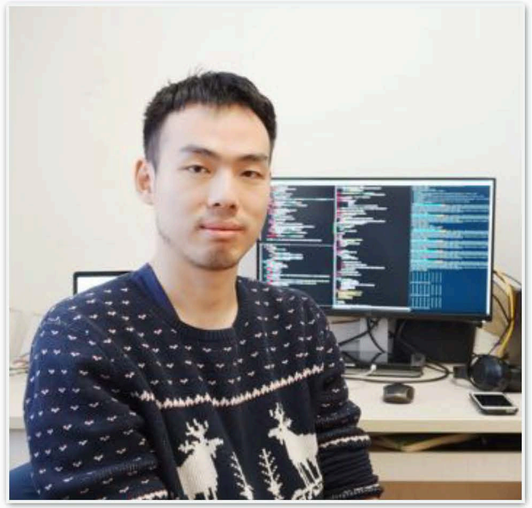
Vulnerability of deep learning, Kenway, 1803.06111 & 1803.10995

and more…

Motivation

RG offers a theoretical understanding of DL

In return, DL helps to solve physics problems

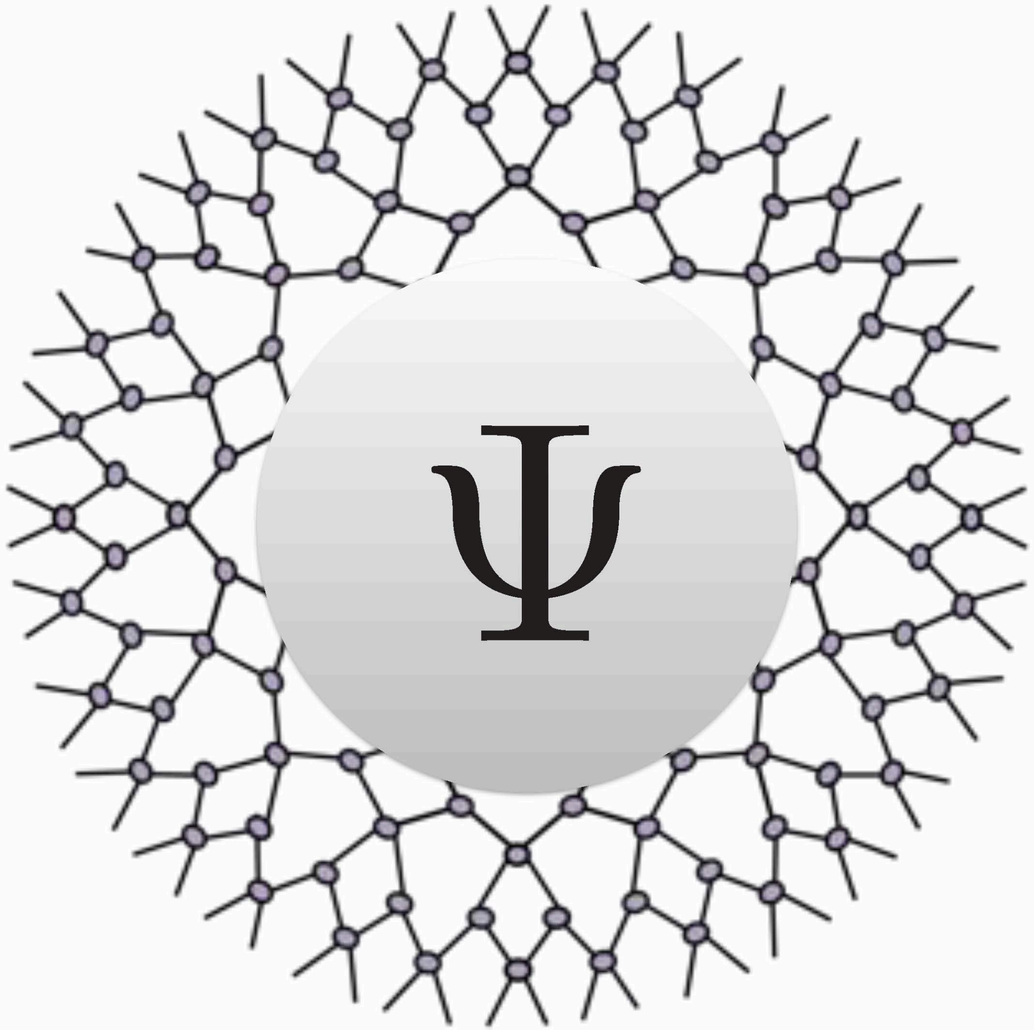


Shuo-Hui Li (李烁辉)

arXiv:1802.02840

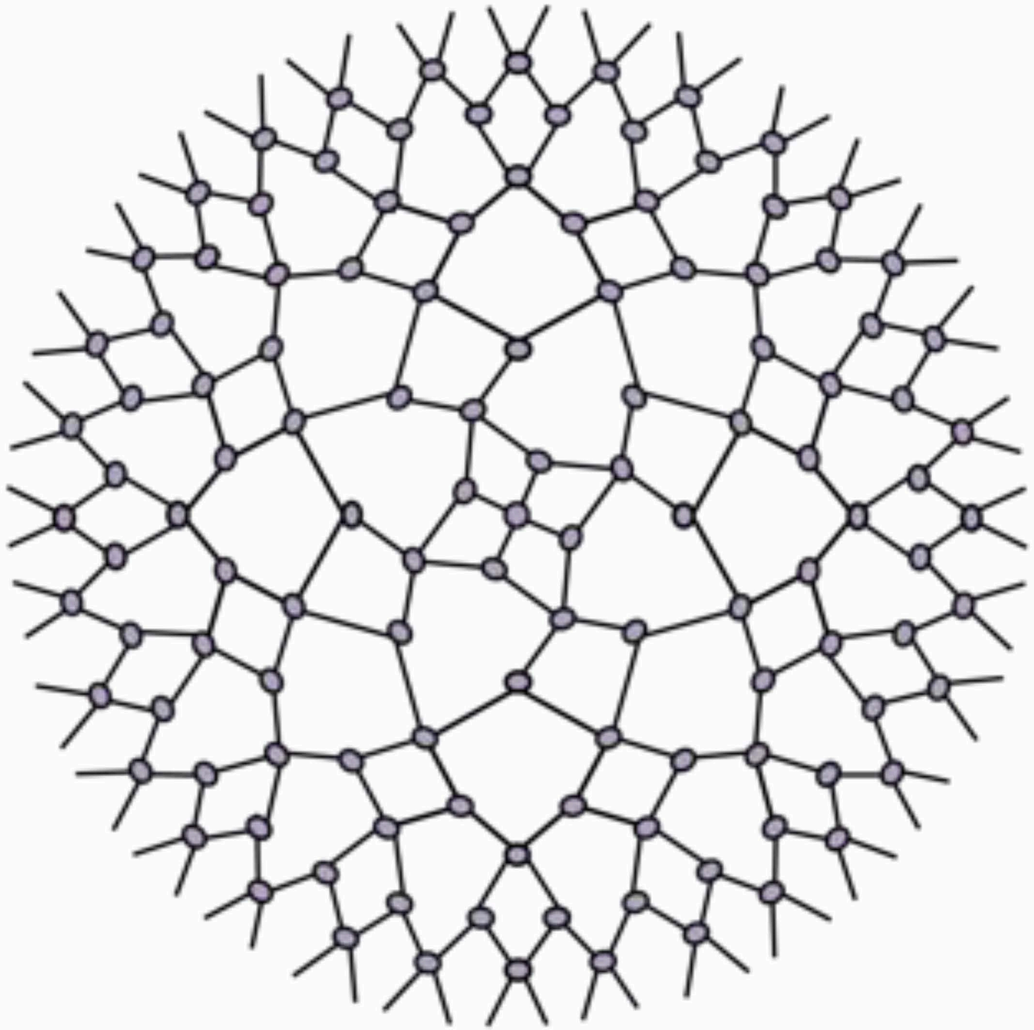
https://github.com/li012589/NeuralRG

Multi-Scale Entanglement Renormalization Ansatz



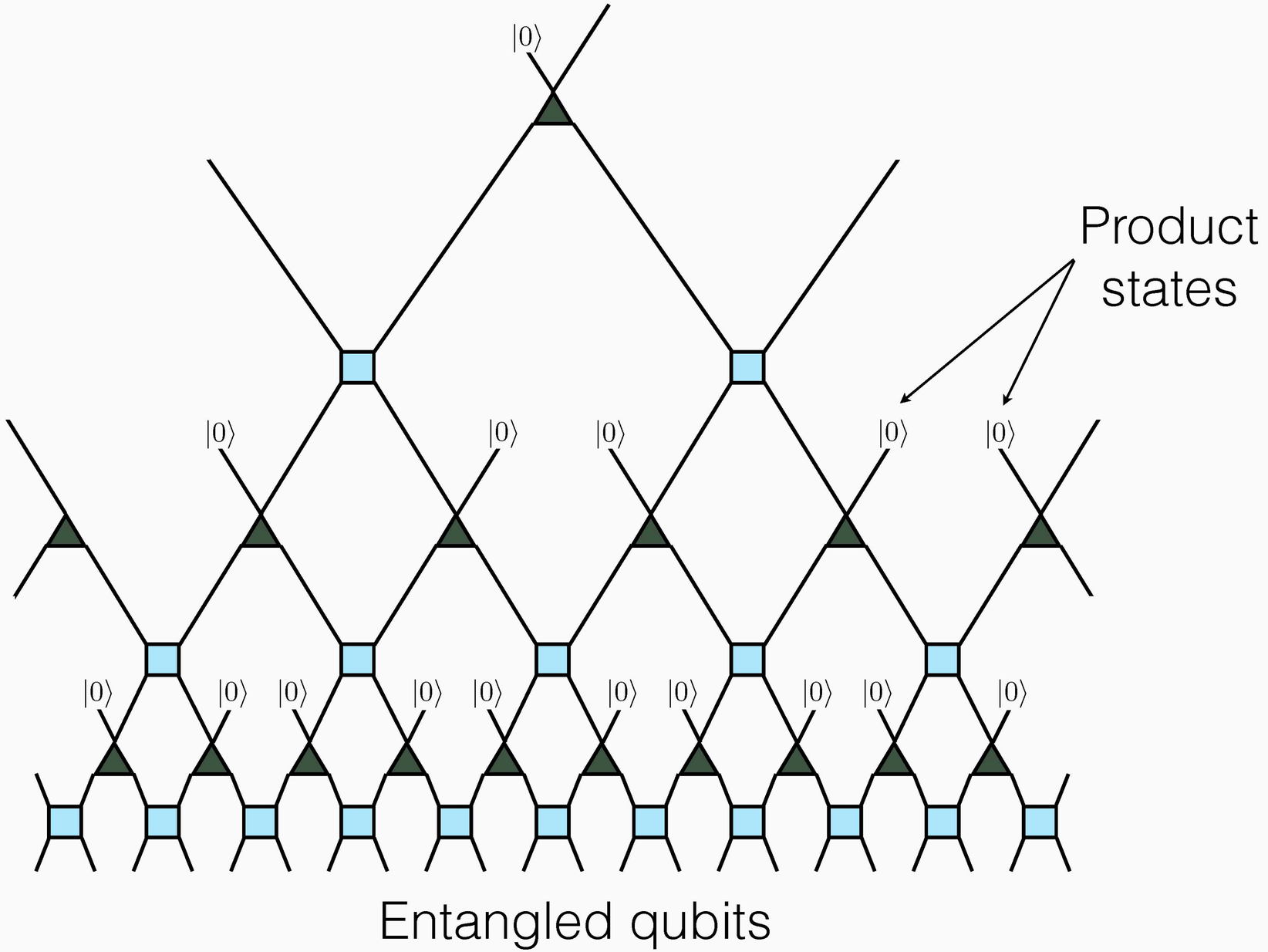
Vidal 2006

Multi-Scale Entanglement Renormalization Ansatz

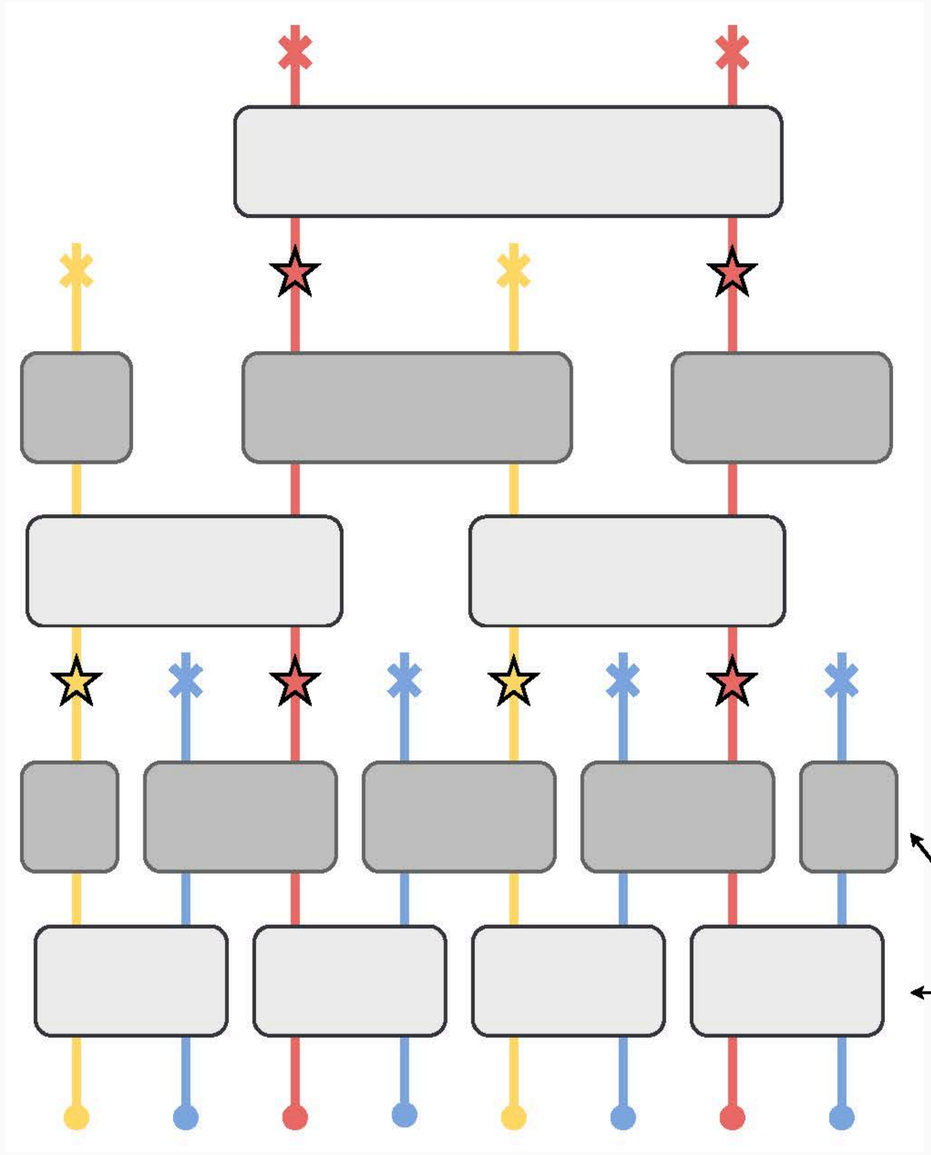


Vidal 2006

MERA as a quantum circuit



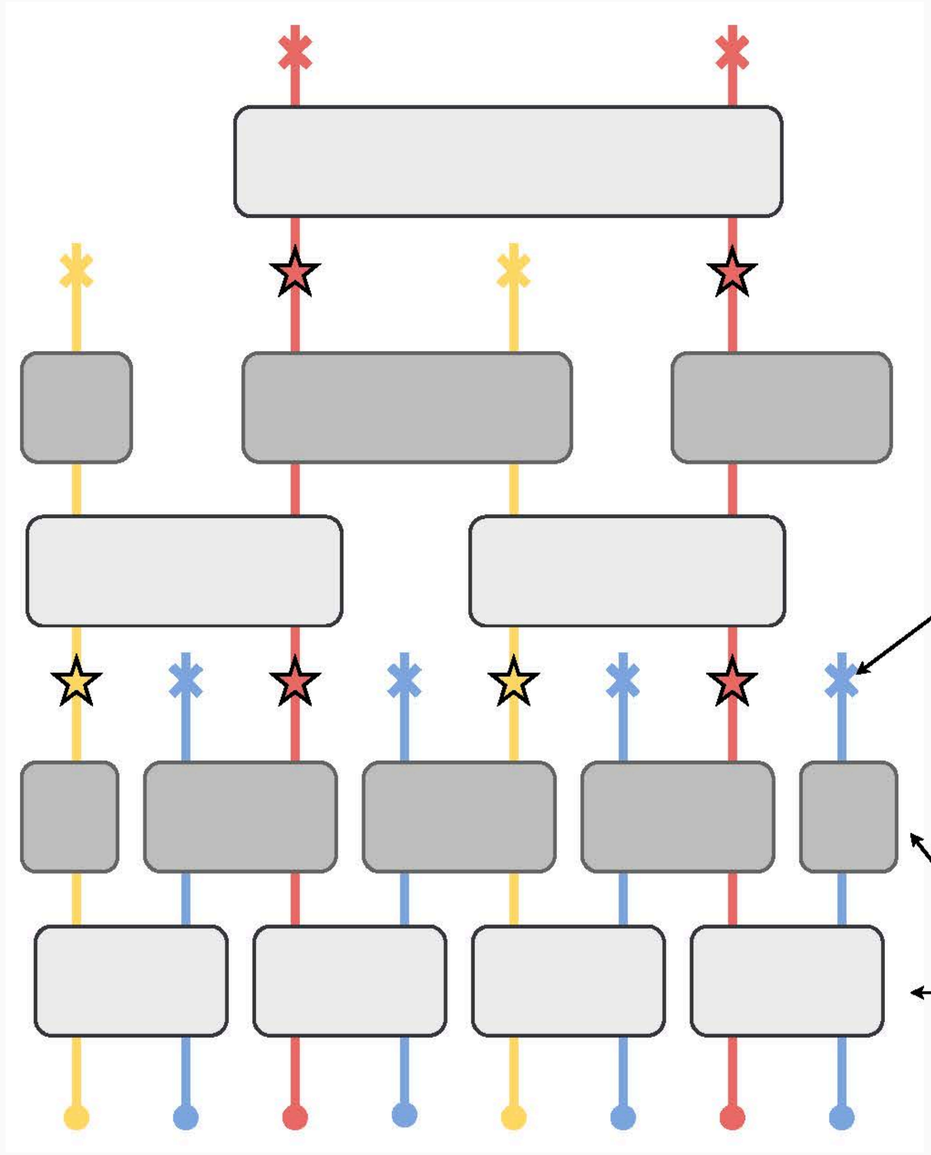
Neural Network Renormalization Group



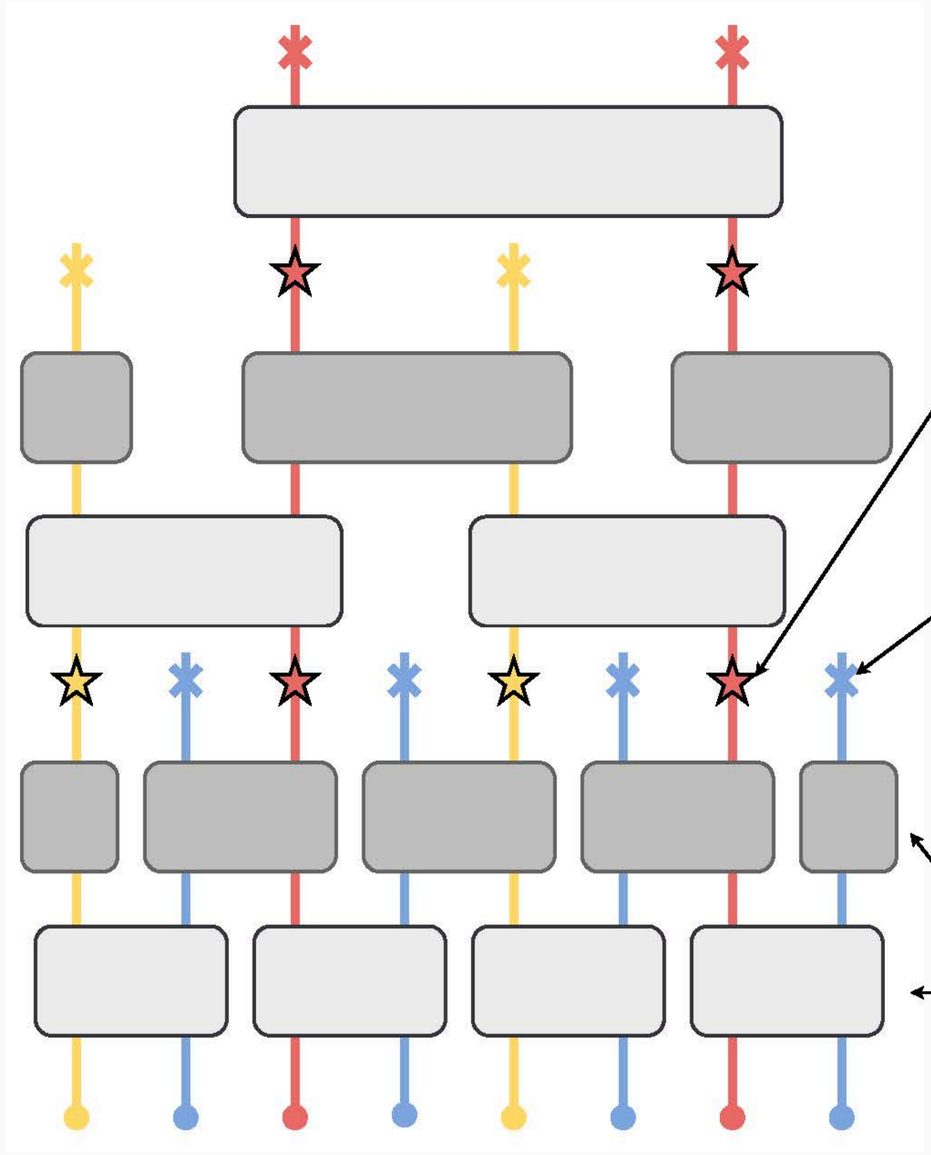
Bijective
neural nets

Correlated classical variables

Neural Network Renormalization Group



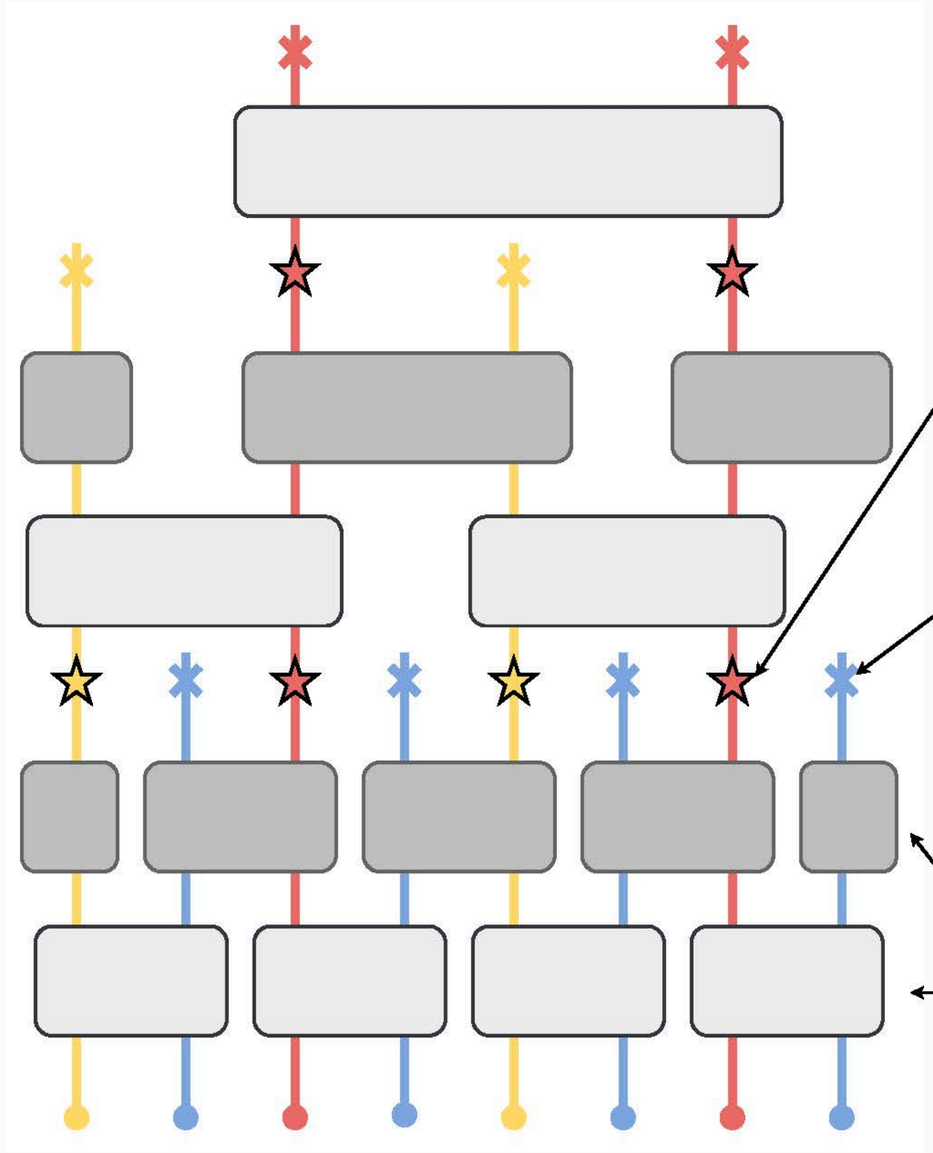
Neural Network Renormalization Group



Neural Network Renormalization Group

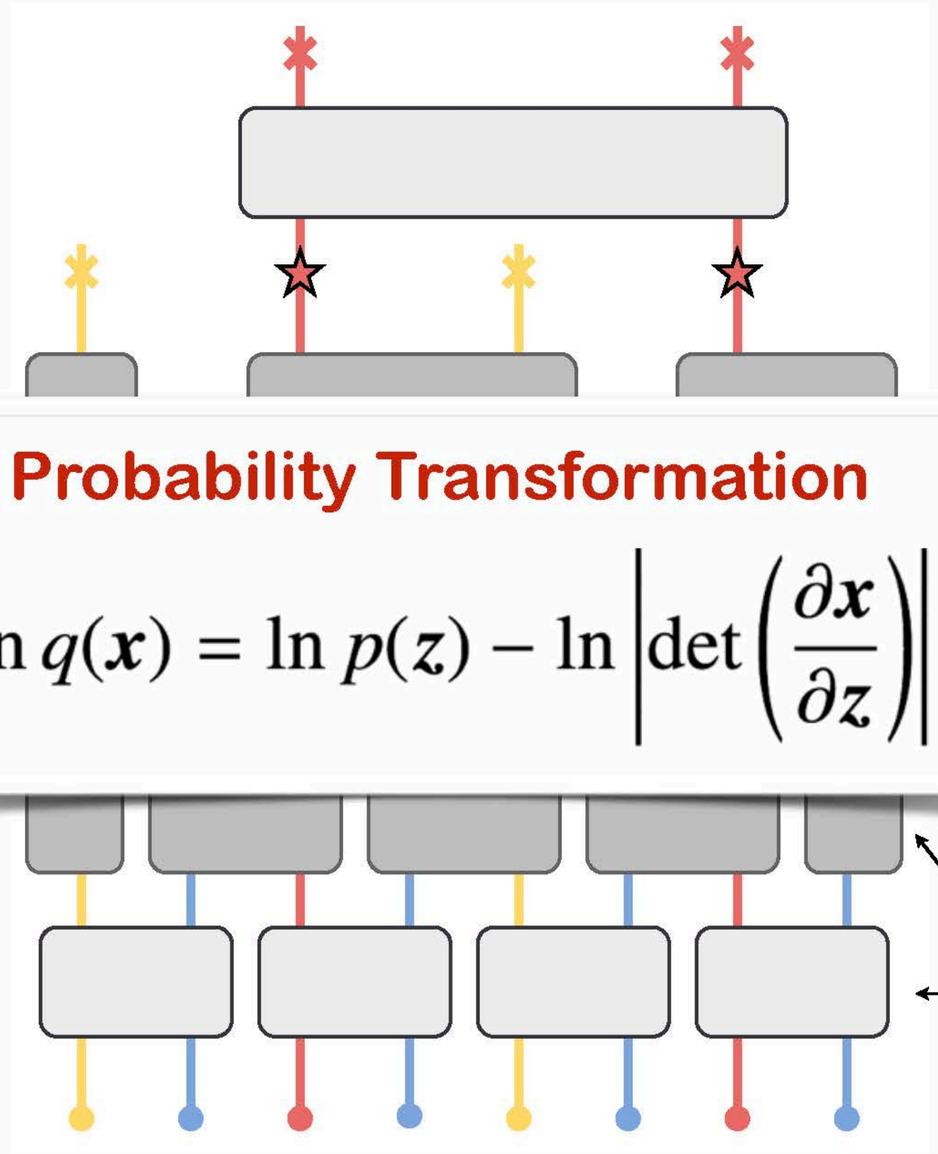
$$z = g^{-1}(x)$$

$$x = g(z)$$



Neural Network Renormalization Group

$$z = g^{-1}(x)$$



Collective variables

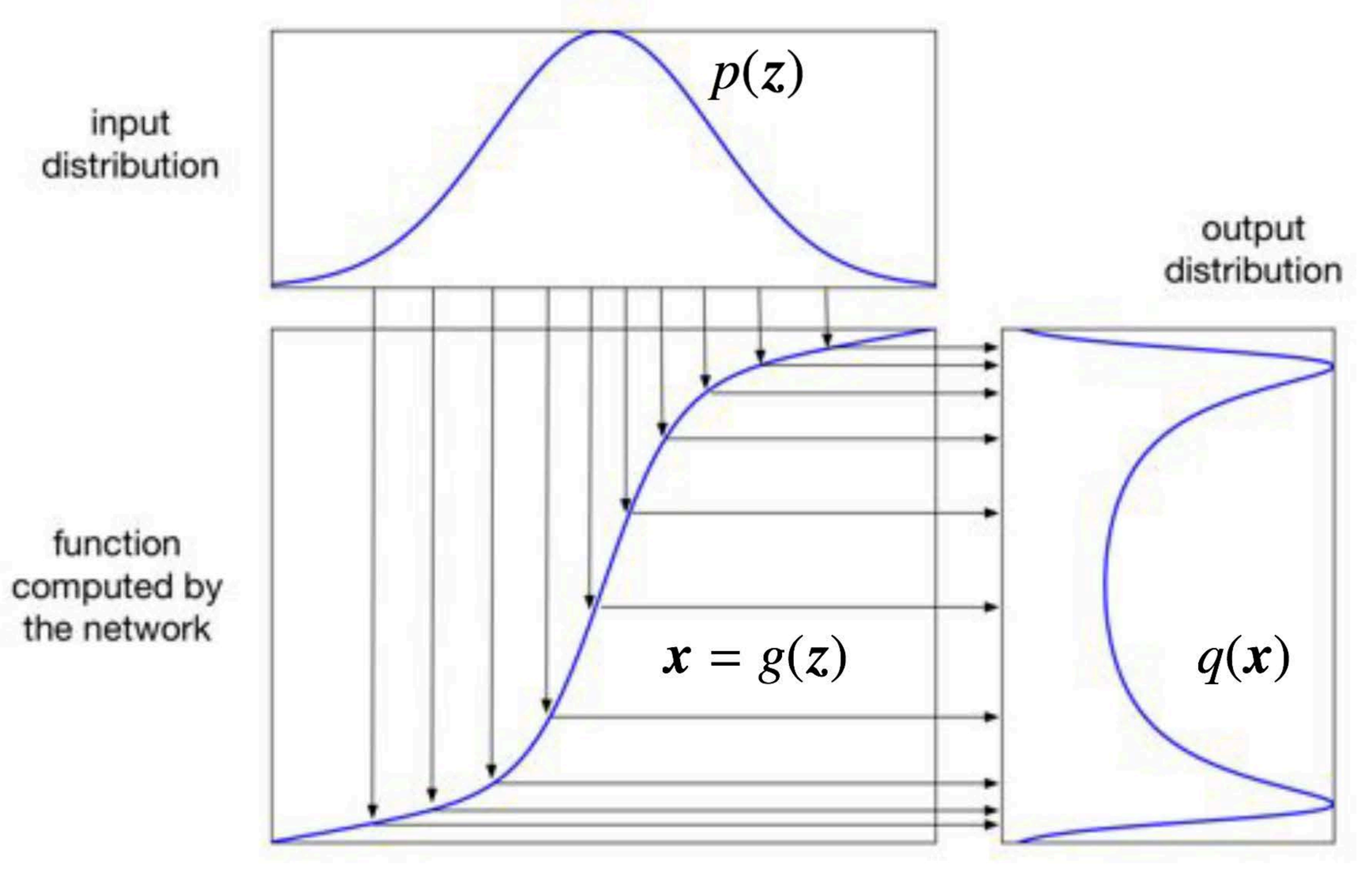
Latent variables

Bijective neural nets

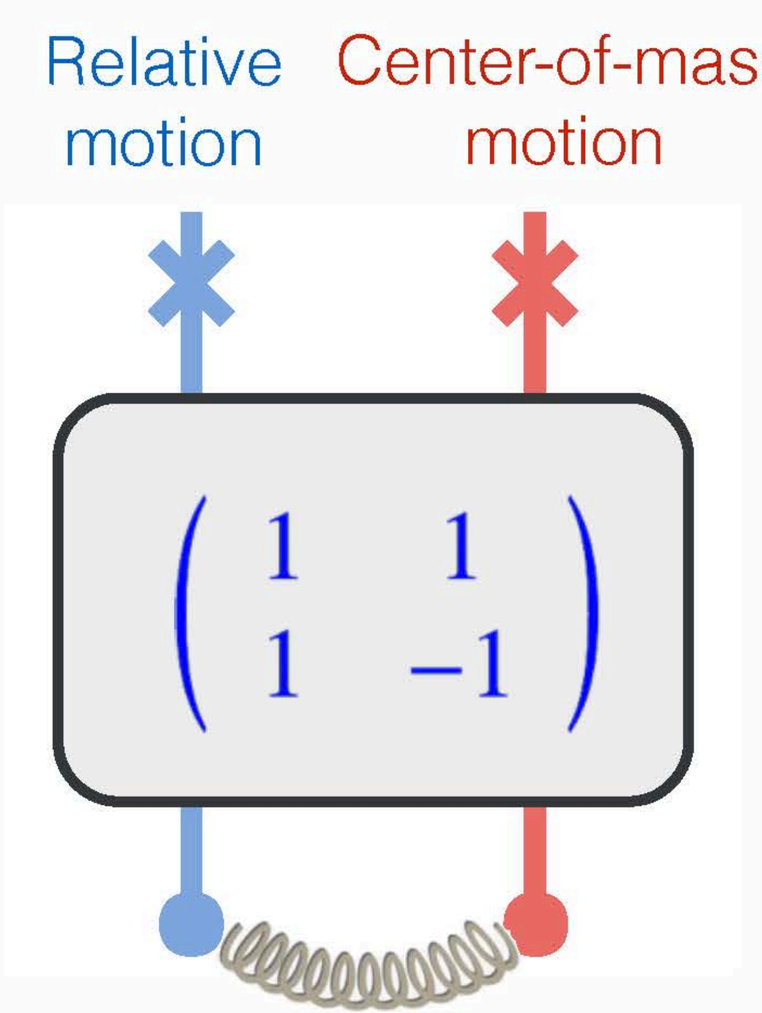
$$x = g(z)$$

Correlated classical variables

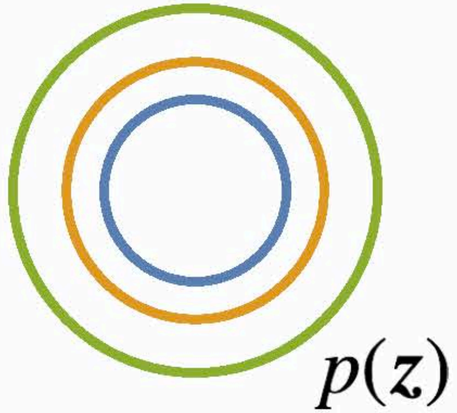
Probability transformation in picture

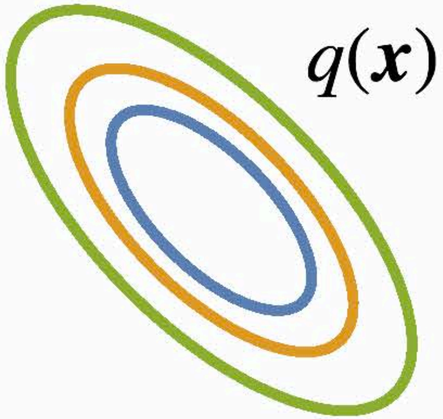


Toy problem: Harmonic oscillator



Coupled harmonic oscillator





Neural Bijectors

Bijective & Differentiable map, i.e., Diffeomorphism

Forward

Arbitrary neural nets

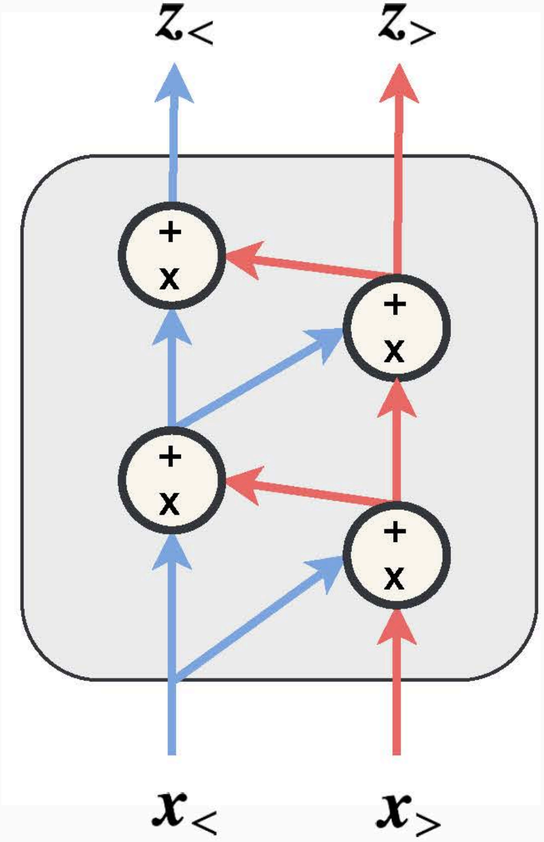
$$\begin{cases} \boldsymbol{x}_< = \boldsymbol{z}_< \\ \boldsymbol{x}_> = \boldsymbol{z}_> \odot e^{s(\boldsymbol{z}_<)} + t(\boldsymbol{z}_<) \end{cases}$$

Inverse

$$\begin{cases} \boldsymbol{z}_< = \boldsymbol{x}_< \\ \boldsymbol{z}_> = (\boldsymbol{x}_> - t(\boldsymbol{x}_<)) \odot e^{-s(\boldsymbol{x}_<)} \end{cases}$$

Log-Abs-Jacobian-Det

$$\ln \left| \det \left(\frac{\partial \boldsymbol{x}}{\partial \boldsymbol{z}} \right) \right| = \sum_i [s(\boldsymbol{z}_<)]_i$$



Normalizing flow, Rezende et al, 1505.05770

Real NVP, Dinh et al, 1605.08803

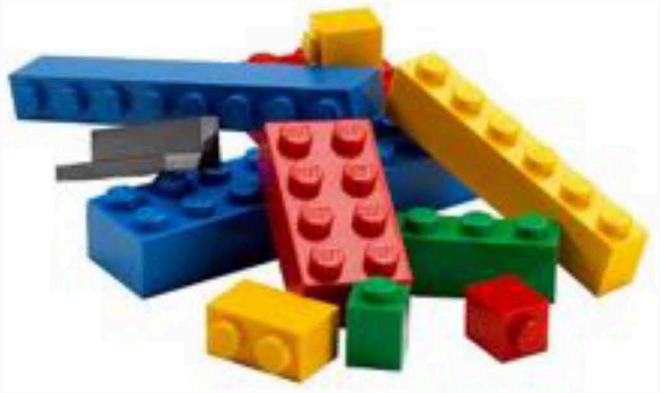
https://www.tensorflow.org/api_docs/python/tf/distributions/bijectors/Bijector

http://pytorch.org/docs/master/distributions.html#transformeddistribution

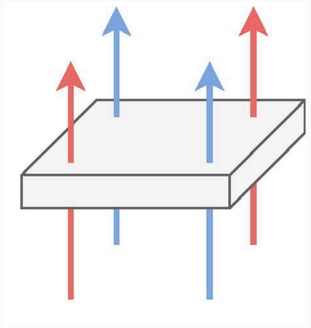
Bijectors form a group

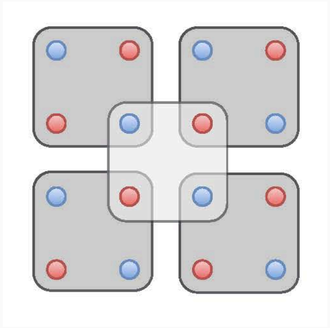
$$\boldsymbol{x} = g(\boldsymbol{z})$$
$$g = \cdots \circ g_2 \circ g_1$$

$$\ln \left| \det \left(\frac{\partial \boldsymbol{x}}{\partial \boldsymbol{z}} \right) \right| = \sum_i \ln \left| \det \left(\frac{\partial g_{i+1}}{\partial g_i} \right) \right|$$



Modular design





Flexible structure

Training: Probability Density *Distillation*

Minimizing the variational free energy

Parallel WaveNet
1711.10433

$$\mathcal{L}_{\theta} = \int \mathrm{d}\boldsymbol{x}\, q_{\theta}(\boldsymbol{x}) \left[\ln q_{\theta}(\boldsymbol{x}) + E(\boldsymbol{x})\right]$$

Training: Probability Density *Distillation*

Minimizing the variational free energy

Parallel WaveNet
1711.10433

$$\mathcal{L}_{\theta} = \int \mathrm{d}\boldsymbol{x}\, q_{\theta}(\boldsymbol{x}) \left[\ln q_{\theta}(\boldsymbol{x}) + E(\boldsymbol{x})\right]$$

Energy
(input)

Training: Probability Density *Distillation*

Minimizing the variational free energy

Parallel WaveNet
1711.10433

$$\mathcal{L}_{\theta} = \int \mathrm{d}\boldsymbol{x}\, q_{\theta}(\boldsymbol{x}) \left[\ln q_{\theta}(\boldsymbol{x}) + E(\boldsymbol{x})\right]$$

Entropy (ansatz)

Energy (input)

Training: Probability Density *Distillation*

Minimizing the variational free energy

Parallel WaveNet
1711.10433

$$\mathcal{L}_{\boldsymbol{\theta}} = \int \mathrm{d}\boldsymbol{x}\, q_{\boldsymbol{\theta}}(\boldsymbol{x}) \left[\ln q_{\boldsymbol{\theta}}(\boldsymbol{x}) + E(\boldsymbol{x})\right]$$

"Learn from the samples generated by the network itself!"

Entropy (ansatz)

Energy (input)

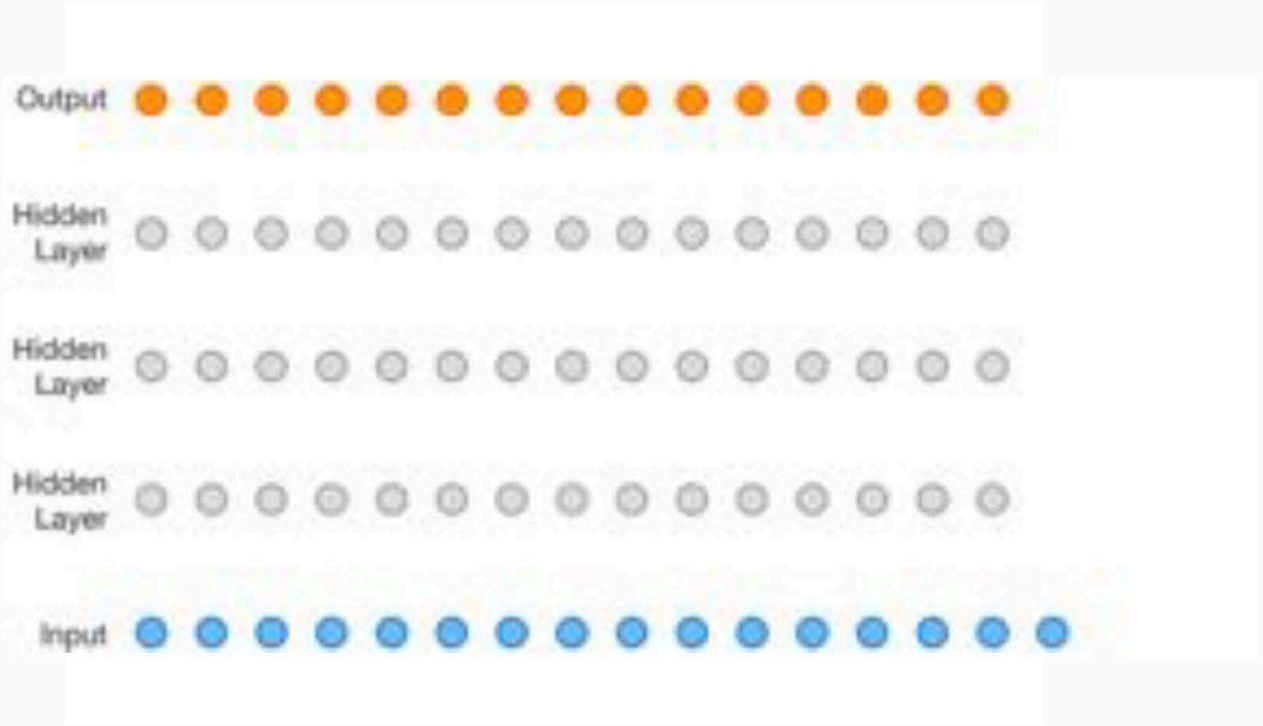
$$\mathcal{L}_{\boldsymbol{\theta}} + \ln Z = \mathbb{KL}\left(q_{\boldsymbol{\theta}}(\boldsymbol{x}) \,\middle\|\, \frac{e^{-E(\boldsymbol{x})}}{Z}\right) \geq 0$$

Reverse Kullback–Leibler divergence

The loss function is lower bounded by the physical free energy (Gibbs-Bogoliubov-Feynman inequality)

Interlude: The WaveNet Story

WaveNet 2016
Autoregressive Flow

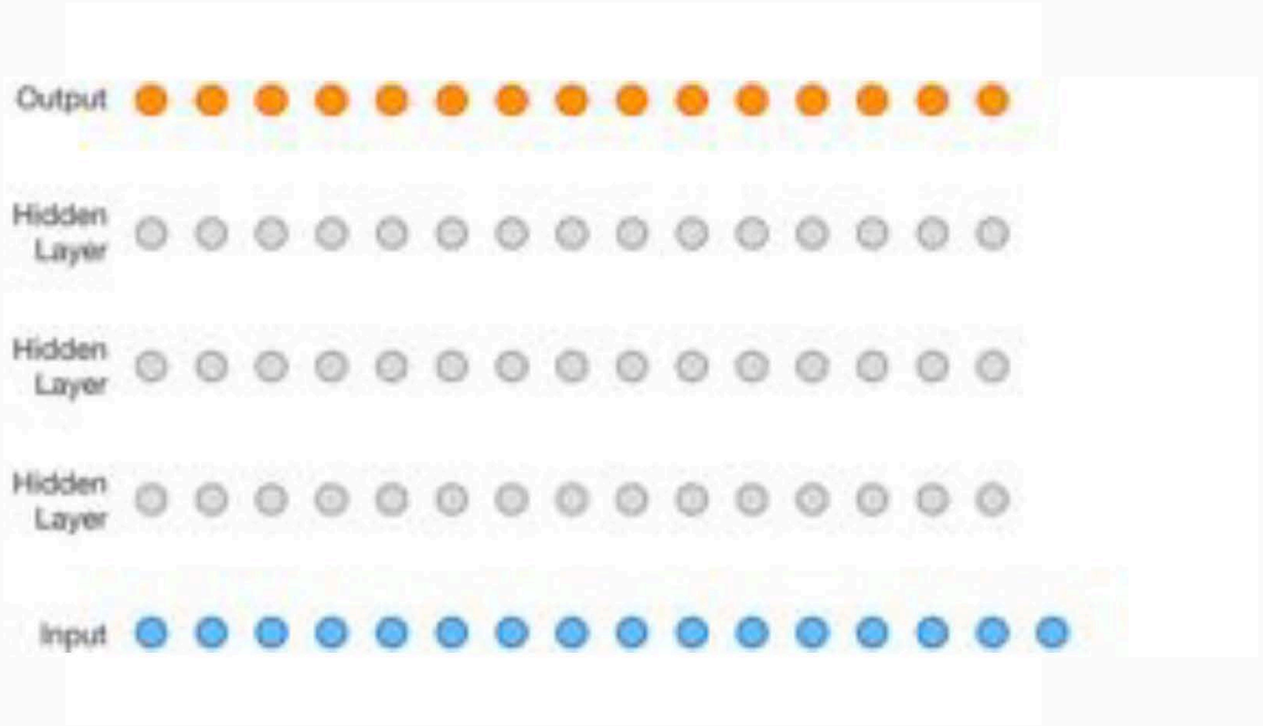


waveforms. The model is fully probabilistic and autoregressive, with the predictive distribution for each audio sample conditioned on all previous ones; nonetheless we show that it can be efficiently trained on data with tens of thousands of samples per second of audio. When applied to text-to-speech, it yields state-of-

https://deepmind.com/blog/wavenet-generative-model-raw-audio/ 1609.03499
https://deepmind.com/blog/high-fidelity-speech-synthesis-wavenet/ 1711.10433

Interlude: The WaveNet Story

WaveNet 2016
Autoregressive Flow



waveforms. The model is fully probabilistic and autoregressive, with the predictive distribution for each audio sample conditioned on all previous ones; nonetheless we show that it can be efficiently trained on data with tens of thousands of samples per second of audio. When applied to text-to-speech, it yields state-of-

https://deepmind.com/blog/wavenet-generative-model-raw-audio/ 1609.03499
https://deepmind.com/blog/high-fidelity-speech-synthesis-wavenet/ 1711.10433

Interlude: The WaveNet Story

speech signal

Parallel WaveNet 2017
Inverse Autoregressive Flow

input noise

Given a parallel WaveNet student $p_S(\boldsymbol{x})$ and WaveNet teacher $p_T(\boldsymbol{x})$ which has been trained on a dataset of audio, we define the *Probability Density Distillation* loss as follows:

$$D_{\mathrm{KL}}\left(P_S||P_T\right) = H(P_S, P_T) - H(P_S) \tag{6}$$

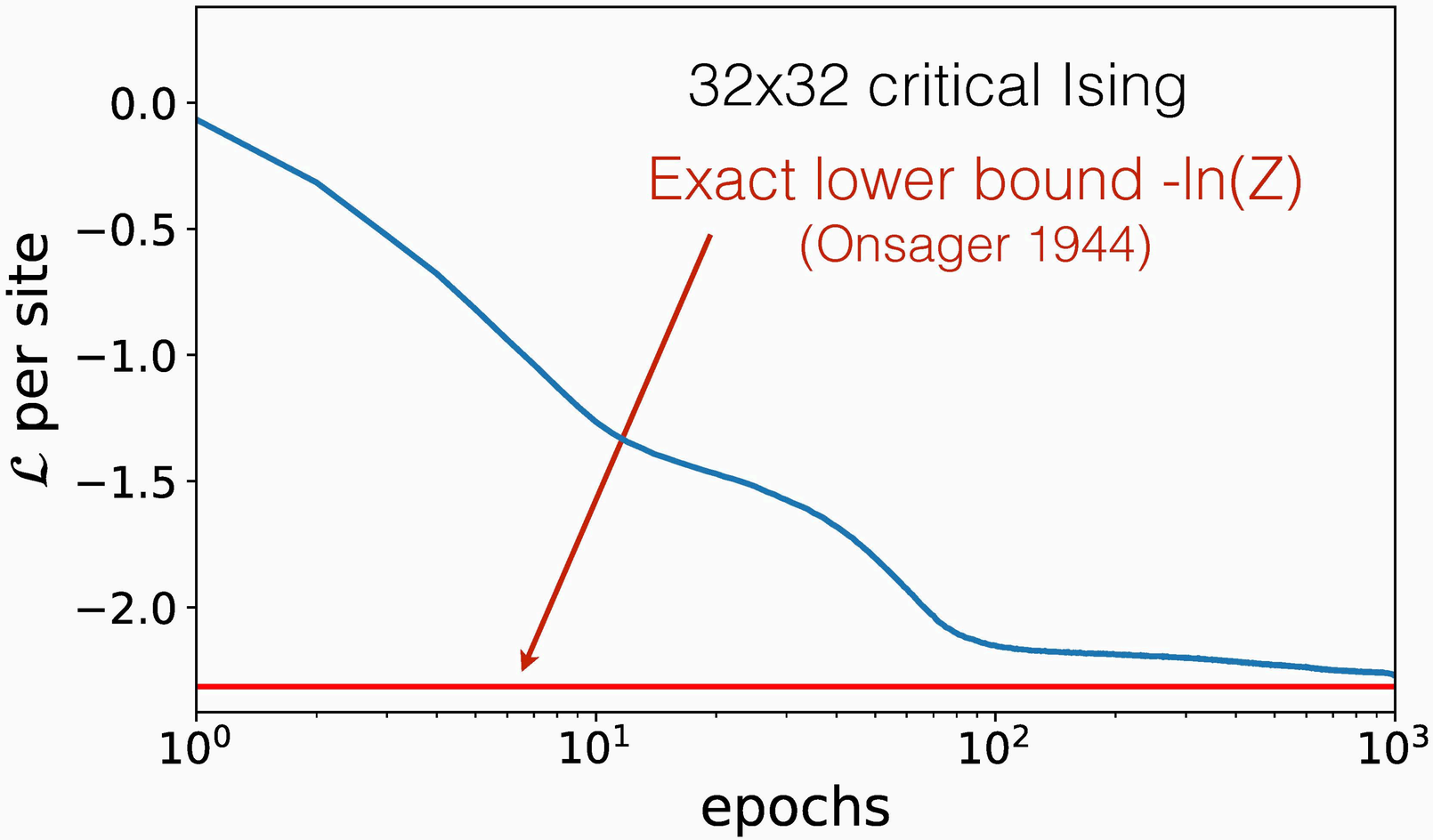
https://deepmind.com/blog/wavenet-generative-model-raw-audio/ 1609.03499
https://deepmind.com/blog/high-fidelity-speech-synthesis-wavenet/ 1711.10433

*Let's apply it to the Ising model**

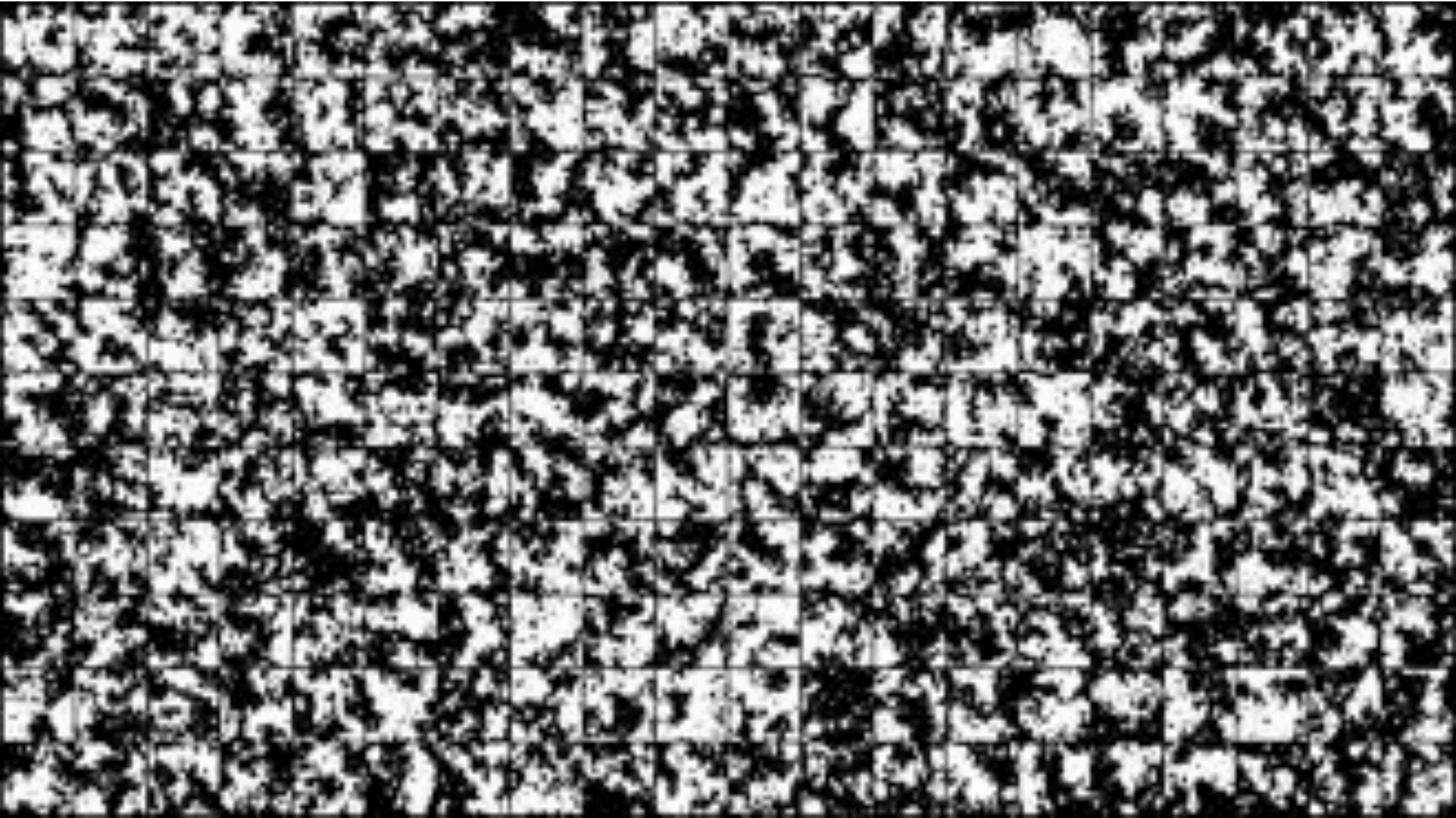
*actually a continuous dual, think of scalar ϕ^4 theory

M. E. Fisher 1983 Zhang, Sutton, Storkey, Ghahramani, NIPS 2012

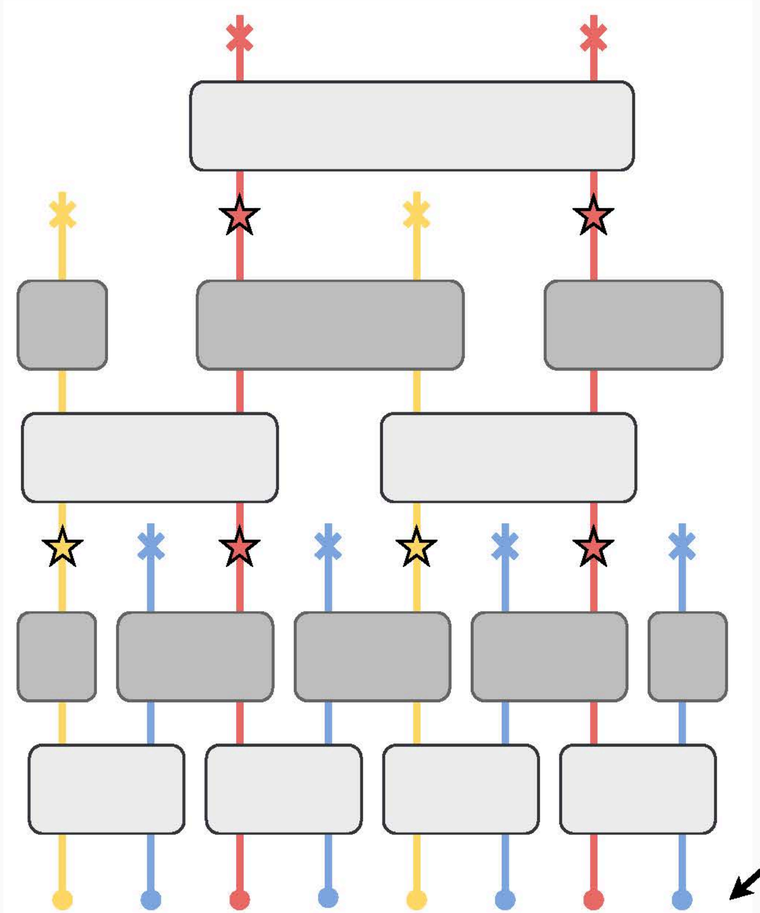
Variational Loss



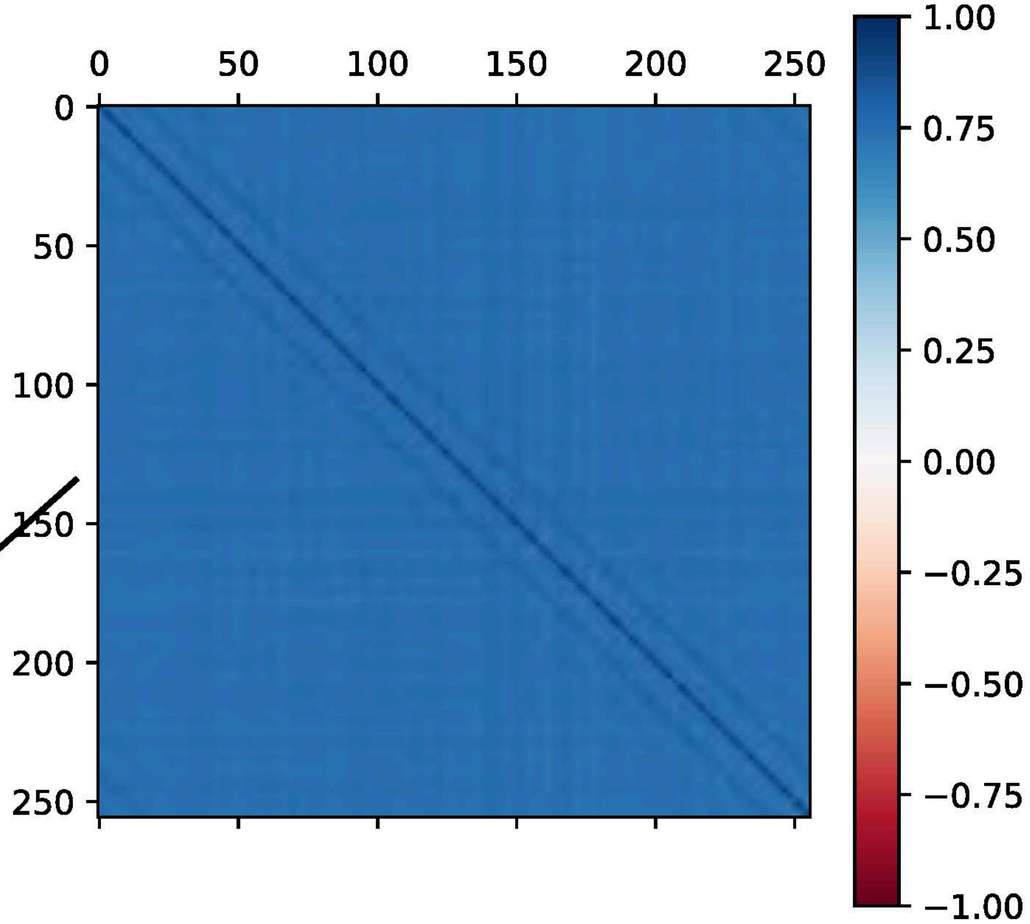
Training = Variational free energy calculation



What is the network learning?



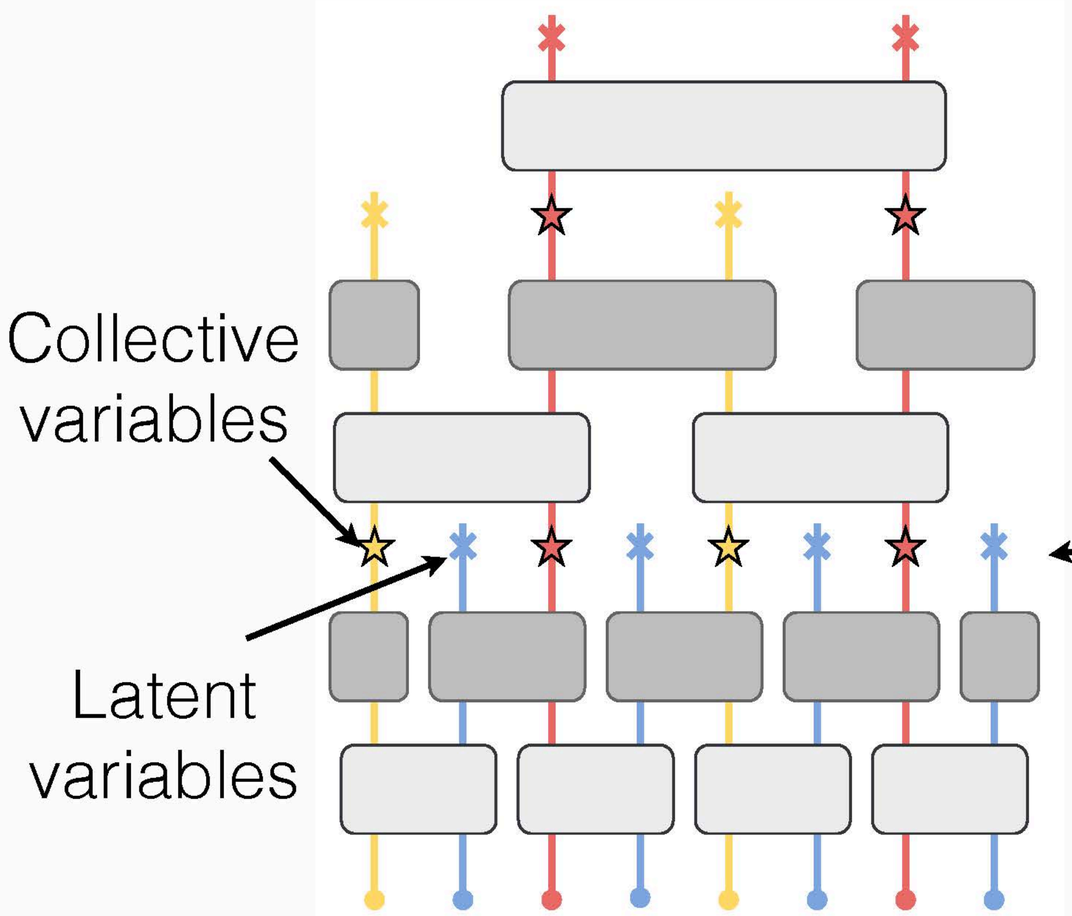
Physical variables



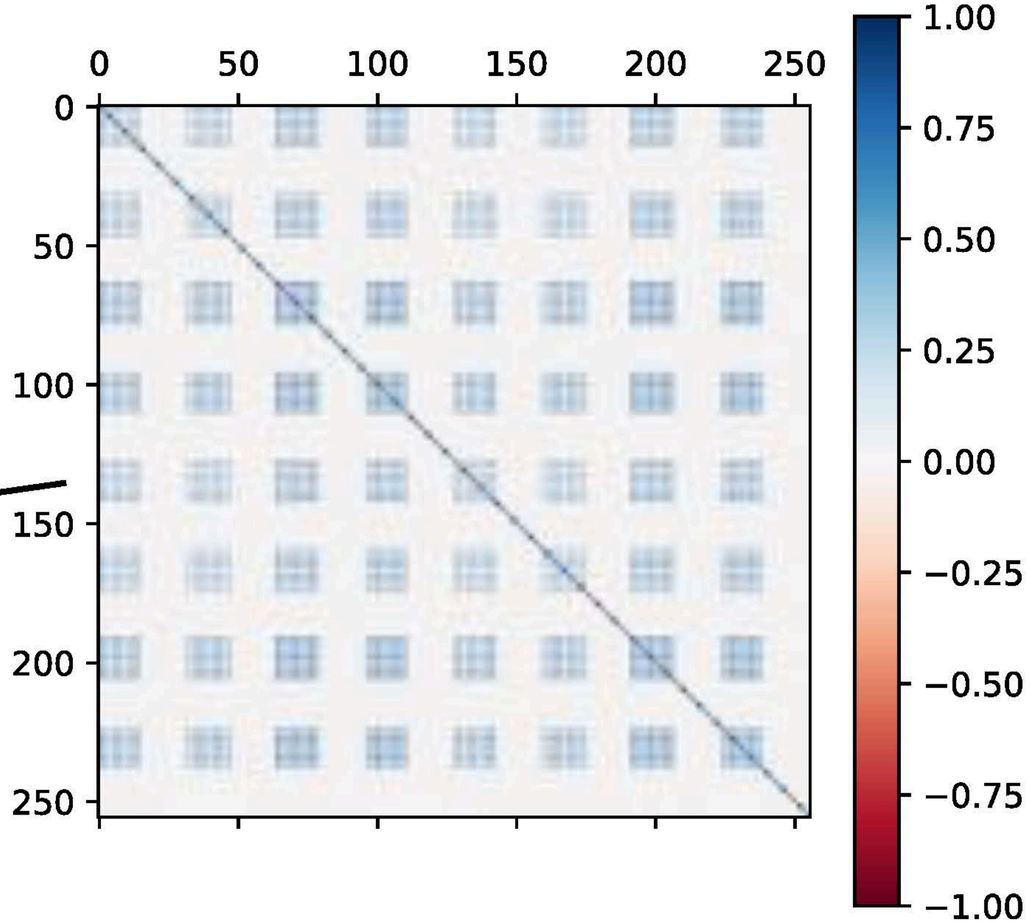
Two-point correlations

Progressively decouple degrees of freedom into Gaussian noise

What is the network learning?



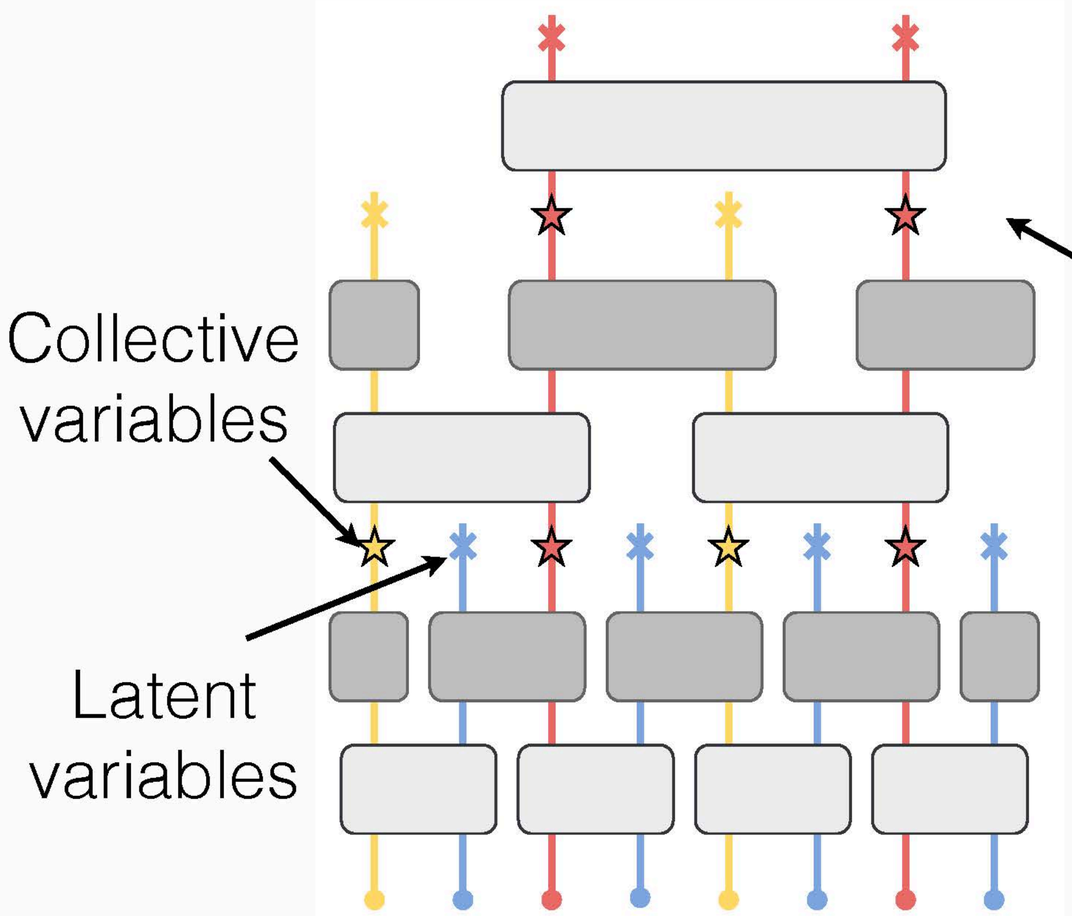
Physical variables



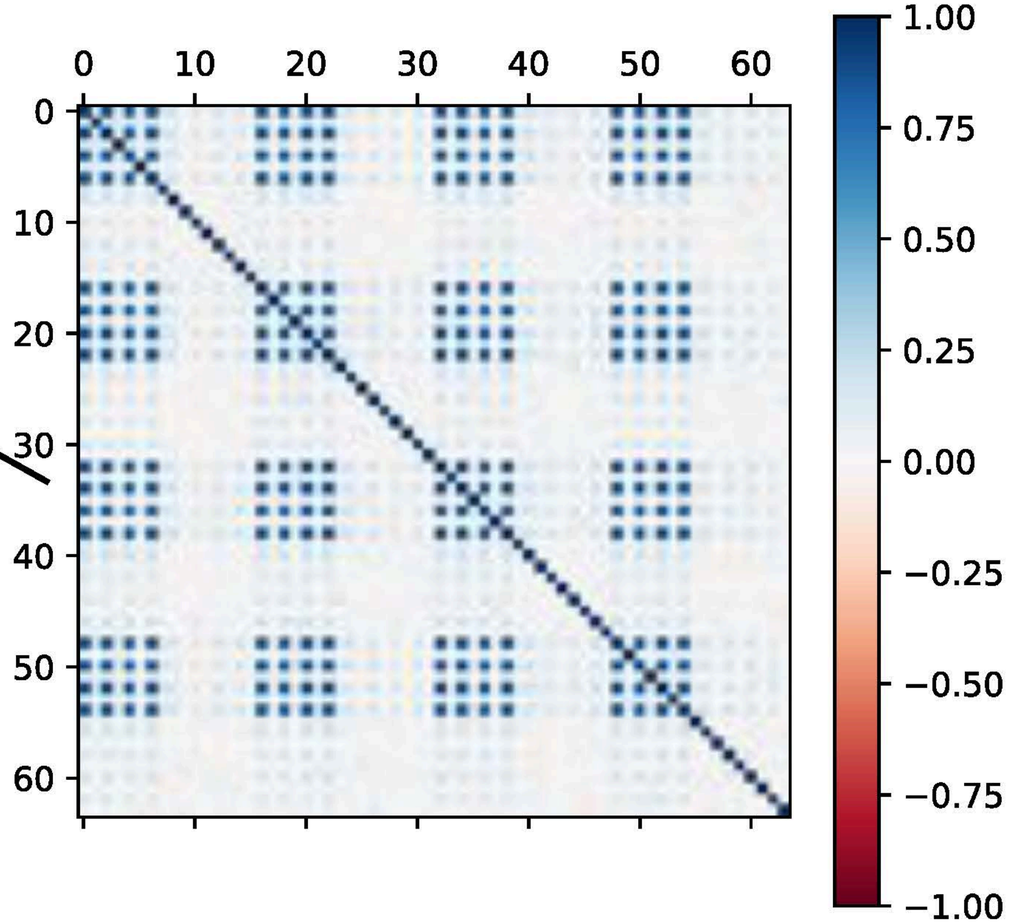
Two-point correlations

Progressively decouple degrees of freedom into Gaussian noise

What is the network learning?



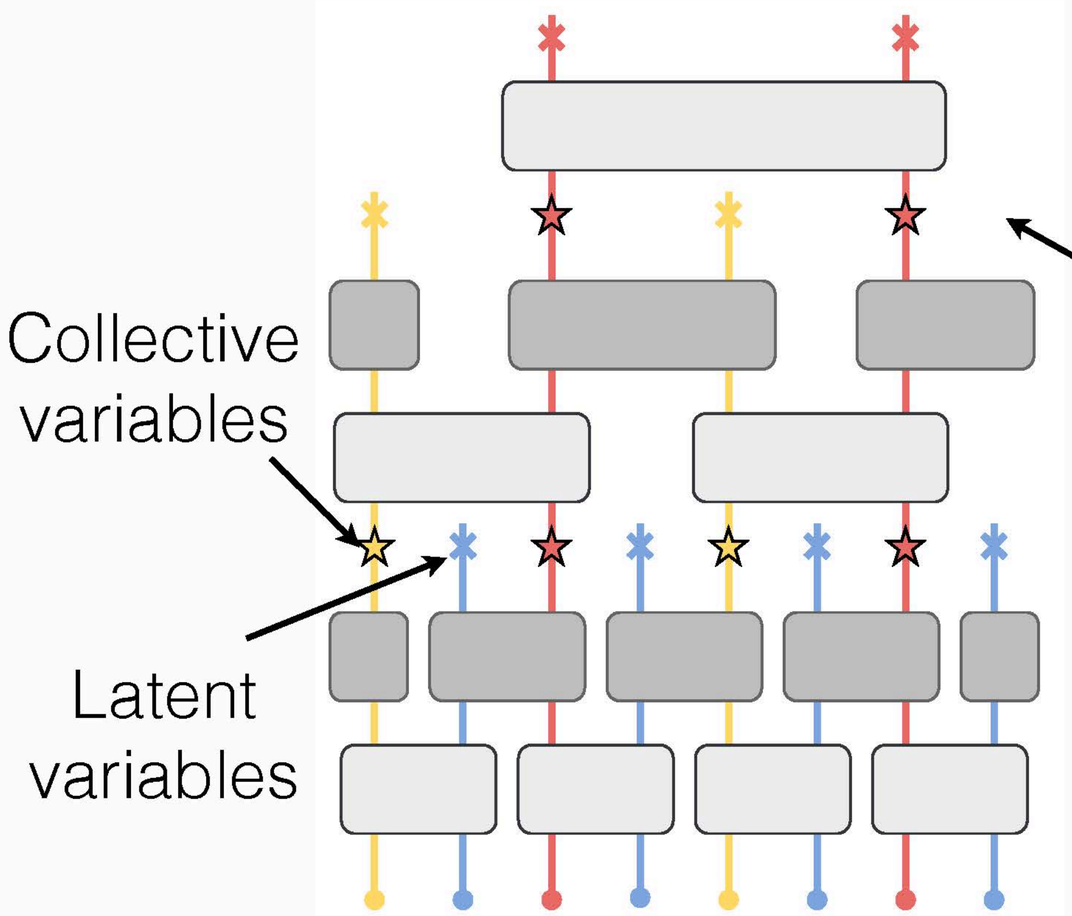
Physical variables



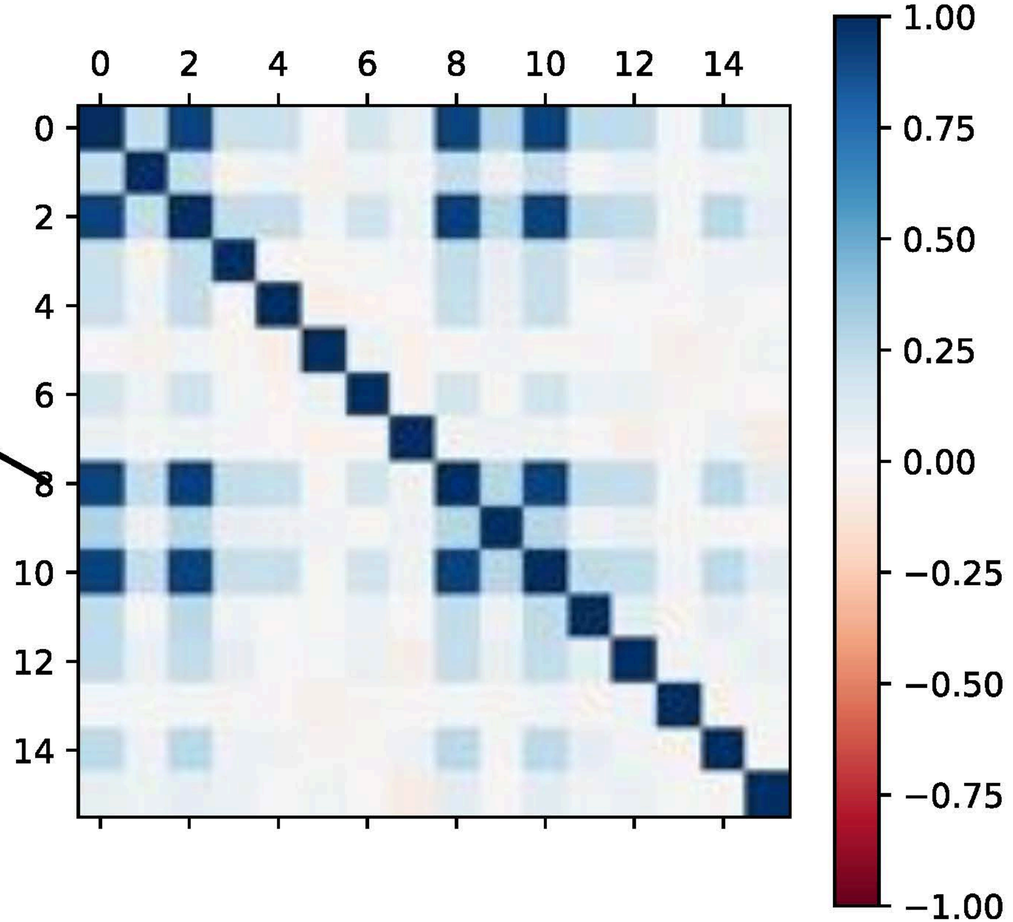
Two-point correlations

Progressively decouple degrees of freedom into Gaussian noise

What is the network learning?



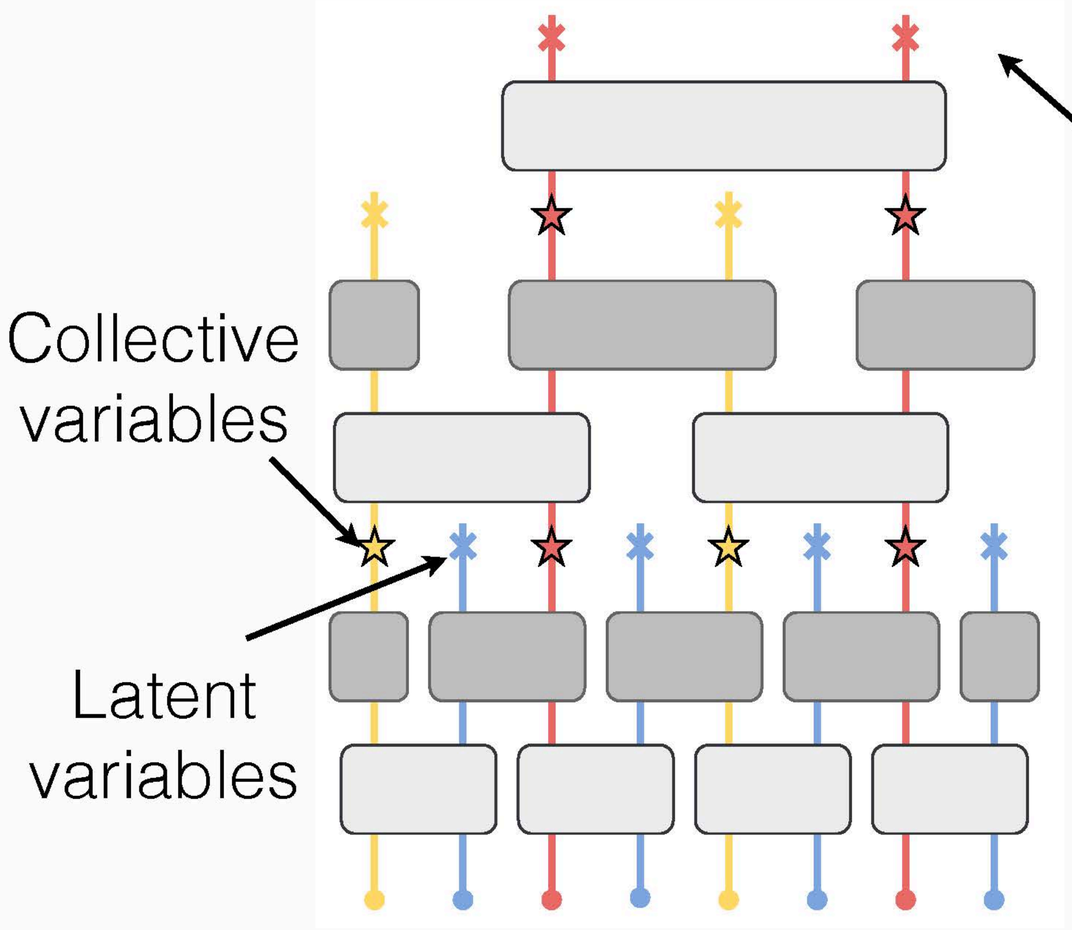
Physical variables



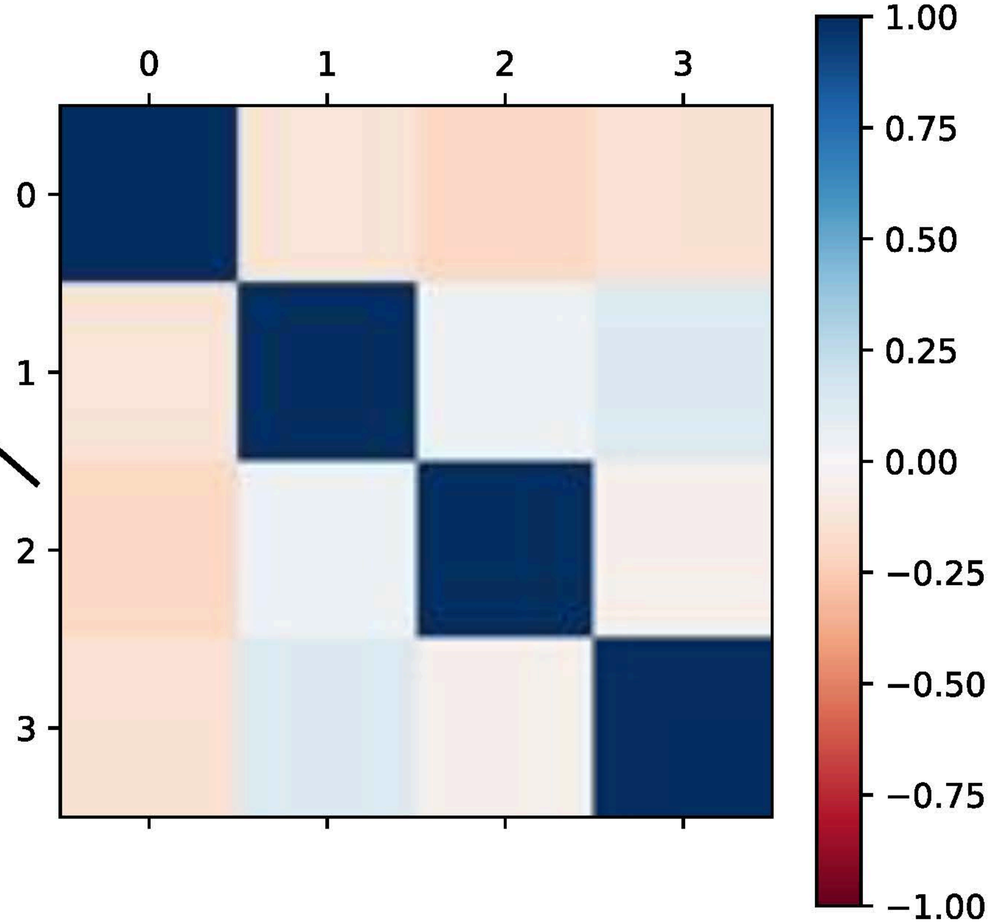
Two-point correlations

Progressively decouple degrees of freedom into Gaussian noise

What is the network learning?



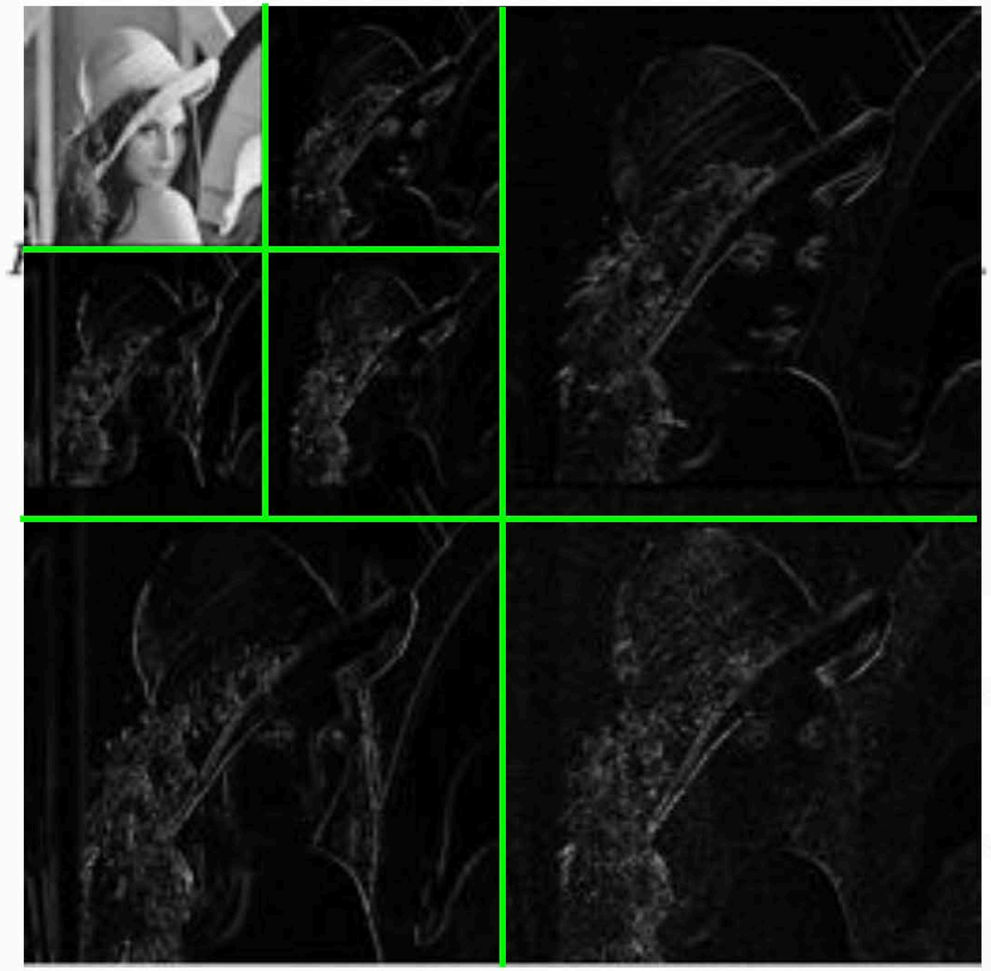
Physical variables

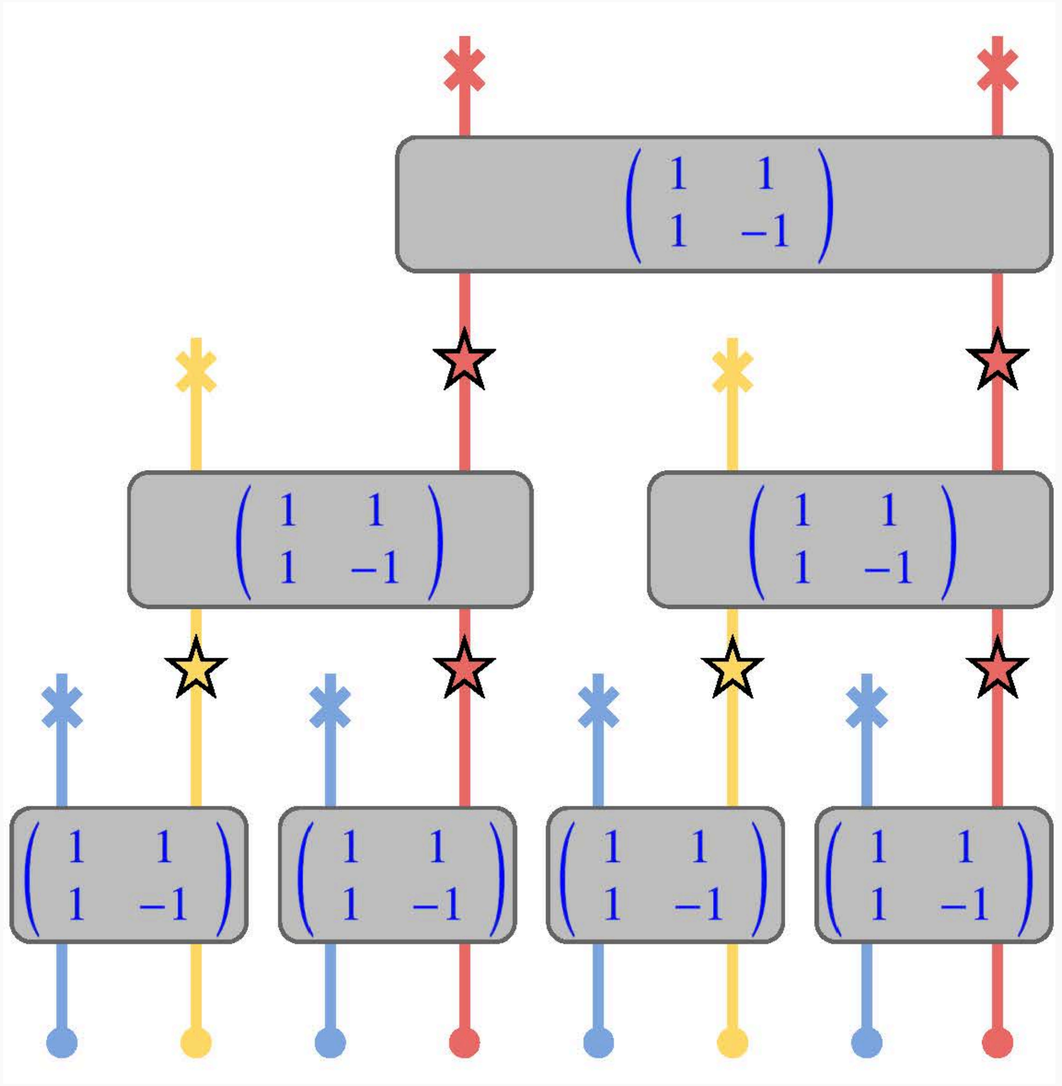


Two-point correlations

Progressively decouple degrees of freedom into Gaussian noise

How to interpret the latent variables ?

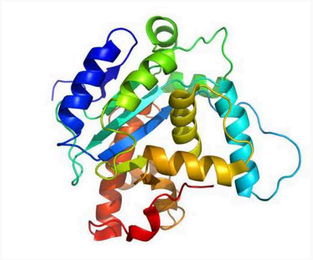




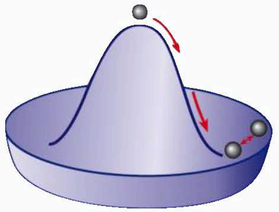
Nonlinear & adaptive generalizations of wavelets!

Guy, Wavelets & RG1999+ White, Evenbly, Qi, Wavelets, MERA, and holographic mapping 2013+

How is this useful ?

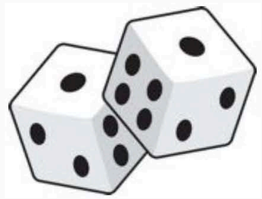


Identifying mutually independent collective variables
(molecular simulation, PIMC, PIMD)



Deriving effective field theory for collective variables

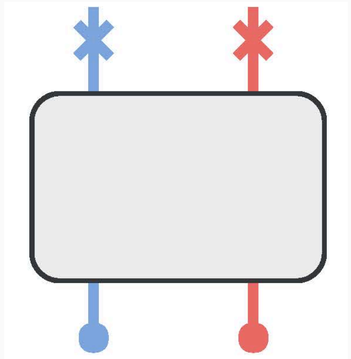
Information preserving RG for holographic mapping



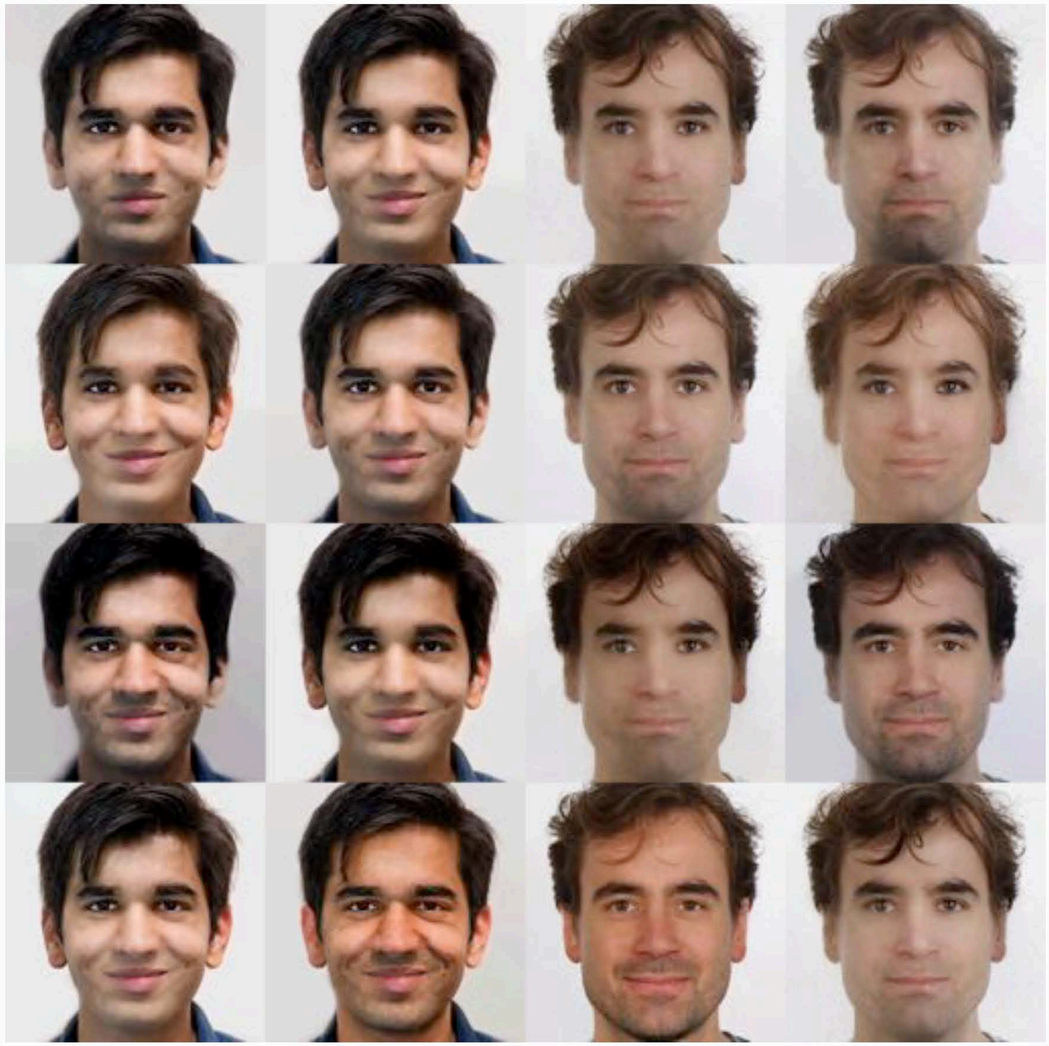
Accelerated Monte Carlo simulation

Wander in the latent space

Latent space



Physical space

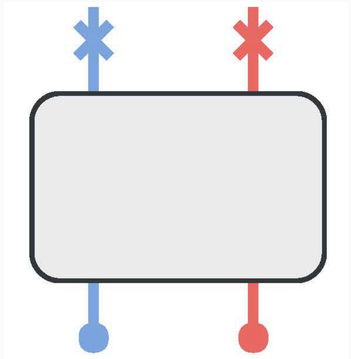


Glow 1807.03039

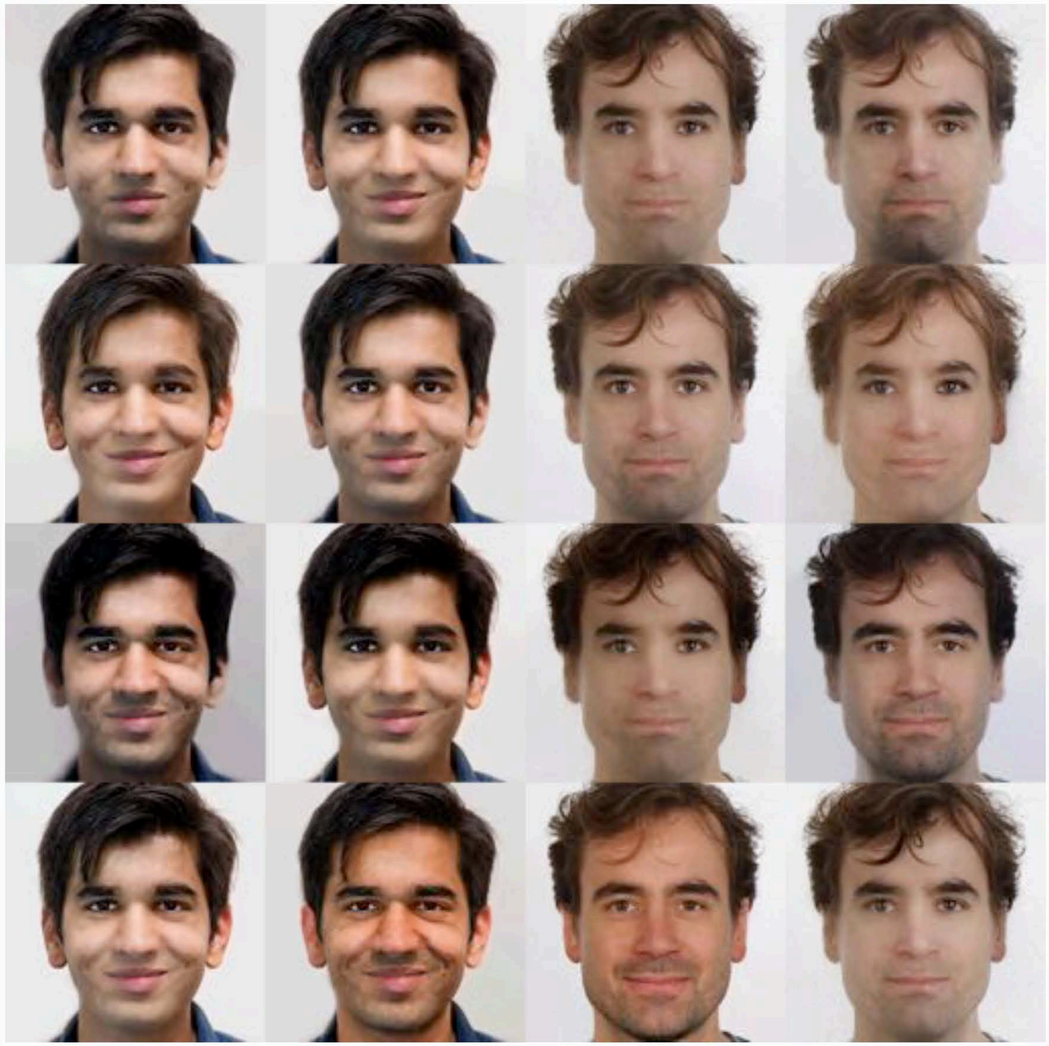
https://blog.openai.com/glow/

Wander in the latent space

Latent space



Physical space

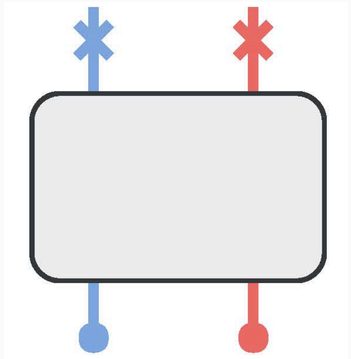


Glow 1807.03039

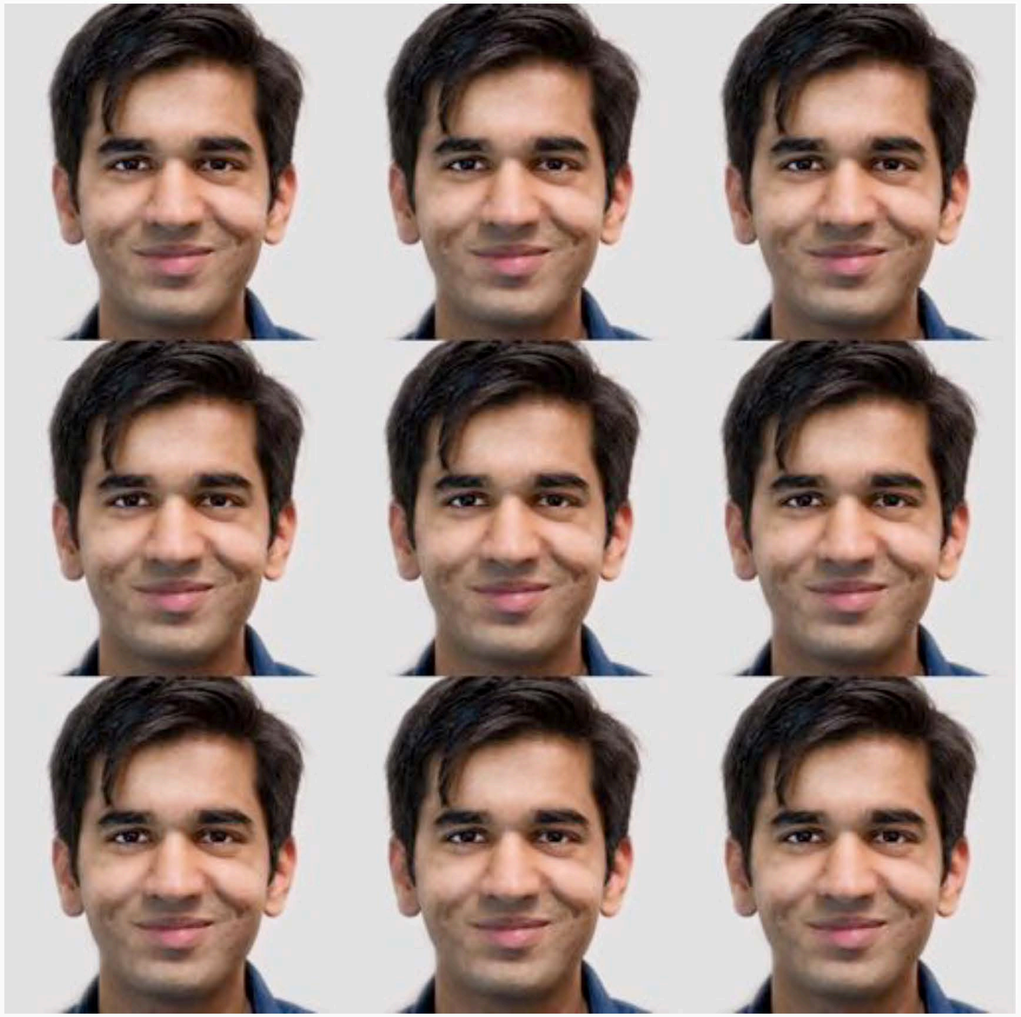
https://blog.openai.com/glow/

Wander in the latent space

Latent space



Physical space

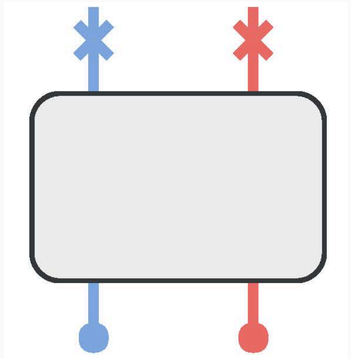


Glow 1807.03039

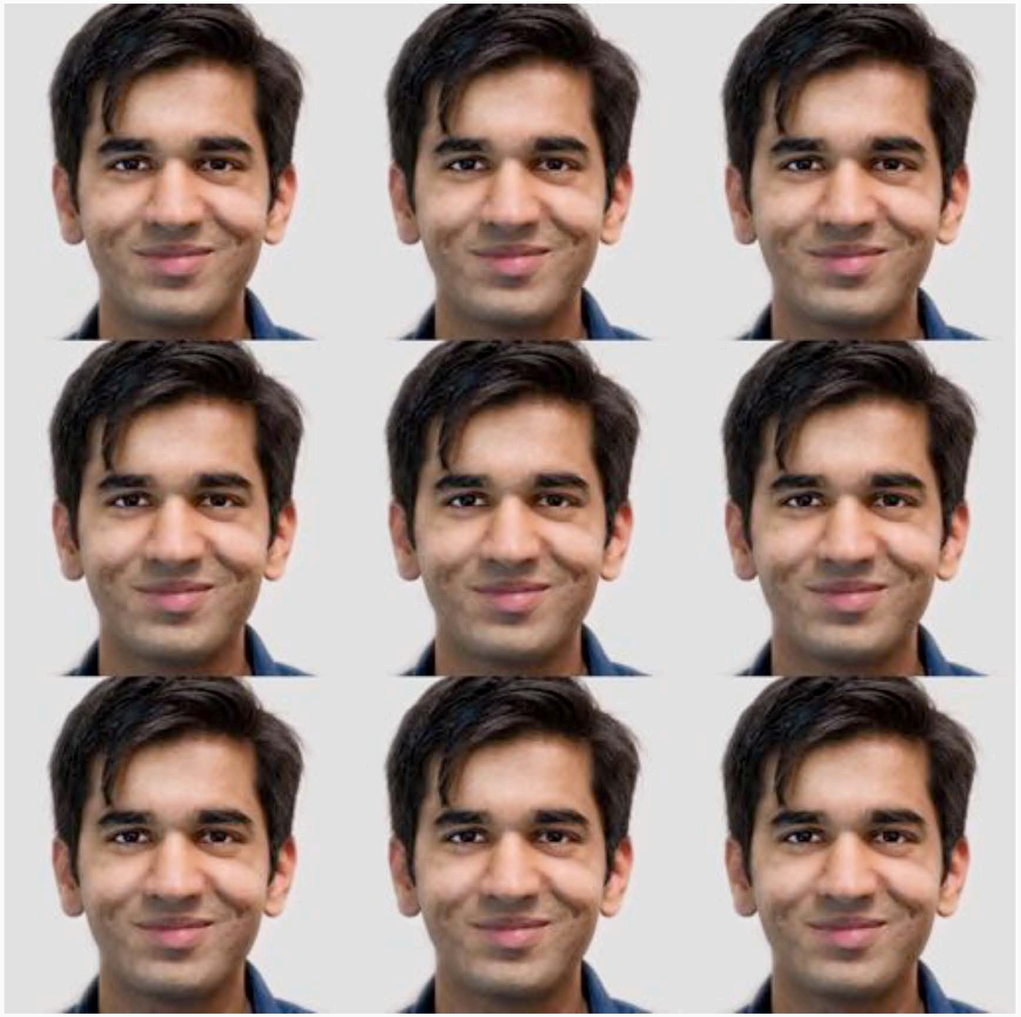
https://blog.openai.com/glow/

Wander in the latent space

Latent space



Physical space



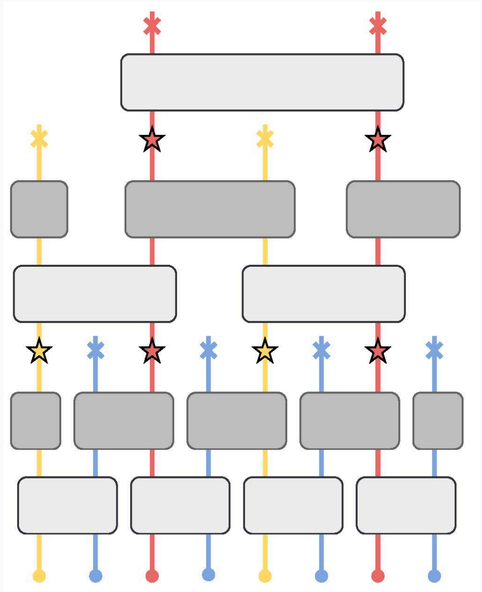
Glow 1807.03039

https://blog.openai.com/glow/

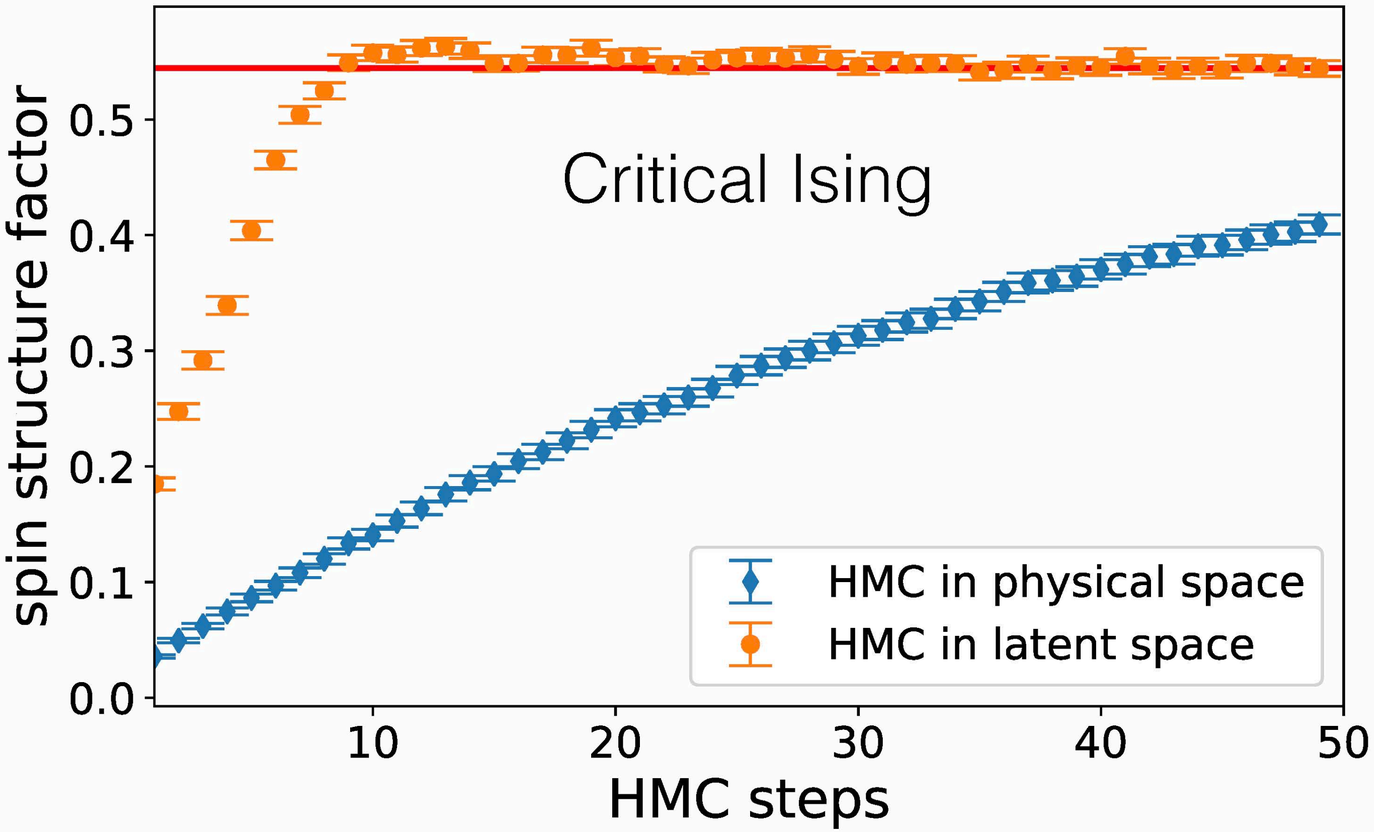
Latent space Hybrid MC

Latent space energy function

$$E_{\text{eff}}(z) = E(g(z)) + \ln q(g(z)) - \ln p(z)$$



Physical energy function $E(x)$

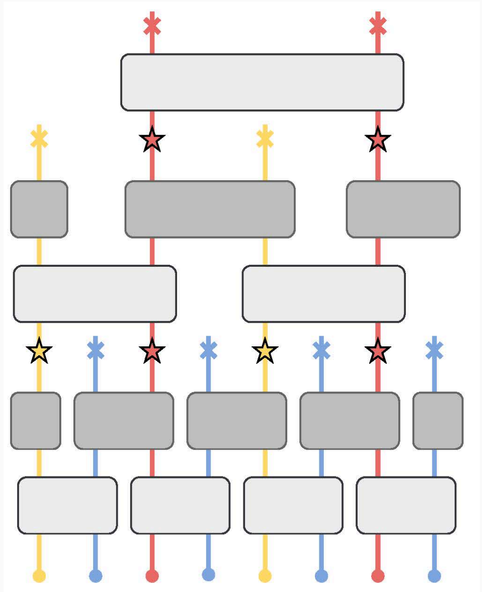


HMC thermalizes faster in the latent space

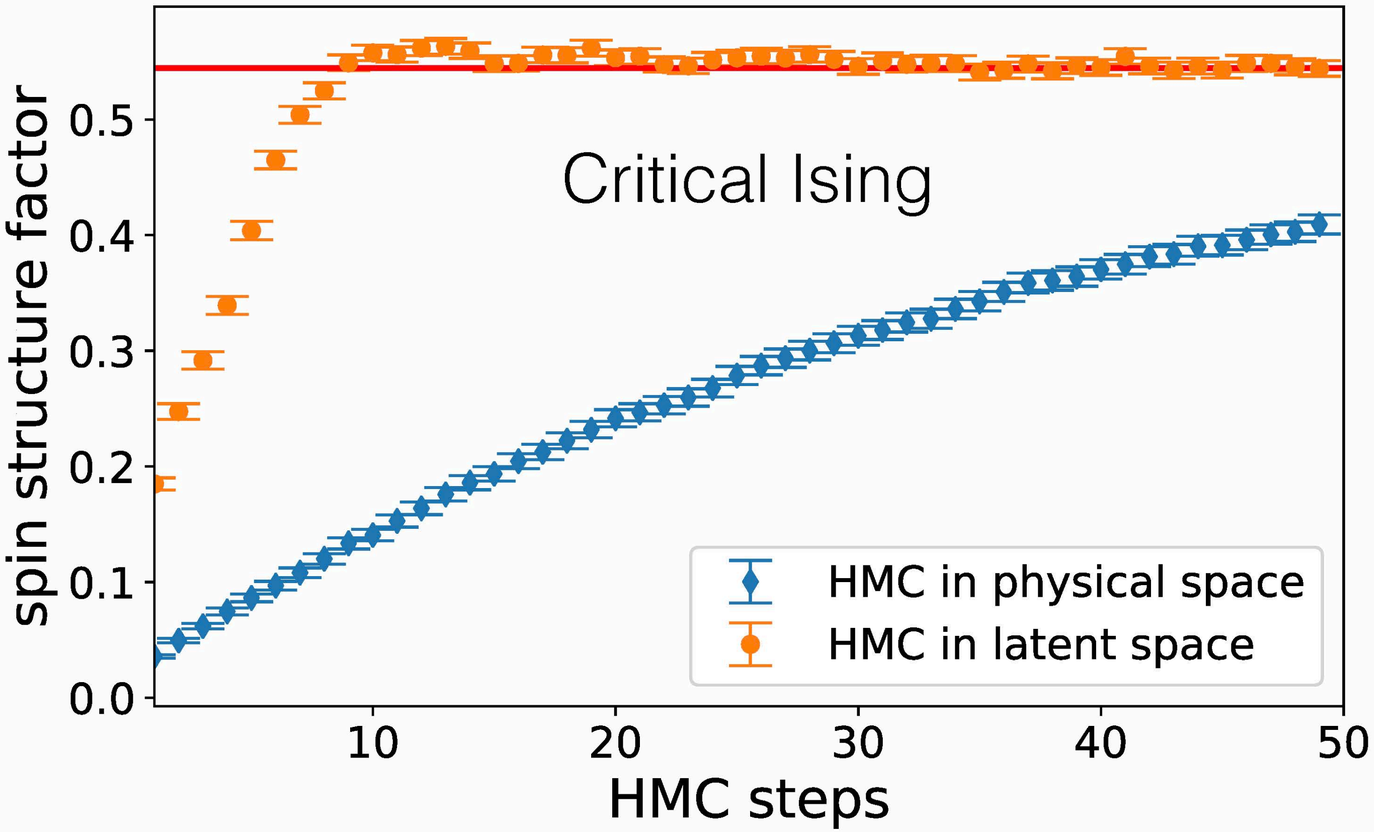
Latent space Hybrid MC

Latent space energy function

$$E_{\text{eff}}(z) = E(g(z)) + \ln q(g(z)) - \ln p(z)$$



Physical energy function $E(x)$

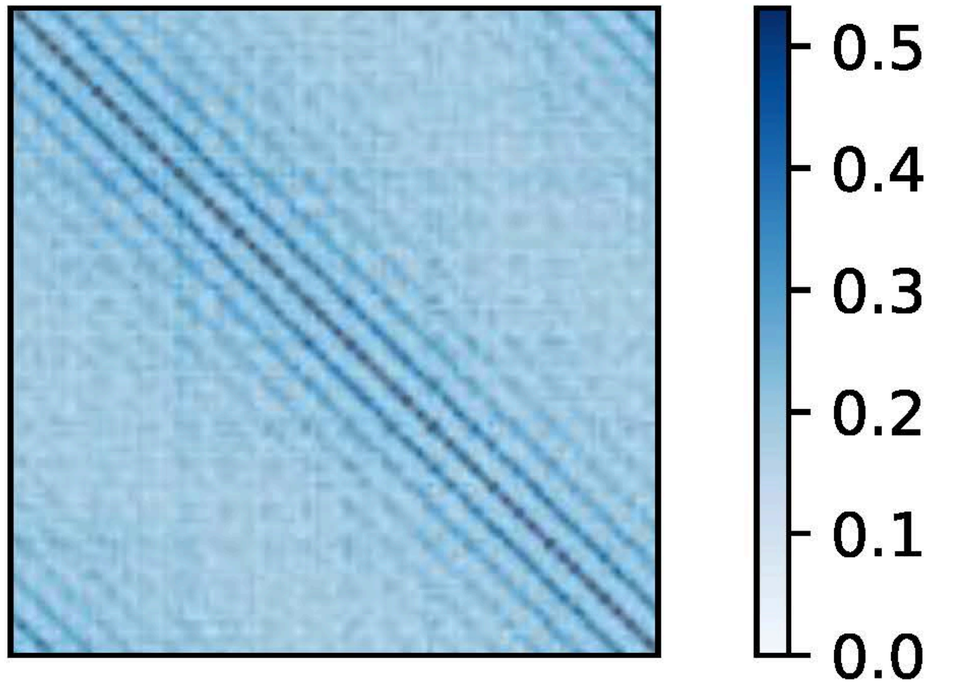


HMC thermalizes faster in the latent space

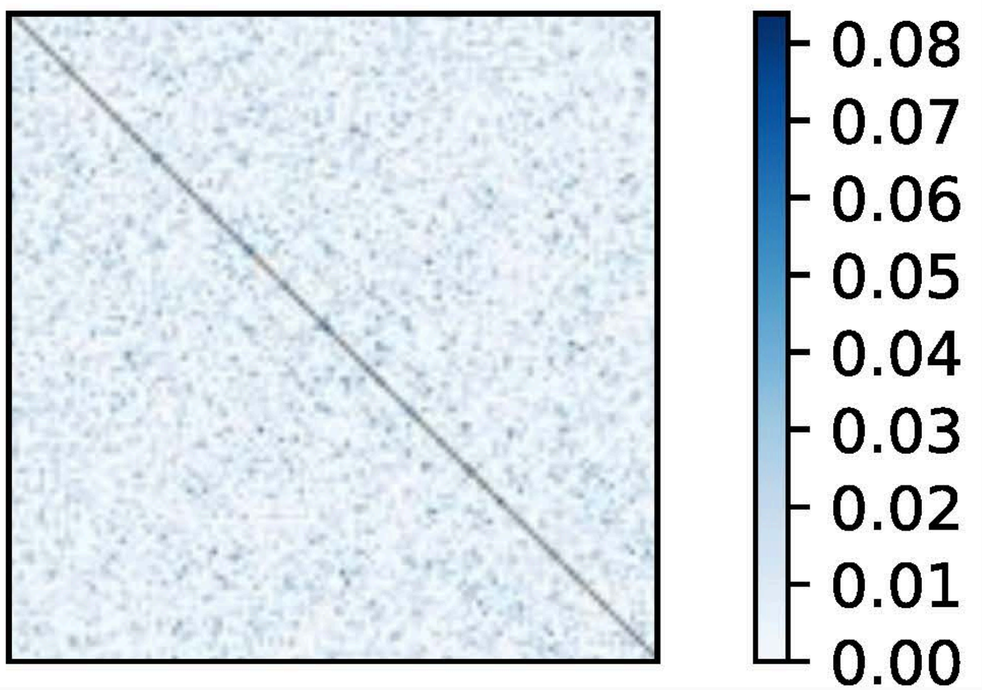
Mutual Information

KSG MI estimator
Phys. Rev. E 69, 066138 (2004)

$I(x_i : x_j)$

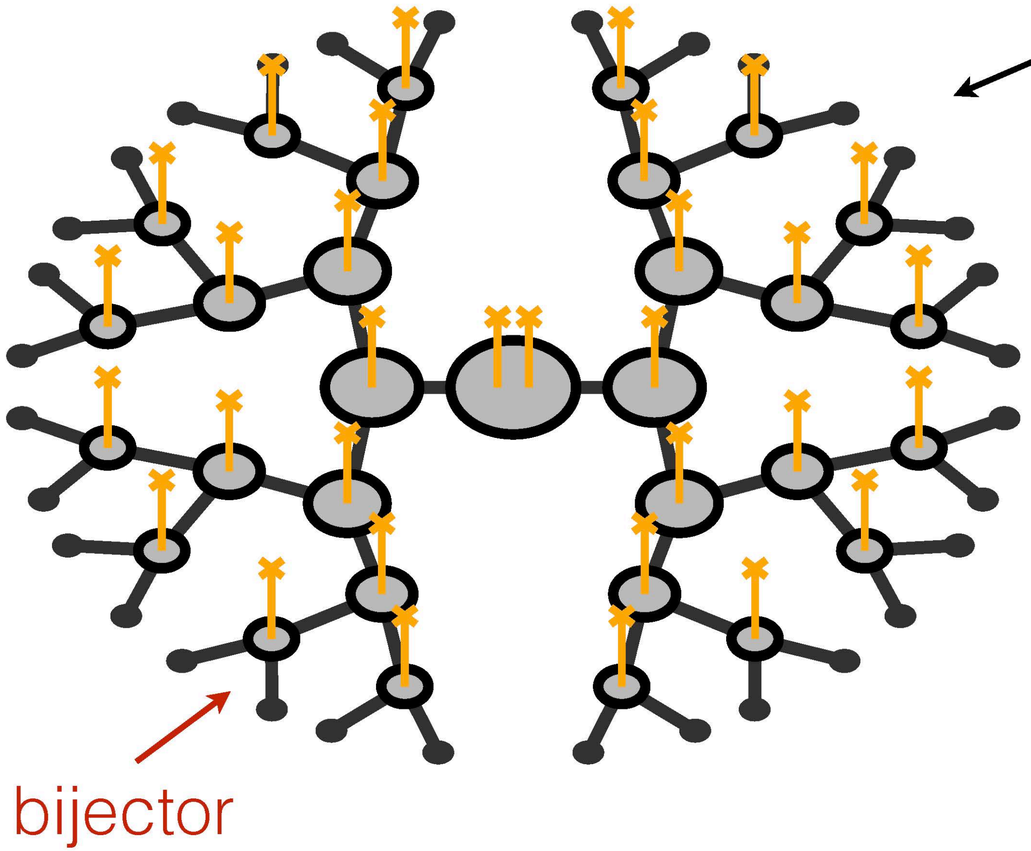


$I(z_i : z_j)$



Reduced Mutual Information in the latent space

MI and holographic RG



This is a neural network

Physical variables on the boundary

Latent variables in the bulk

RG flows along the radial direction

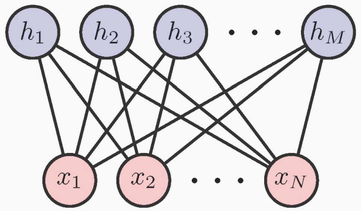
Information is preserved by the flow

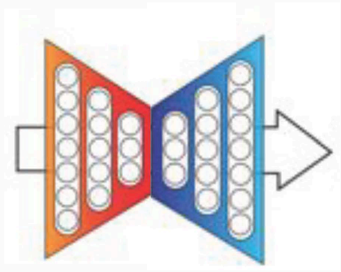
Swingle 0905.1317, Qi 1309.6282 and more

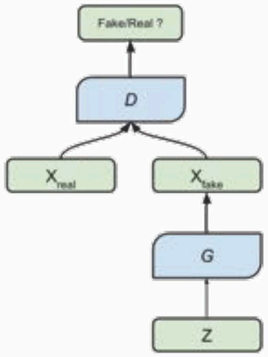
Normalizing flow implements an invertible RG flow

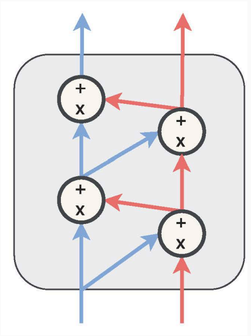
Mutual information reveals the emergent geometry in the bulk

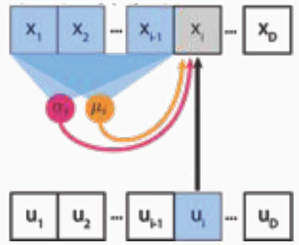
Timeline of Generative Models

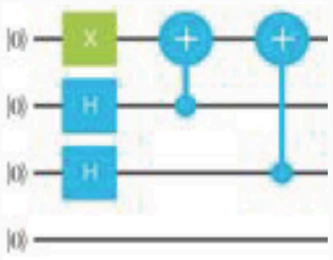












Boltzmann Machines 1980s

Variational Autoendoer 2013

Adversarial Network 2014

Normalizing Flows 2015

Autoregressive Flows 2016

Born Machines 2017

Han, Wang, Fan, LW, Zhang
PRX 8, 031012

1. **Leverage the power of modern generative models for physics**
2. **Statistical, quantum, and fluid physics inspired generative models**

DL as a fluid control problem

$$\frac{p(z)}{q(\nabla u(z))} = \det\left(\frac{\partial^2 u}{\partial z_i \partial z_j}\right)$$

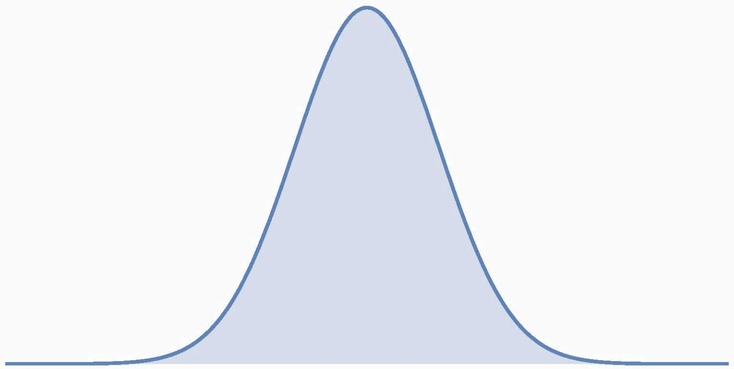
Monge-Ampère equation
optimal transport theory

Continuous-time limit

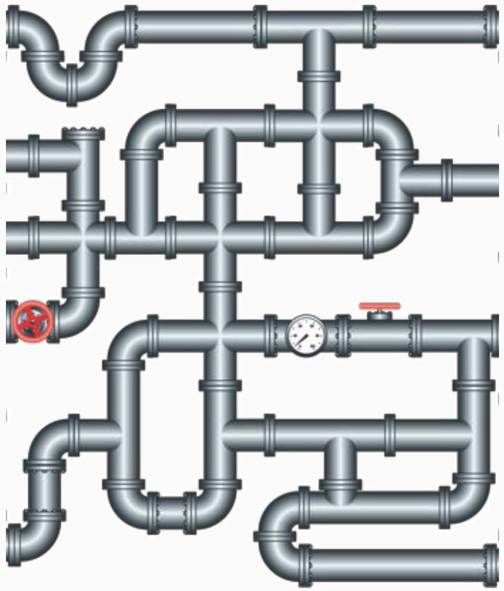
$$u(z) = |z|^2/2 + \epsilon\varphi(z)$$

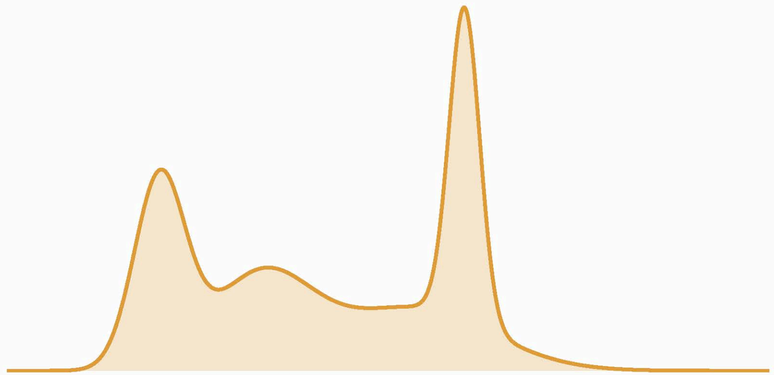
$$\frac{\partial p(\boldsymbol{x}, t)}{\partial t} + \nabla \cdot \left[p(\boldsymbol{x}, t)\nabla\varphi\right] = 0$$

Continuity equation of
compressible fluids



Simple density

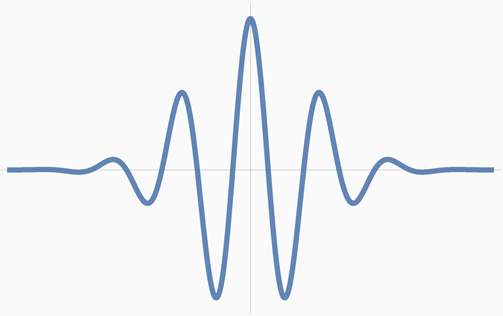


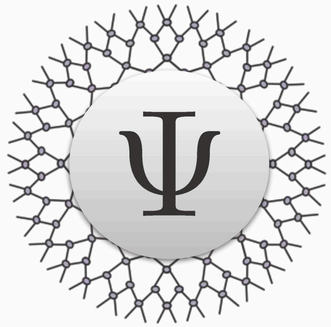


Complex density

Zhang, E, LW, 1809.10188
c.f. Neural ODE, 1806.07366

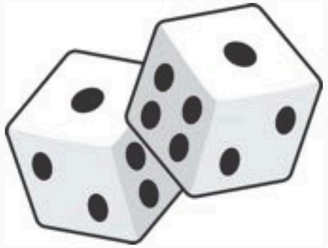
Wavelets

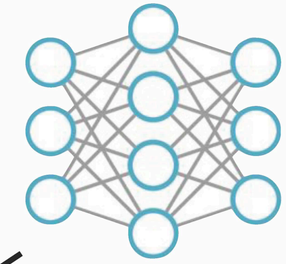


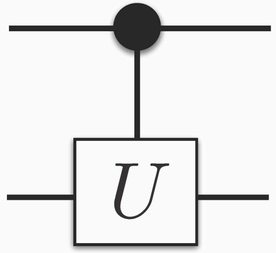


Tensor Networks

Monte Carlo

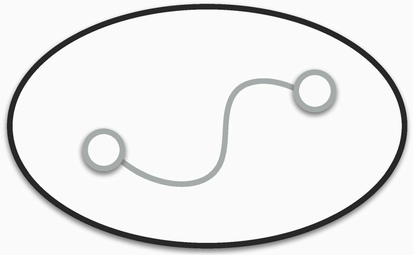


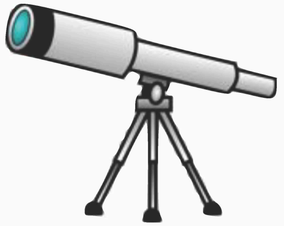




Quantum Circuits

Variational Inference





Holographic RG

Thank You!

Shuo-Hui Li Pan Zhang Yi-Zhuang You Linfeng Zhang Weinan E

IOP, CAS ITP, CAS UCSD Princeton