Winter School on Numerical Methods for Strongly Correlated Quantum Systems, February 19-23 2018, Marburg

Deep Learning and Quantum Many-Body Computation

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

Fall School 2013 on Advanced Algorithms for Correlated Quantum Matter

Fakher Assaad | Announcement, Fall School 2013, Schools/Conferences | April 9, 2013 | No Comments »

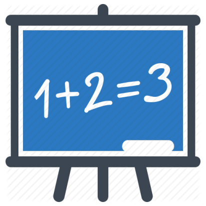
Our first Fall school on

Advanced Algorithms for Correlated Quantum Matter

took place in Würzburg, during the week of September 30 to October 4.

Plan

Introduction



Deep learning theoretical minimum

Applications to quantum many-body physics and beyond

Hands on session (Jin-Guo Liu, Shuo-Hui Li)

Hands on https://github.com/GiggleLiu/marburg

Deep Learning and Quantum Many-Body Physics - Hands on Session

Table of Contents

We have prepaired four examples

- Computation Graphs and Back Propagation
- RealNVP network for sampling
- Restricted Boltzmann Machine for image restoration
- Deep Neural Network as a Quantum Wave Function Ansatz

They have been uploaded to both Google drive and Github repository. Have fun!

Preparations

You may use **either** local or online accesses to our python notebooks.

If you are not using an Nvidia GPU or its driver are not properly configured, online access is recommended, otherwise you may loss some fun in this session.

Lecture note http://wangleiphy.github.io/lectures/DL.pdf

Lecture Note on Deep Learning and Quantum Many-Body Computation

Jin-Guo Liu, Shuo-Hui Li, and Lei Wang*

Institute of Physics, Chinese Academy of Sciences
Beijing 100190, China

February 14, 2018

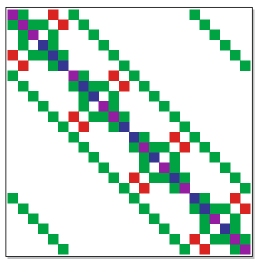
Abstract

This note introduces deep learning from a computational quantum physicist's perspective. The focus is on deep learning's impacts to quantum many-body computation, and vice versa. The latest version of the note is at http://wangleiphy.github.io/. Please send comments, suggestions and corrections to the email address in below.

* wanglei@iphy.ac.cn

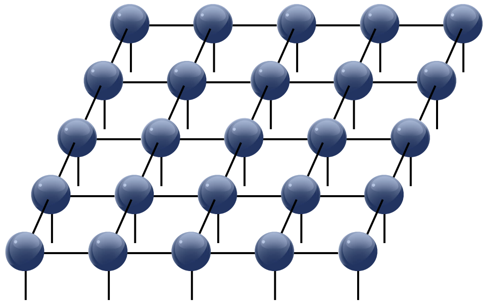
CONTENTS

Quantum Many-Body Computation

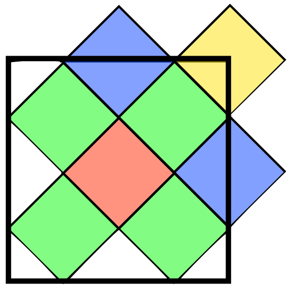


exact diagonalization

quantum Monte Carlo

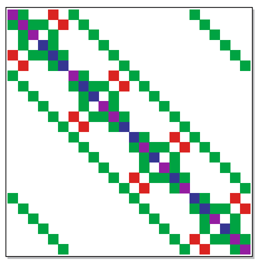


tensor network states



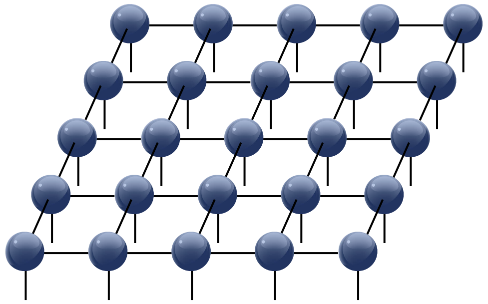
dynamical mean field theories

Quantum Many-Body Computation

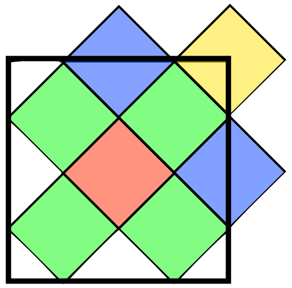


exact diagonalization

quantum Monte Carlo



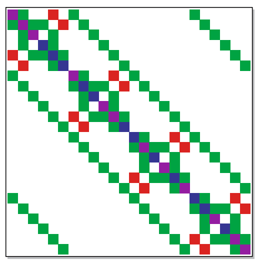
tensor network states



dynamical mean field theories

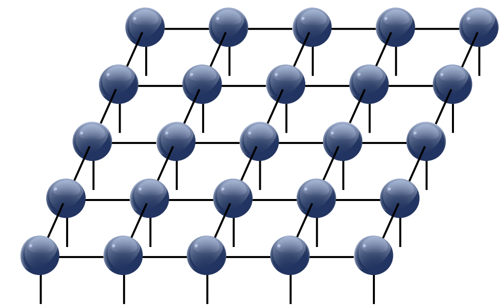
Algorithmic improvement in past 20 years outperformed Moore's law

Quantum Many-Body Computation

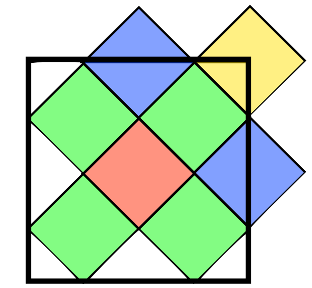


exact diagonalization

quantum Monte Carlo



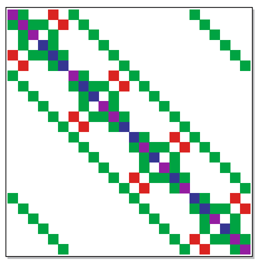
tensor network states



dynamical mean field theories

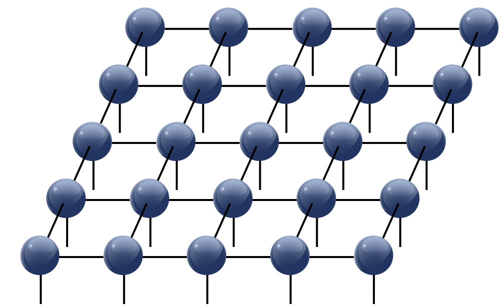
is faster than

Quantum Many-Body Computation

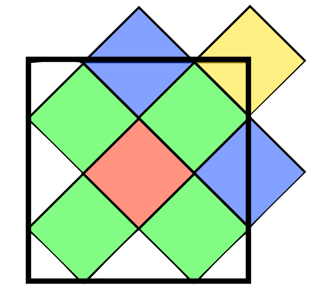


exact diagonalization

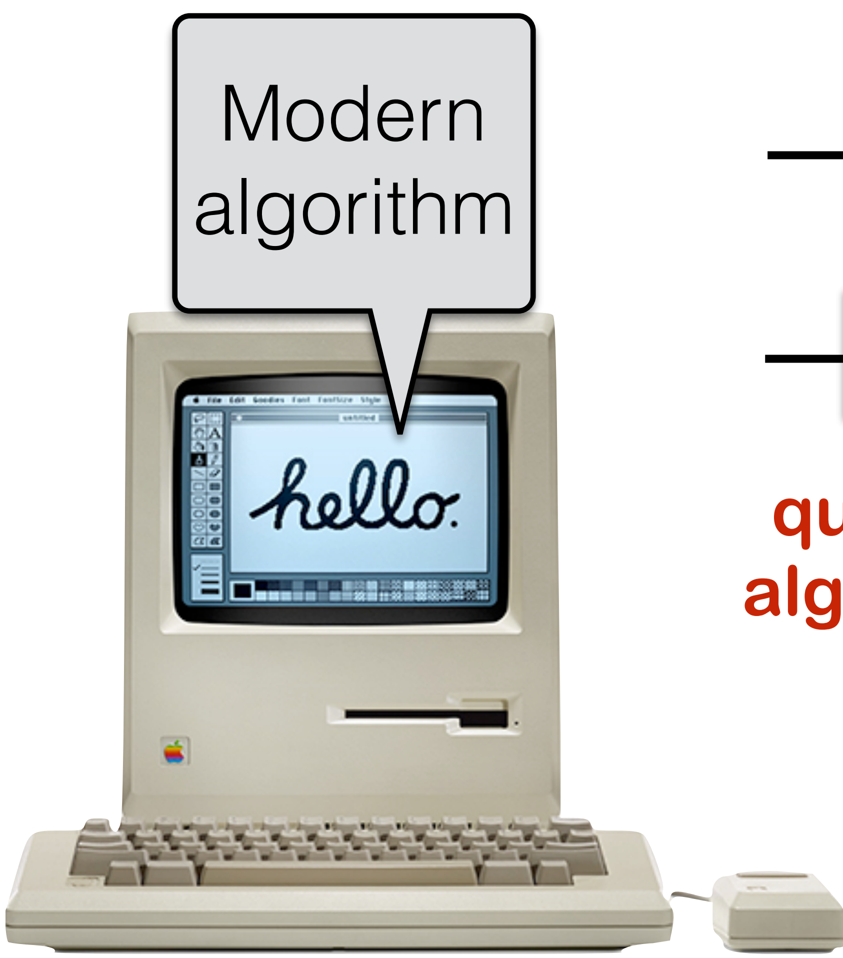
quantum Monte Carlo

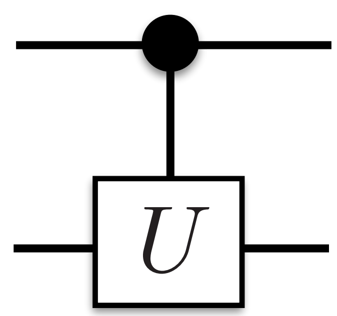


tensor network states

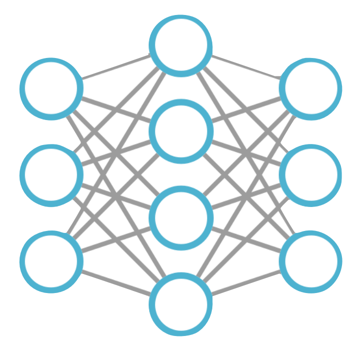


dynamical mean field theories

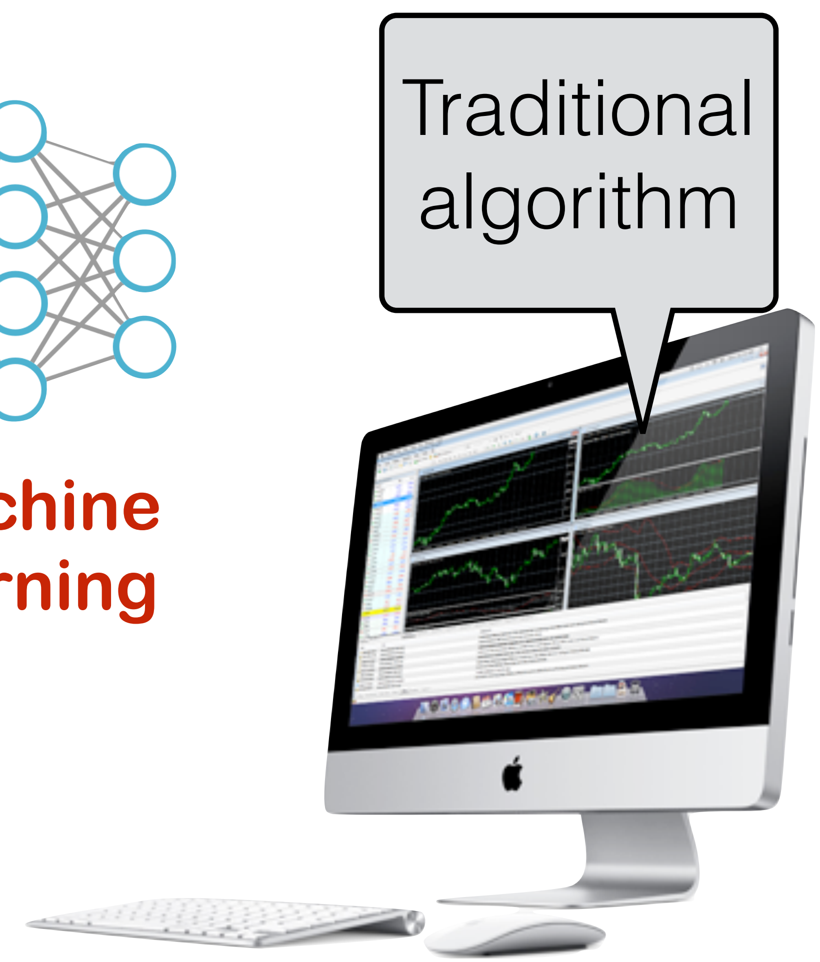




quantum algorithms



machine learning



Materials Discovery/Ab Initio/DFT

Microscopic Composition

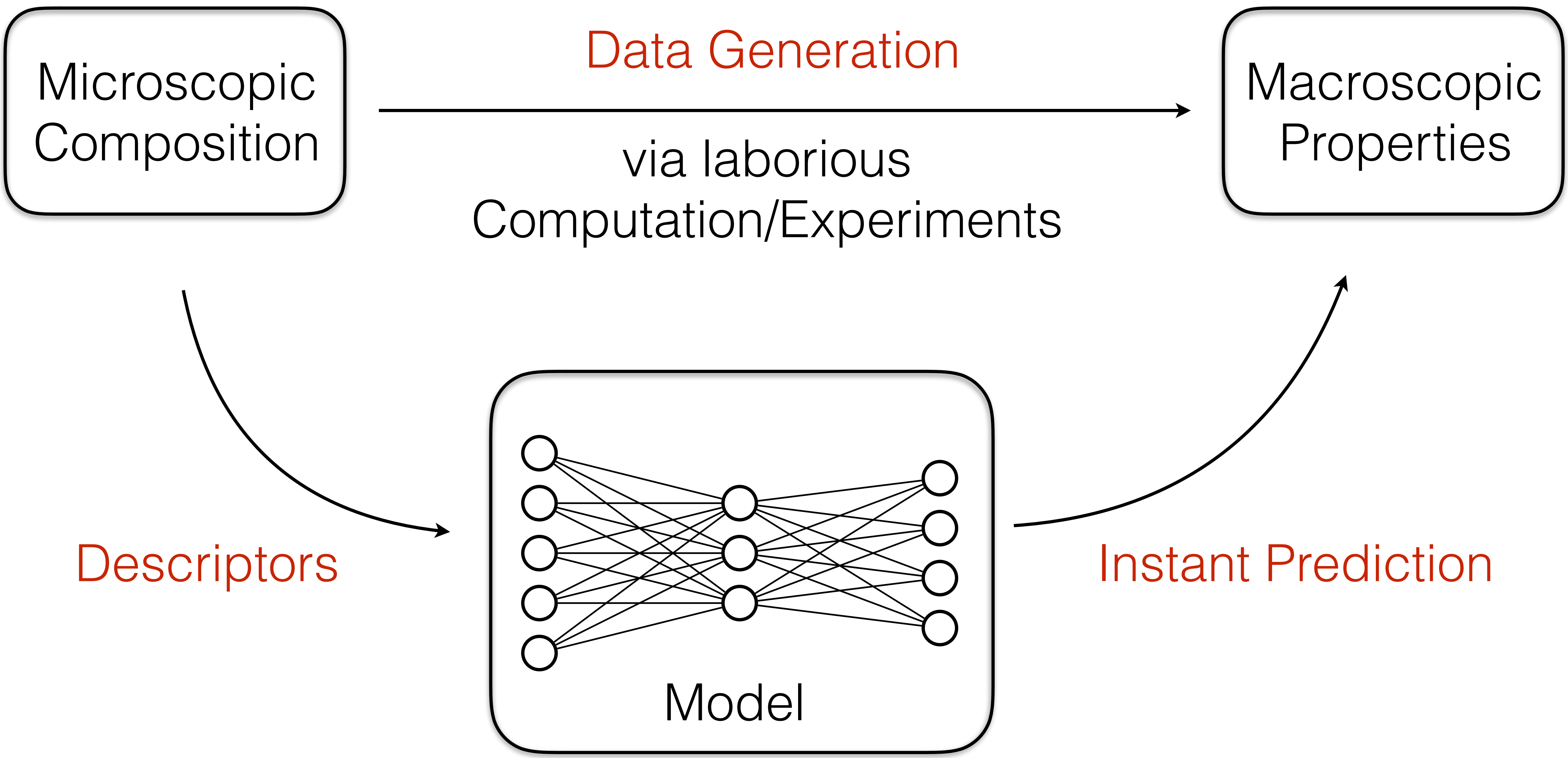
Data Generation

via laborious Computation/Experiments

Macroscopic Properties

“Machine Learning in Materials Science: Recent Progress and Critical Next Steps”
Rampi Ramprasad @ IPAM program on Understanding Many-Particle Systems with ML, 2016

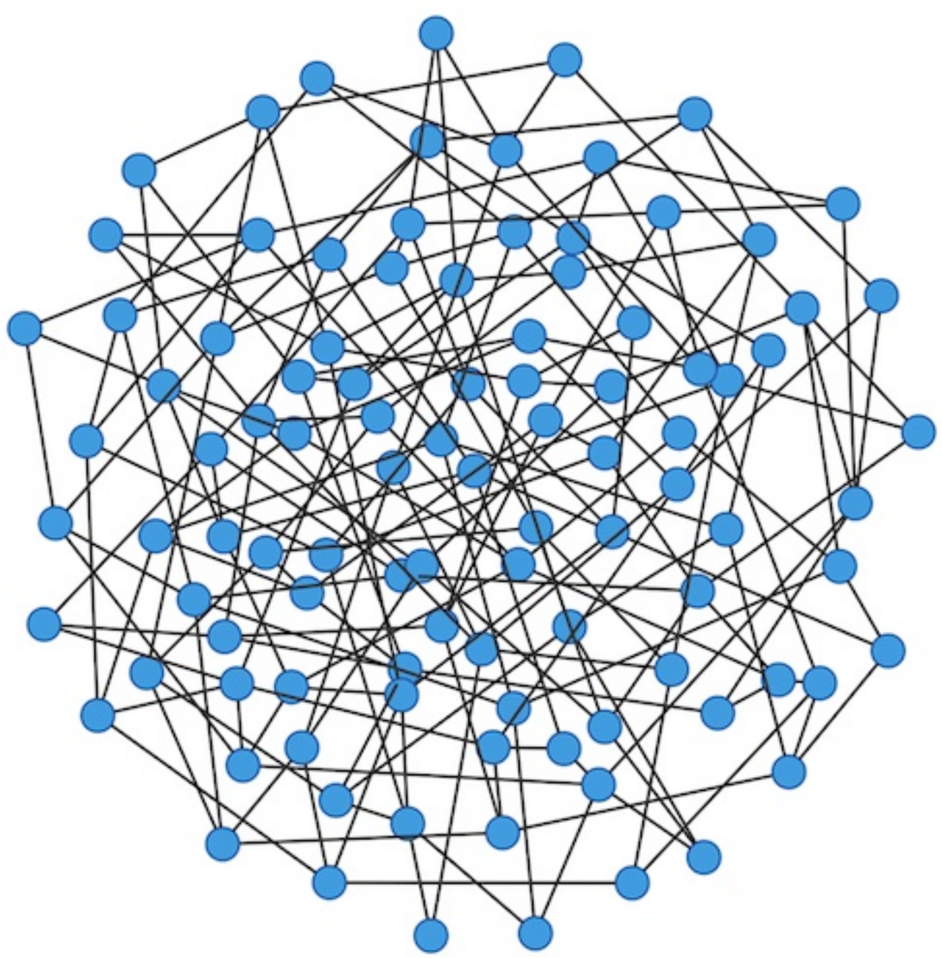
Materials Discovery/Ab Initio/DFT



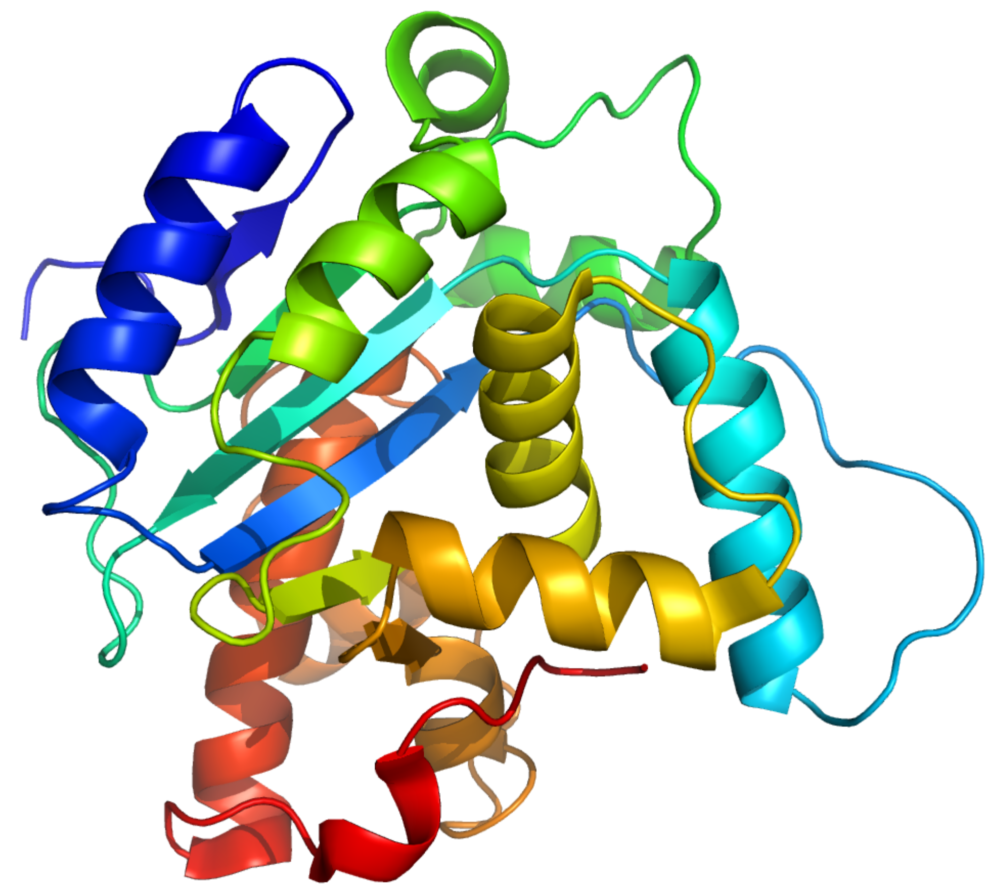
"Machine Learning in Materials Science: Recent Progress and Critical Next Steps"
Rampi Ramprasad @ IPAM program on Understanding Many-Particle Systems with ML, 2016

Statistical Physics

Spin Glasses/Complex Networks

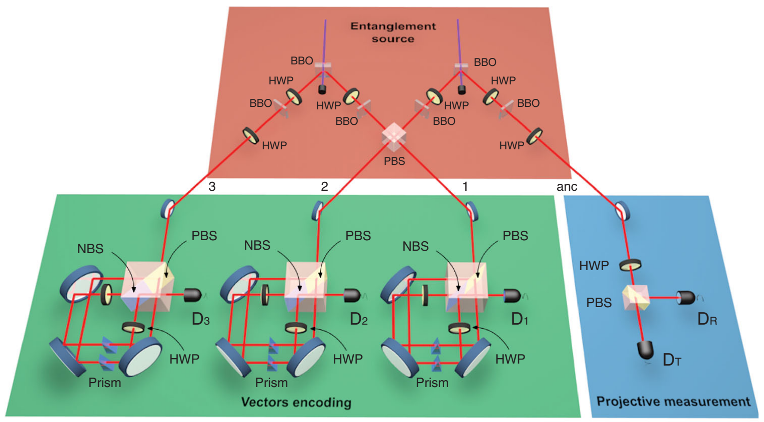


Soft Matter/Biophysics



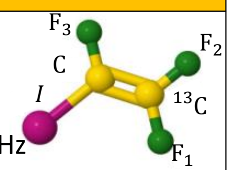
Deep learning has its statistical physics gene

Quantum Information & Computation

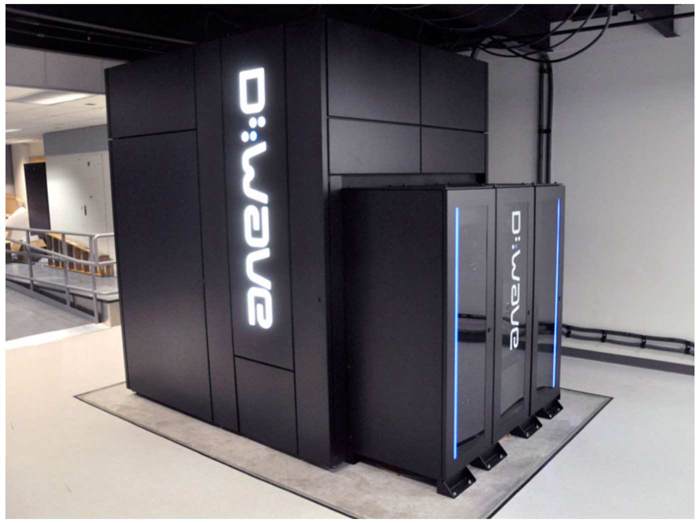


Cai et al, PRL 114, 110504 (2015)

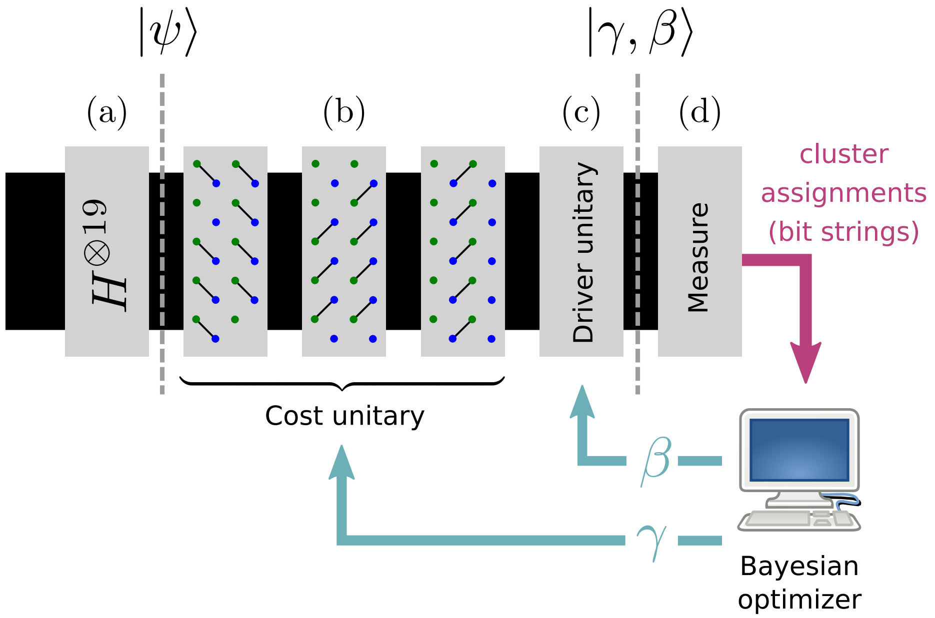
	^{13}C	F_1	F_2	F_3
^{13}C	15479.9Hz			
F_1	-297.7Hz	-33130.1Hz		
F_2	-275.7Hz	64.6Hz	-42681.4Hz	
F_3	39.1Hz	51.5Hz	-129.0Hz	-56443.5Hz
T_2^*	1.22s	0.66s	0.63s	0.61s
T_2	7.9s	4.4s	6.8s	4.8s



Li et al, PRL 114, 140504 (2015)

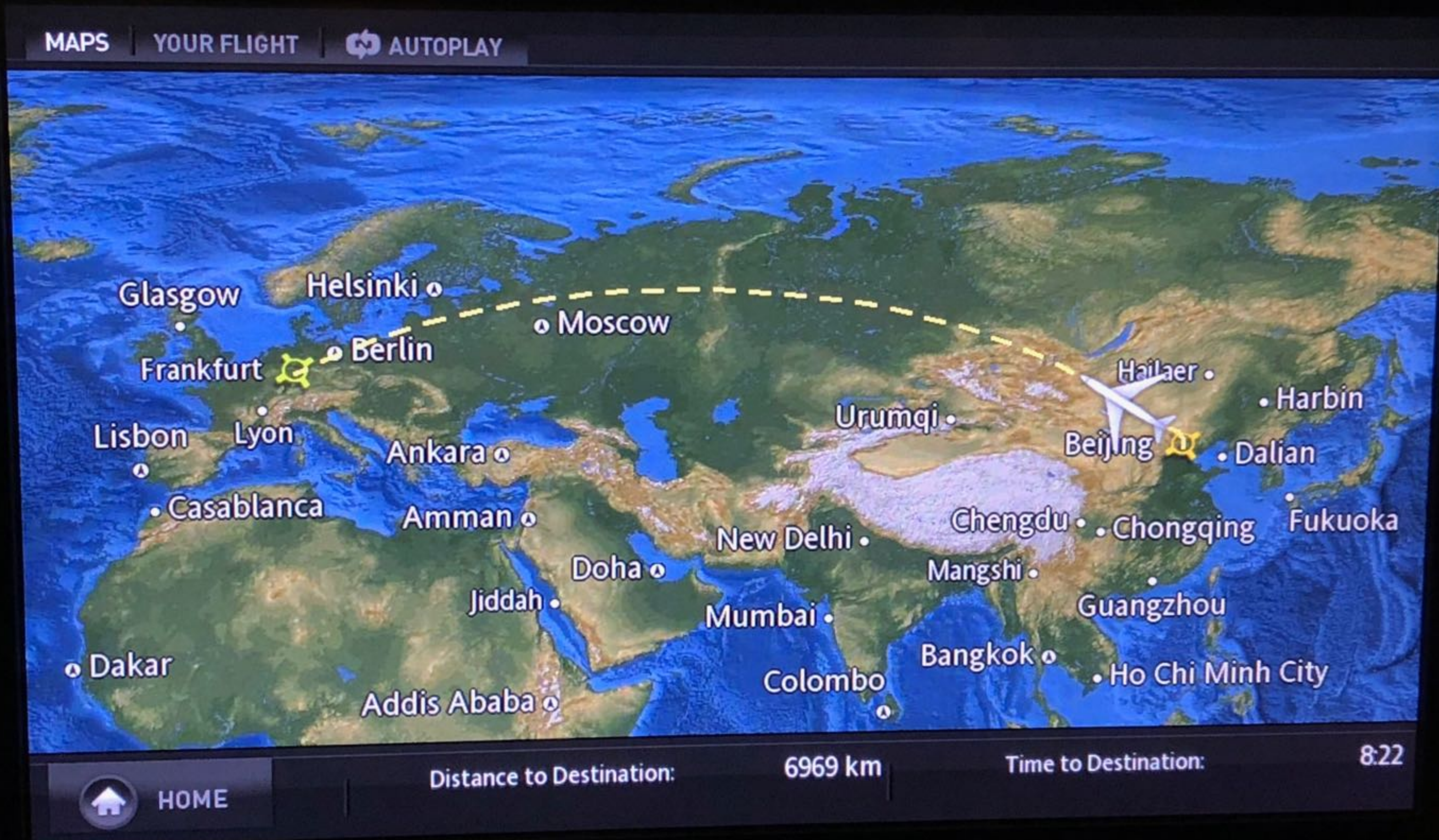


Perdomo-Ortiz et al, 1708.09757

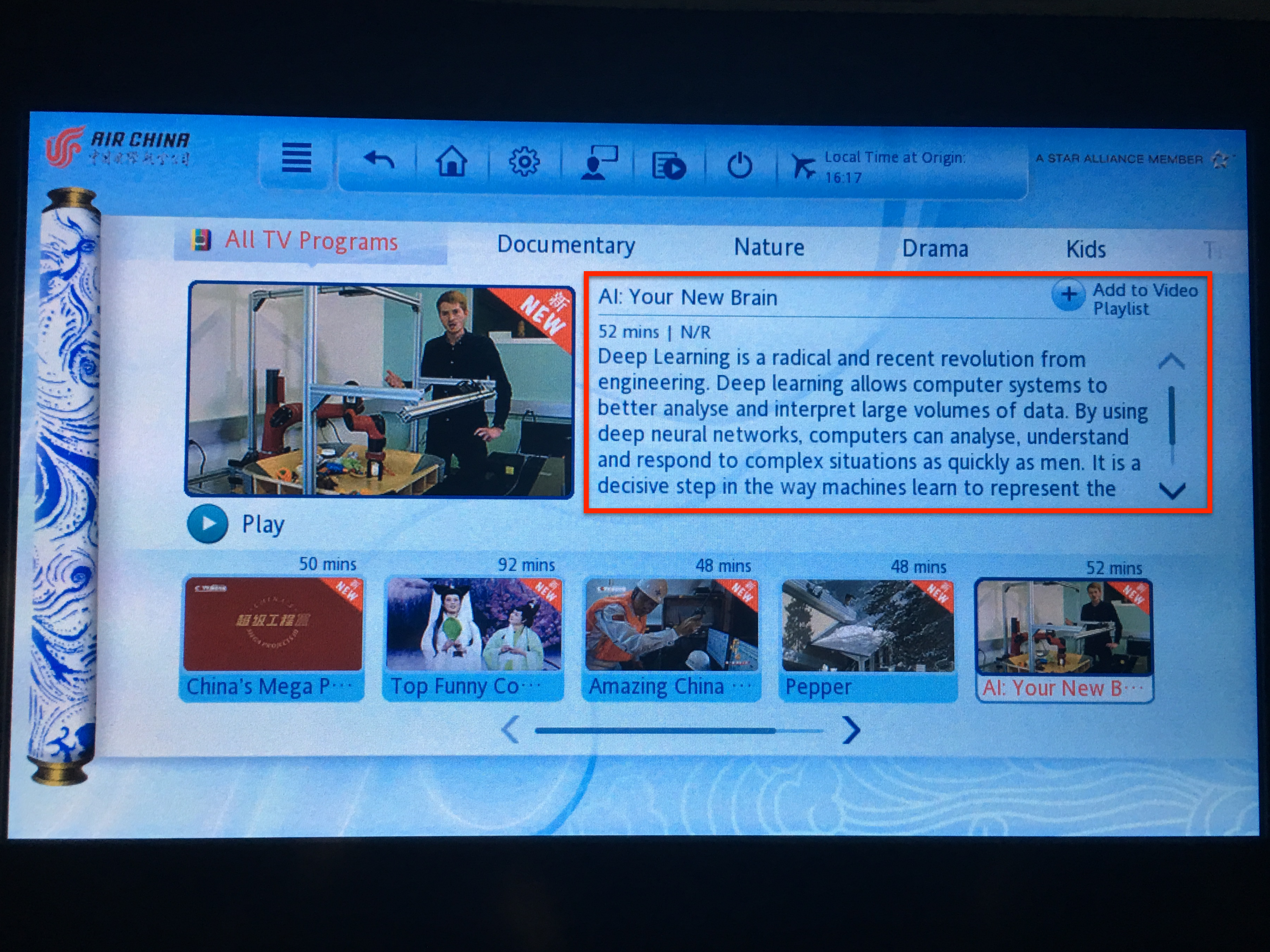


Rigetti Computing, 1712.05771

Review “Quantum machine learning”, Biamonte,Wittek et al, Nature 2017


MAPS
YOUR FLIGHT
AUTOPLAY
Glasgow
Helsinki
Moscow
Berlin
Frankfurt
Lisbon
Lyon
Ankara
Casablanca
Amman
Doha
Jiddah
Dakar
Addis Ababa
Urumqi
Hailaer
Harbin
Beijing
Dalian
Chengdu
Chongqing
Fukuoka
New Delhi
Mangshi
Guangzhou
Mumbai
Bangkok
Colombo
Ho Chi Minh City
HOME
Distance to Destination:
6969 km
Time to Destination:
8:22


AIR CHINA
Local Time at Origin:
16:17
A STAR ALLIANCE MEMBER
All TV Programs
Documentary
Nature
Drama
Kids
NEW
AI: Your New Brain
Add to Video Playlist
52 mins | N/R
Deep Learning is a radical and recent revolution from engineering. Deep learning allows computer systems to better analyse and interpret large volumes of data. By using deep neural networks, computers can analyse, understand and respond to complex situations as quickly as men. It is a decisive step in the way machines learn to represent the
Play
50 mins
China's Mega P···
92 mins
Top Funny Co···
48 mins
Amazing China ···
48 mins
Pepper
52 mins
AI: Your New B···


AIR CHINA
Local Time at Origin:
16:17
A STAR ALLIANCE MEMBER
All TV Programs
Documentary
Nature
Drama
Kids
NEW
AI: Your New Brain
Add to Video Playlist
52 mins | N/R
Deep Learning is a radical and recent revolution from engineering. Deep learning allows computer systems to better analyse and interpret large volumes of data. By using deep neural networks, computers can analyse, understand and respond to complex situations as quickly as men. It is a decisive step in the way machines learn to represent the
Play
50 mins
China's Mega P···
92 mins
Top Funny Co···
48 mins
Amazing China ···
48 mins
Pepper
52 mins
AI: Your New B···

AIR CHINA

Local Time at Origin:

A STAR ALLIANCE MEMBER

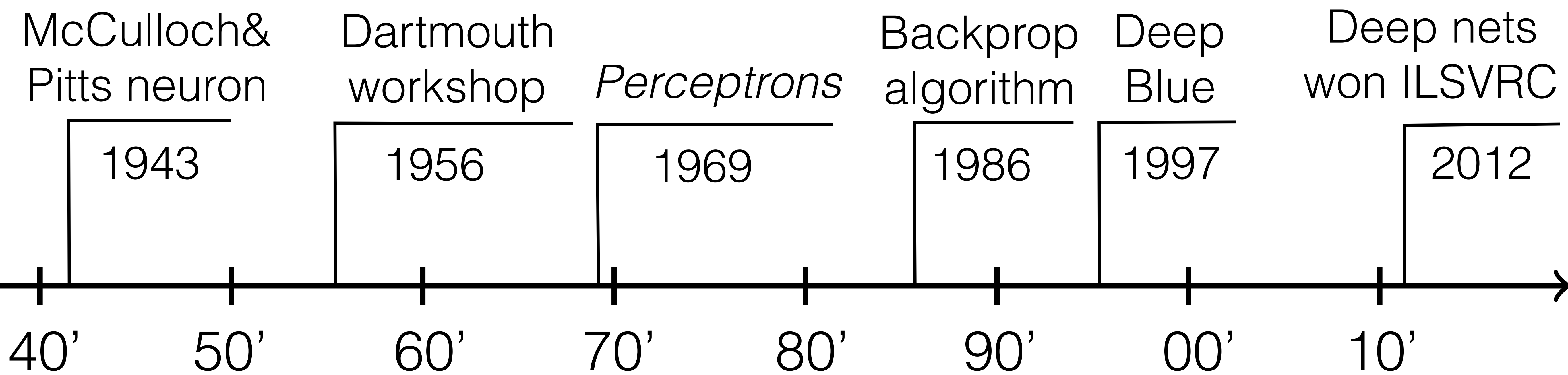
AI: Your New Brain

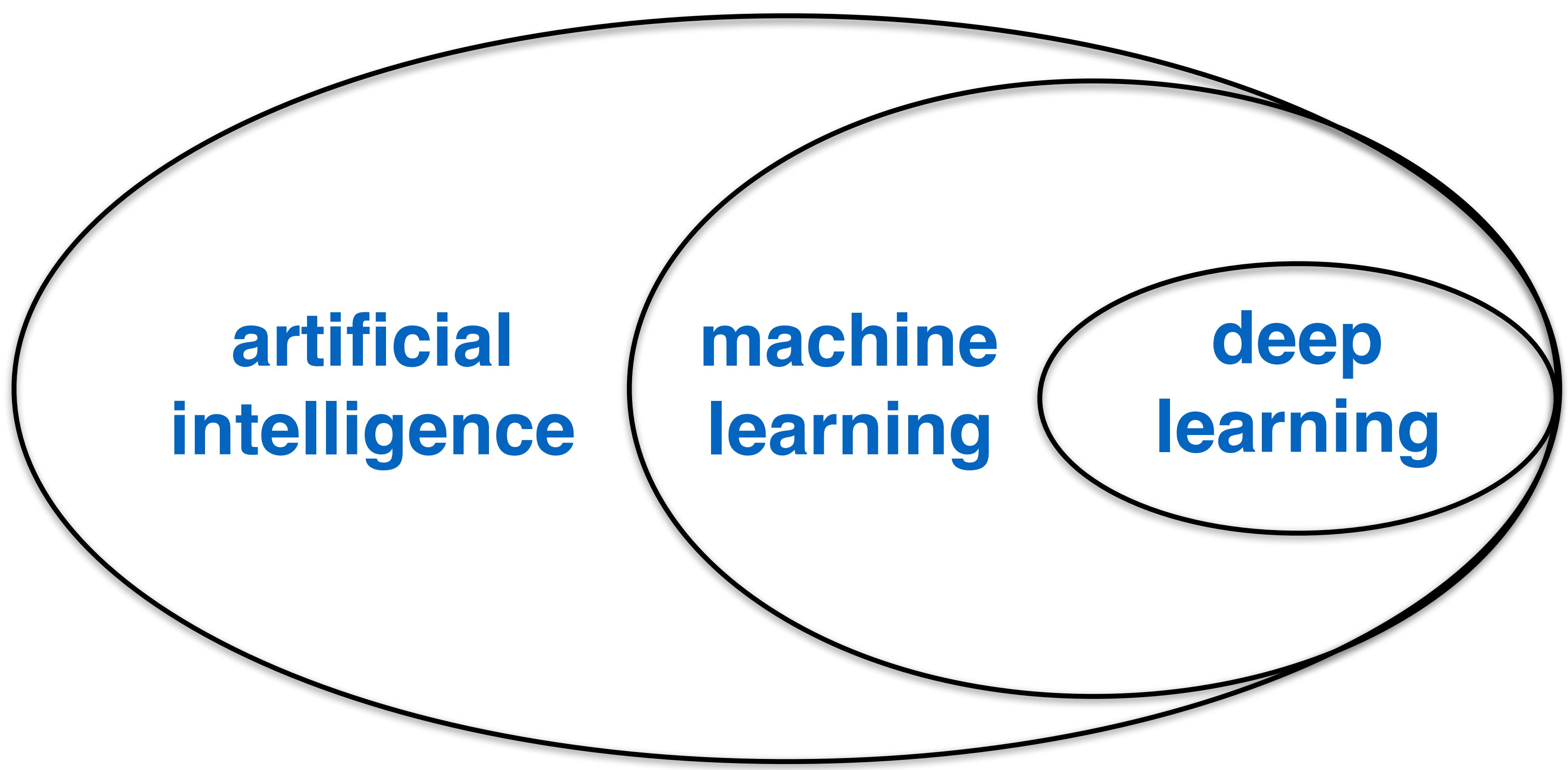
Add to Playlist

52 mins | N/R

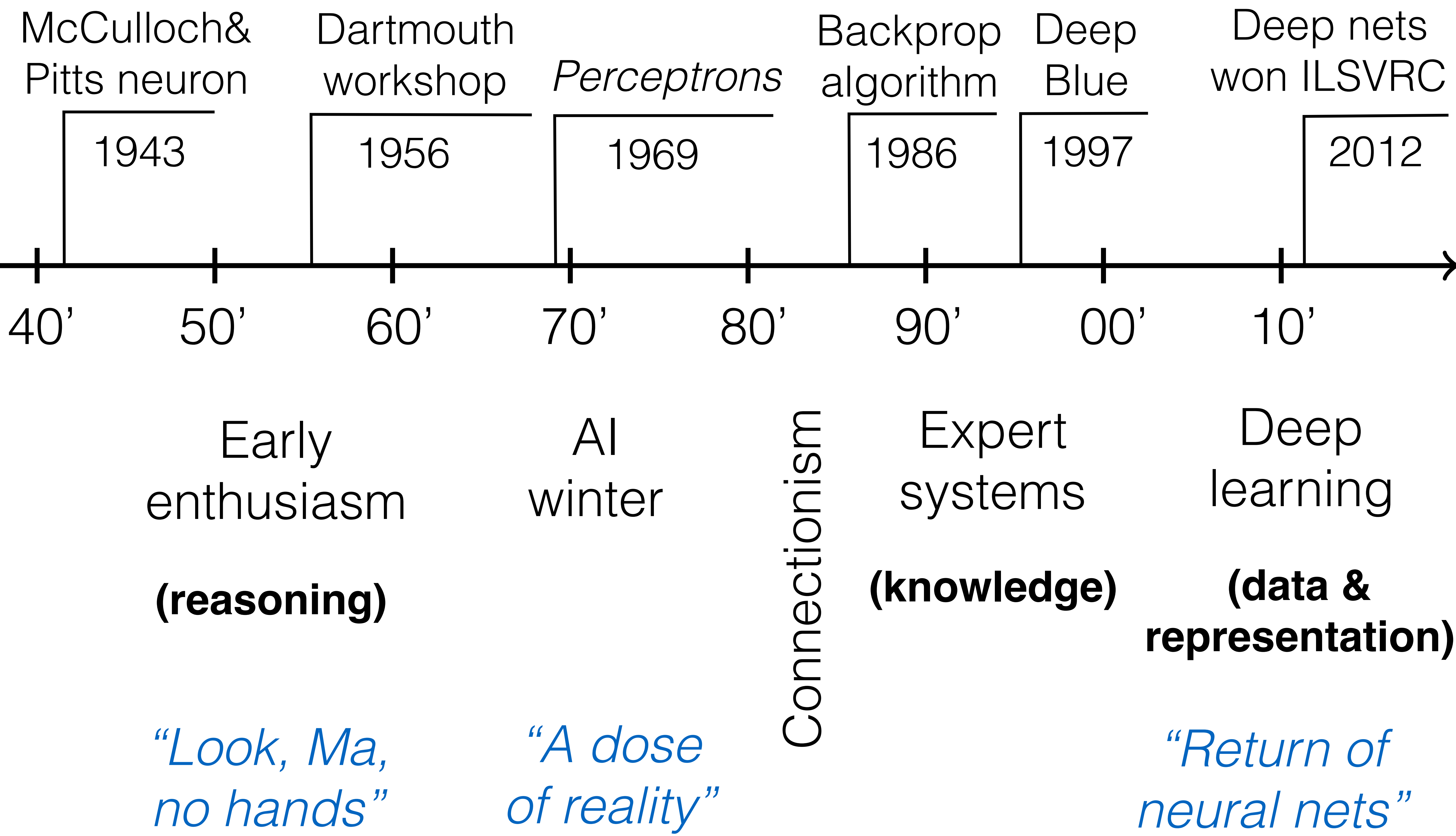
Deep Learning is a radical and recent revolution from engineering. Deep learning allows computer systems to better analyse and interpret large volumes of data. By using deep neural networks, computers can analyse, understand and respond to complex situations as quickly as men. It is a decisive step in the way machines learn to represent the world as humans do. Deep learning brings back to the forefront the search for artificial intelligence.

Timeline of AI research

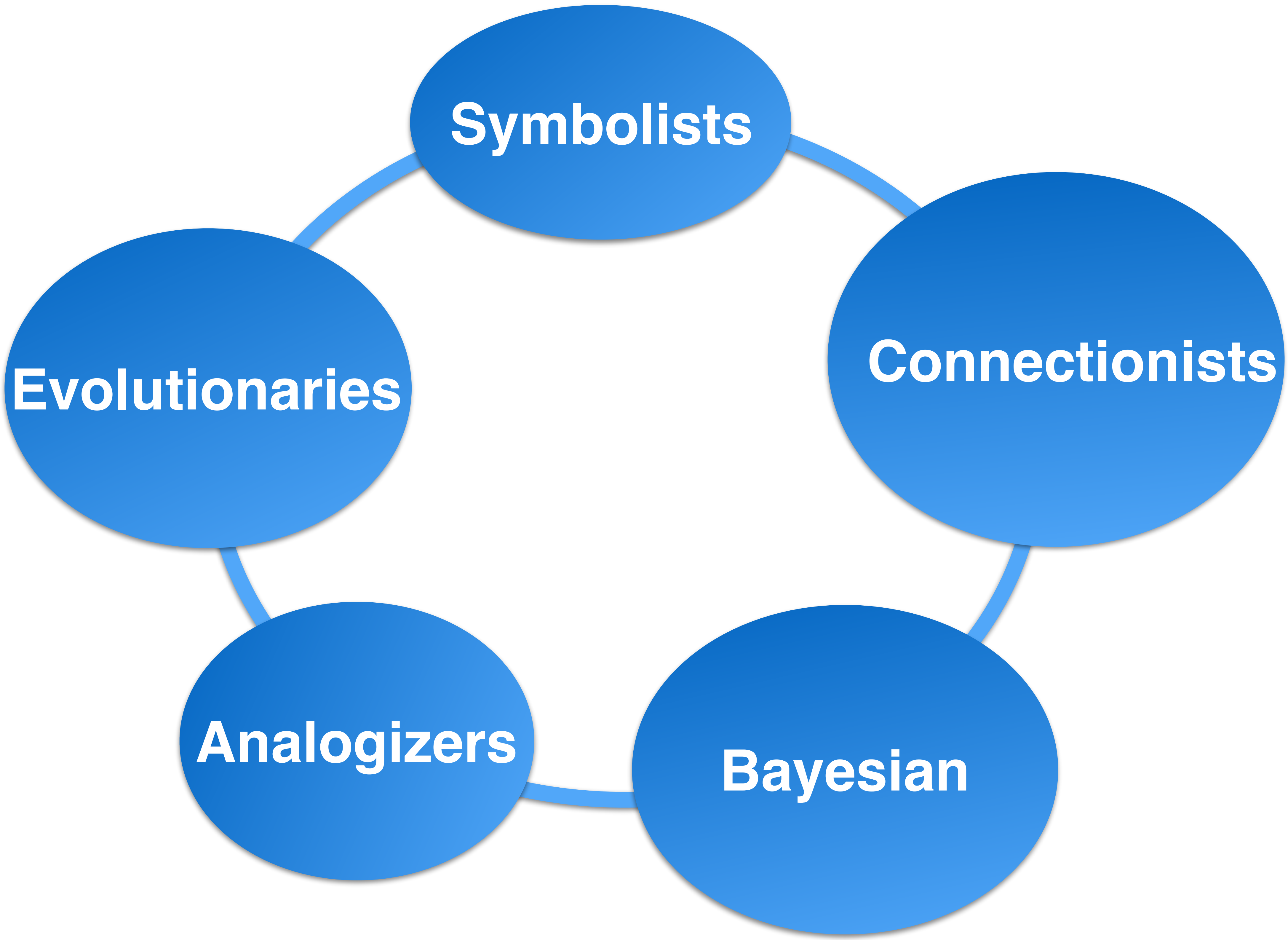




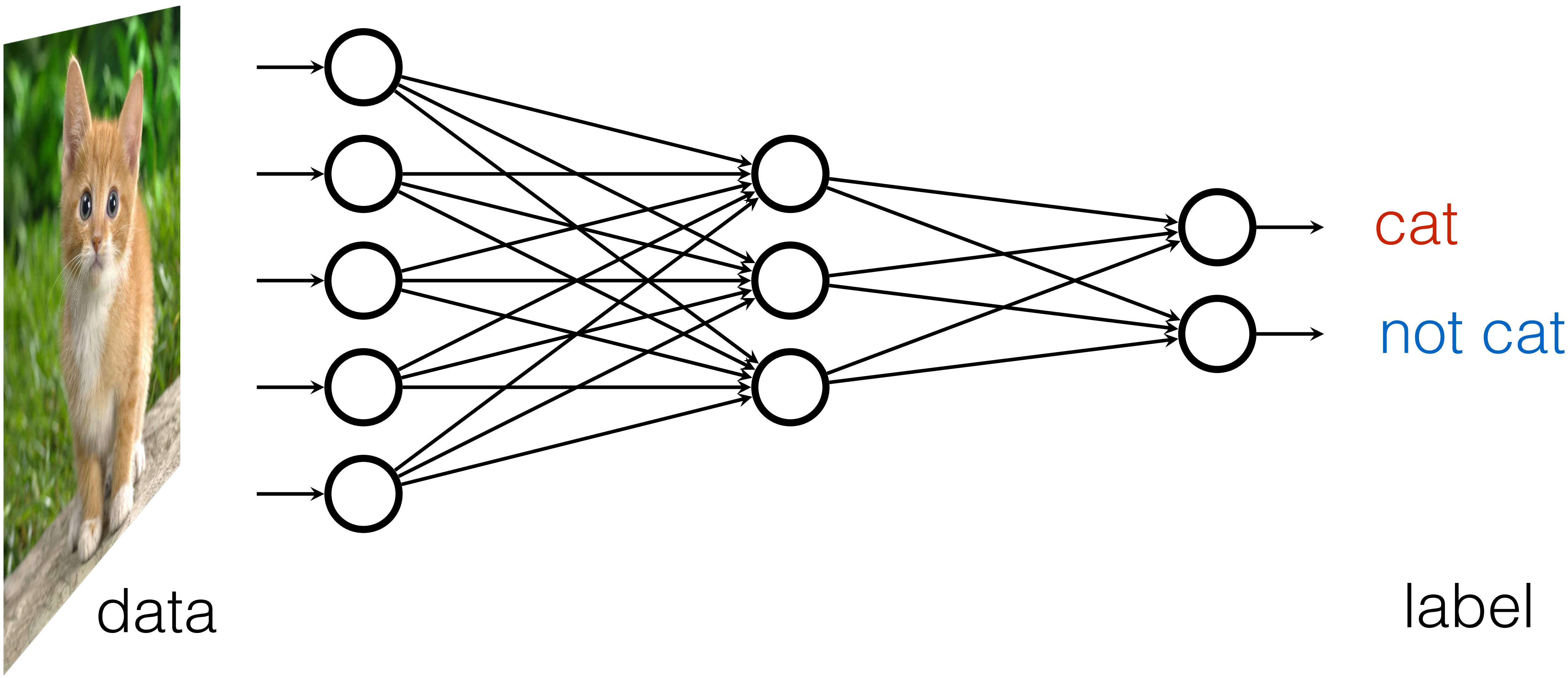
Timeline of AI research



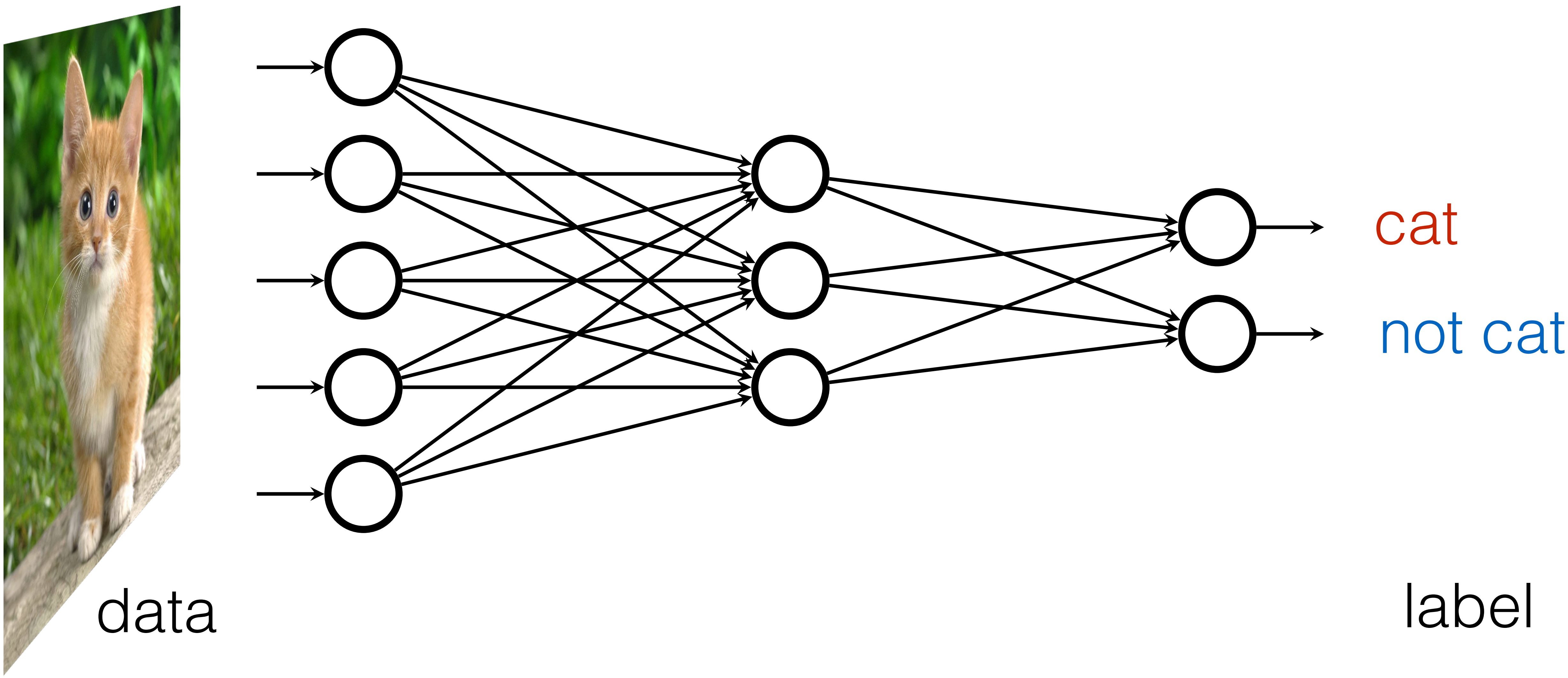
Five schools of ML



Artificial Neural Networks



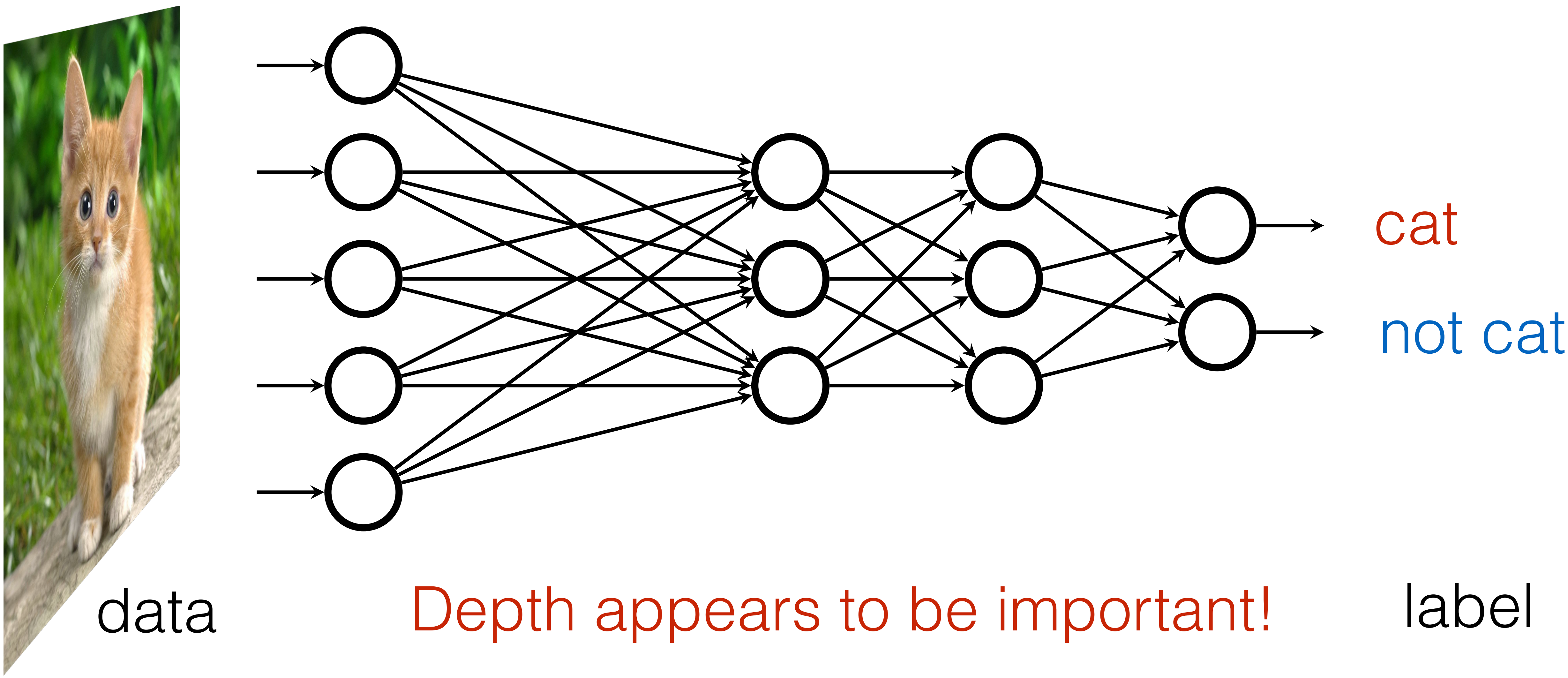
Artificial Neural Networks



Universal Function Approximator

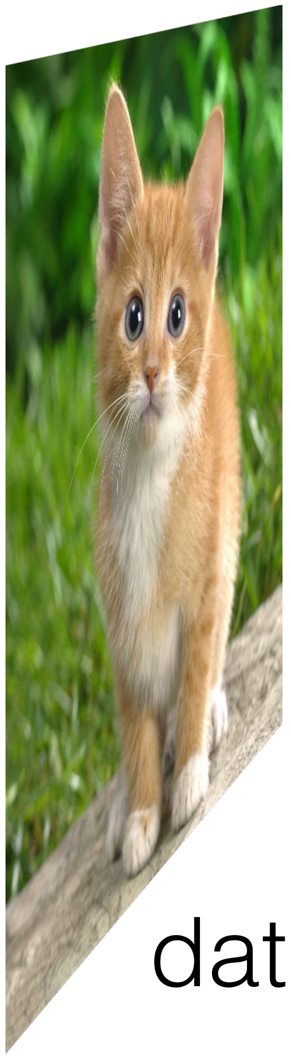
Cybenko 1989
Hornik, Stinchcombe, White 1989

Artificial Neural Networks

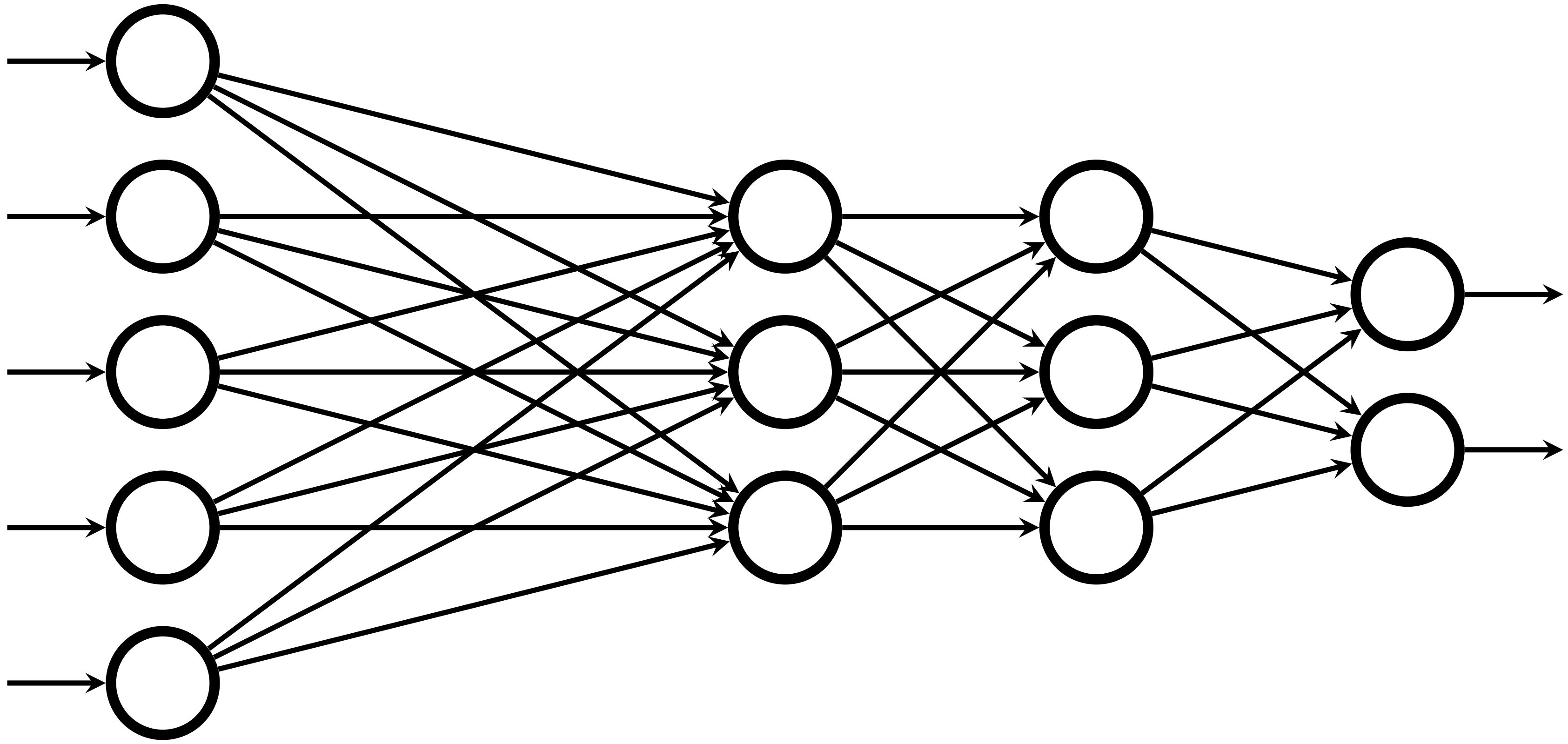


Depth appears to be important!

Artificial Neural Networks



data



cat

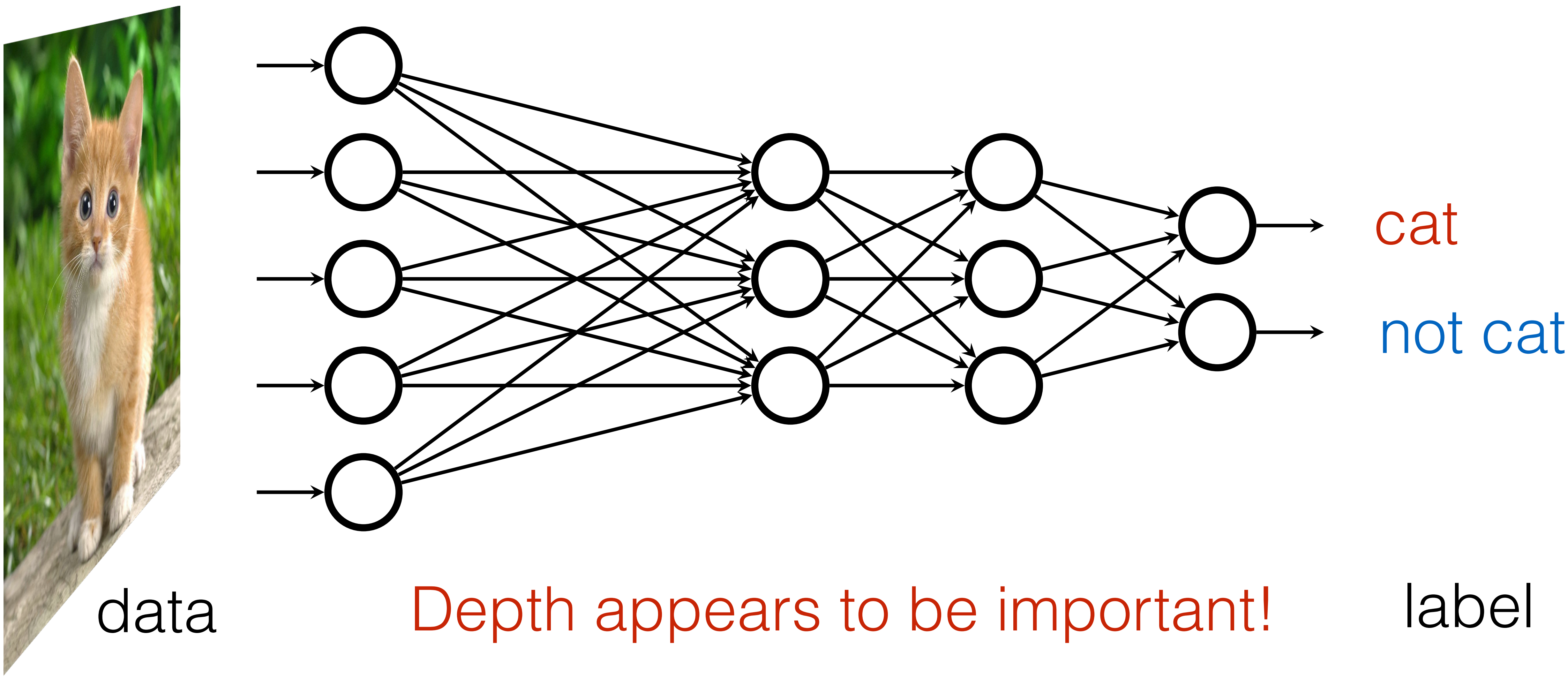
not cat

label

Depth appears to be important!

Q: Why does deep learning work?

Artificial Neural Networks

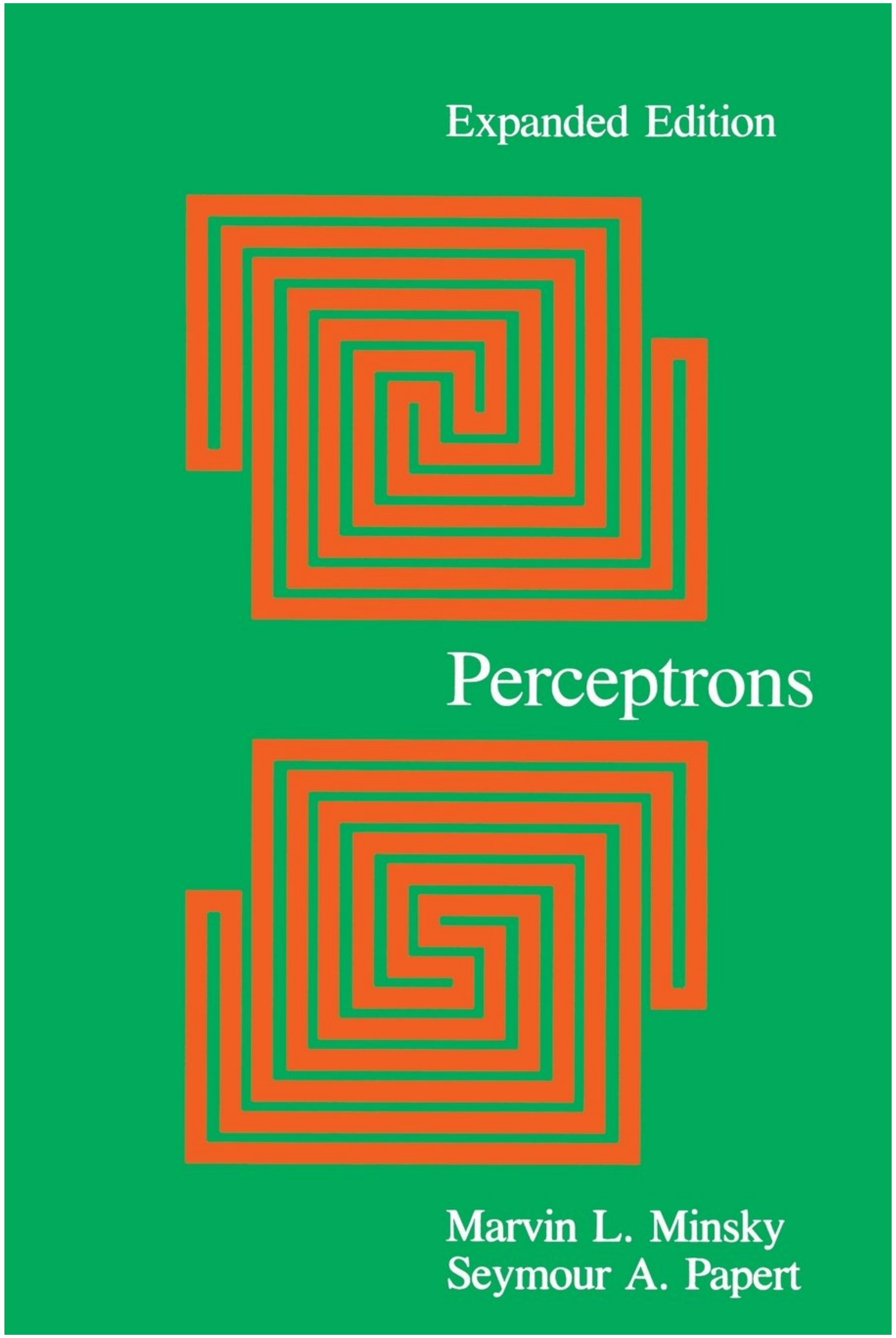


Depth appears to be important!

Q: Why does deep learning work?

A: Law of physics: symmetry, locality, compositionality, renormalization group, and quantum entanglement.

Lin, Tegmark, Rolnick ,1608.08225 Mehta, Schwab, 1410.3831 Levine et al, 1704.01552 …



1st edition 1969
expanded 1972
commented 1988

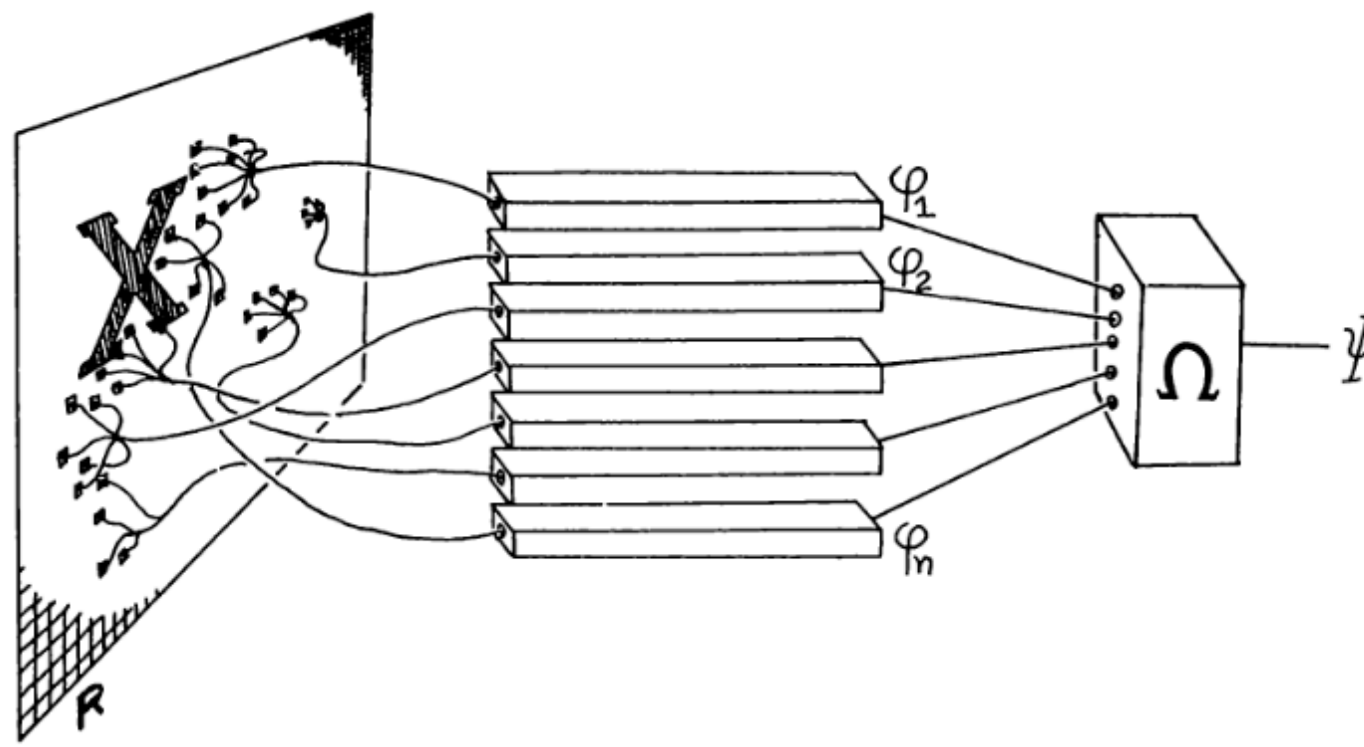
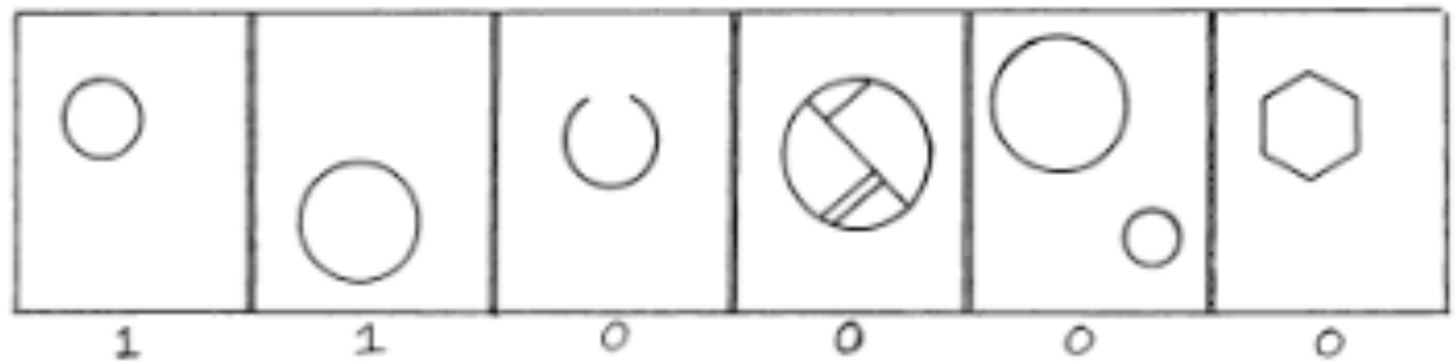


Figure 0.1

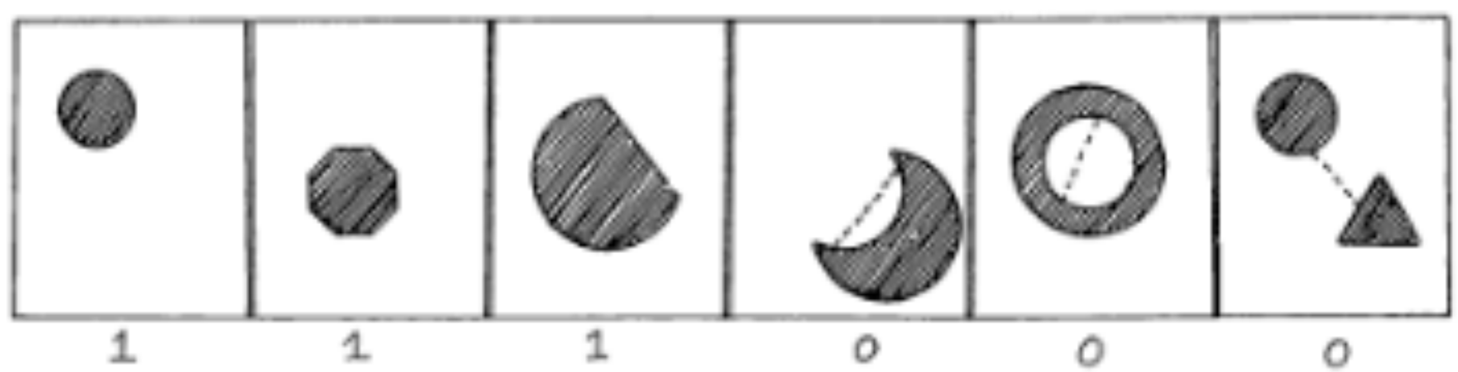
0.3 Cybernetics and Romanticism

Our discussion will include some rather sharp criticisms of earlier work in this area. Perceptrons have been widely publicized as "pattern recognition" or "learning" machines and as such have been discussed in a large number of books, journal articles, and voluminous "reports." Most of this writing (some exceptions are mentioned in our bibliography) is without scientific value and we will not usually refer by name to the works we criticize. The sciences of computation and cybernetics began, and it seems quite rightly so, with a certain flourish of romanticism. They were laden with attractive and exciting new ideas which have already borne rich fruit. Heavy demands of rigor and caution could have held this development to a much slower pace; only the future could tell which directions were to be the best. We feel, in fact, that the solemn experts who most complained about the "exaggerated claims" of the cybernetic enthusiasts were, in the balance, much more in the wrong. But now the time has come for maturity, and this requires us to match our speculative enterprise with equally imaginative standards of criticism.

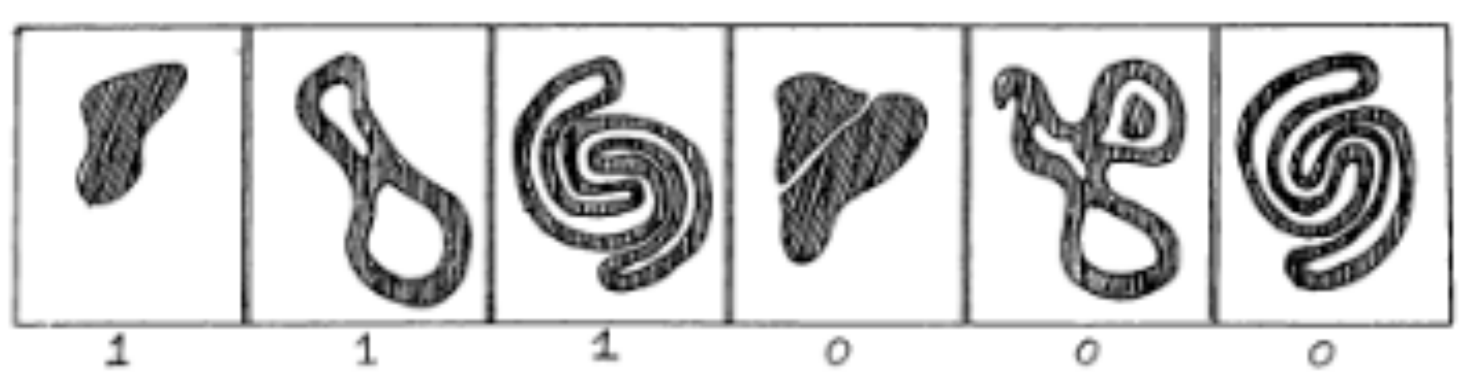
$$\psi_{\text{CIRCLE}}(X) = \begin{cases} 1 \text{ if the figure } X \text{ is a circle,} \\ 0 \text{ if the figure is not a circle;} \end{cases}$$



$$\psi_{\text{CONVEX}}(X) = \begin{cases} 1 \text{ if } X \text{ is a convex figure,} \\ 0 \text{ if } X \text{ is not a convex figure;} \end{cases}$$

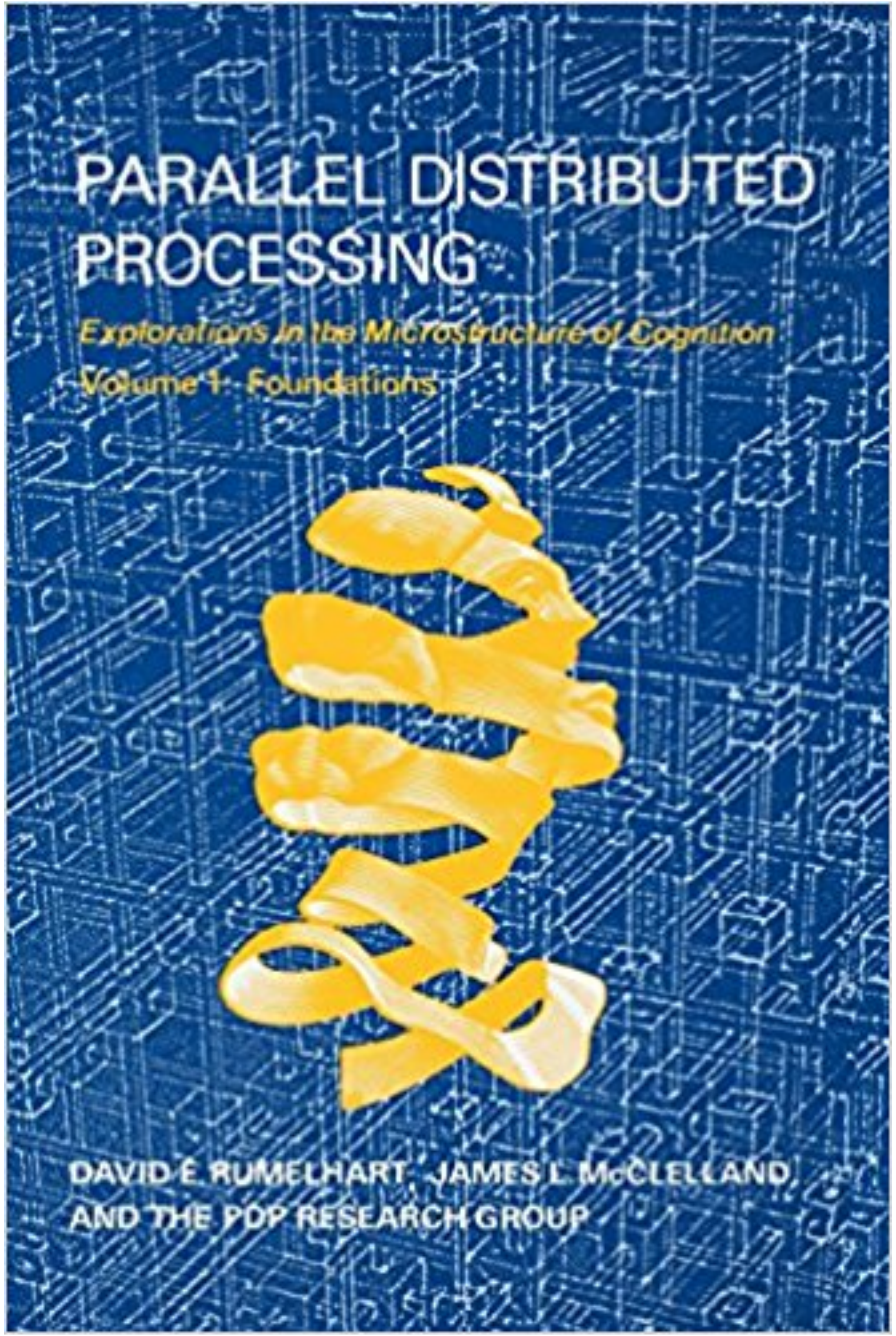


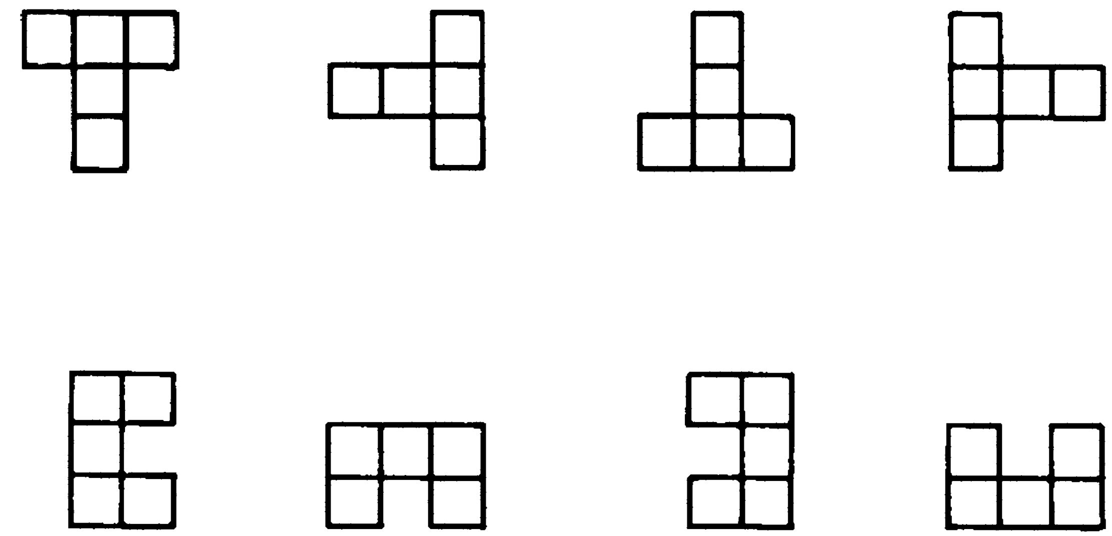
$$\psi_{\text{CONNECTED}}(X) = \begin{cases} 1 \text{ if } X \text{ is a connected figure,} \\ 0 \text{ otherwise.} \end{cases}$$



“Manifesto of Connectionism”

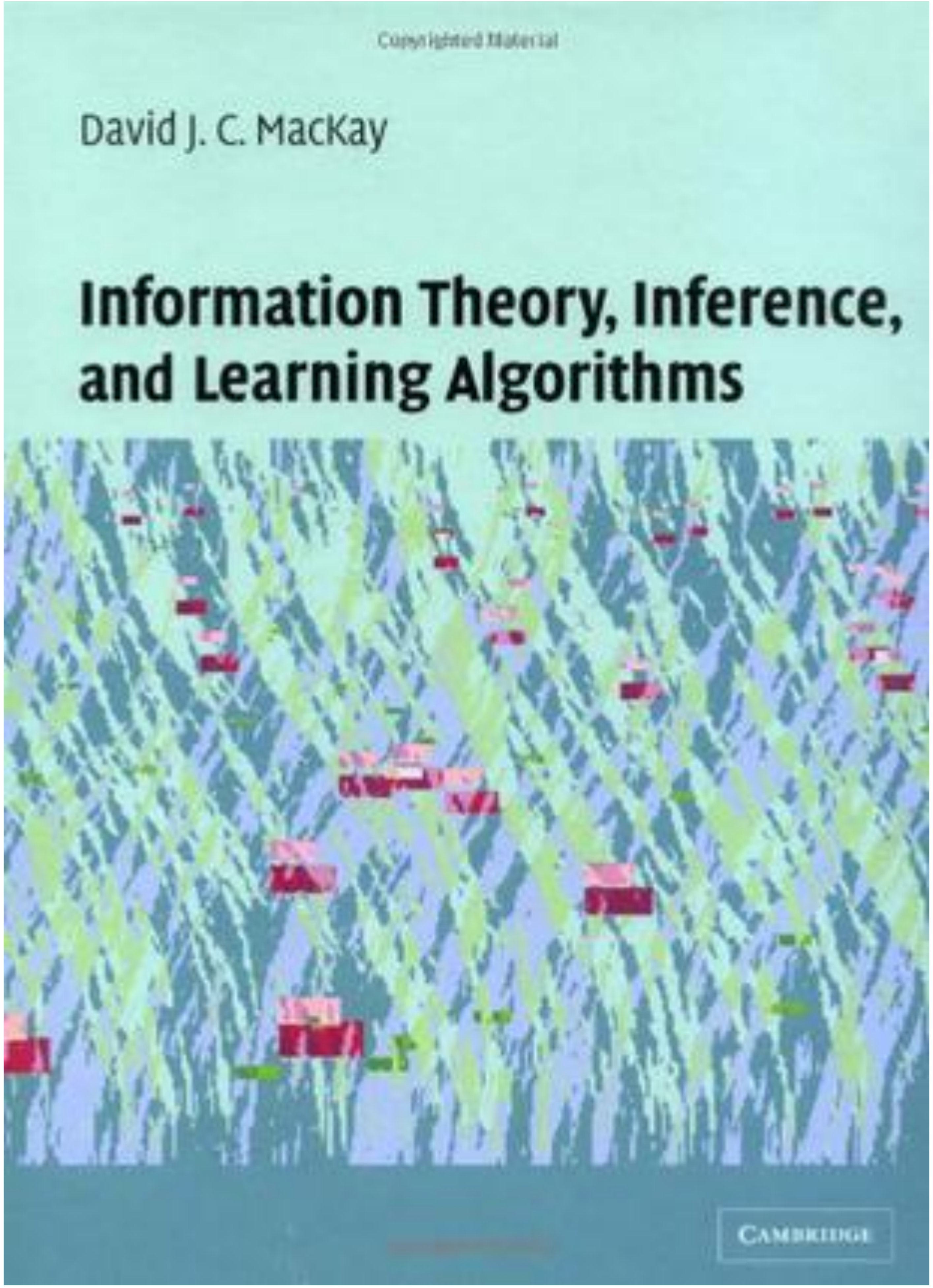
1988





“T-C Problem”

Contains many interesting experiments & theories on toy problems and foundational thoughts by the pioneers



Insightful, fun to read

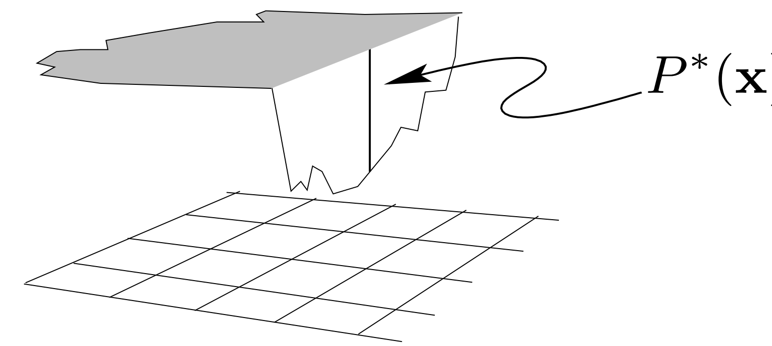


Figure 29.2. A lake whose depth at $\mathbf{x} = (x, y)$ is $P^*(\mathbf{x})$.

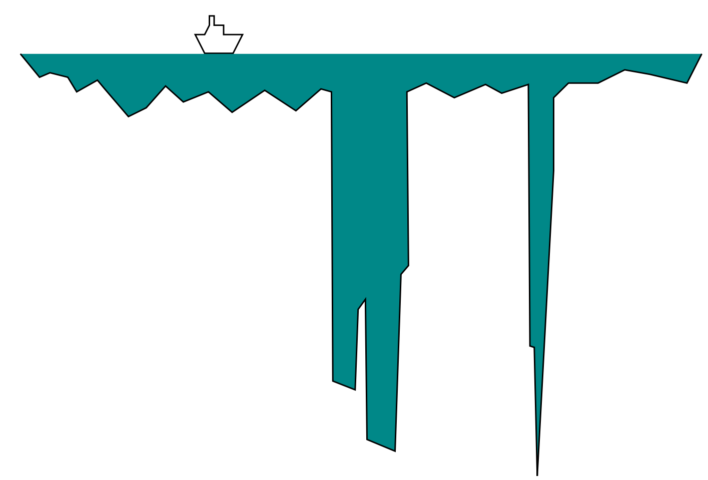
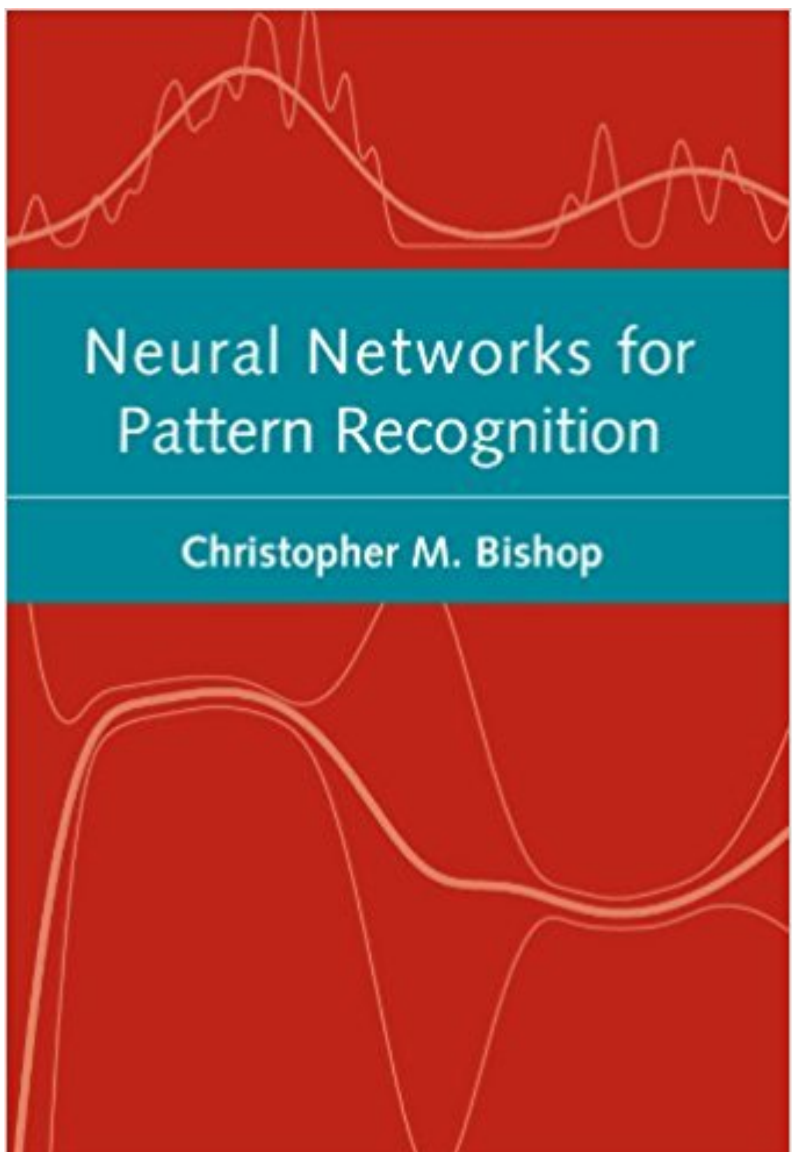


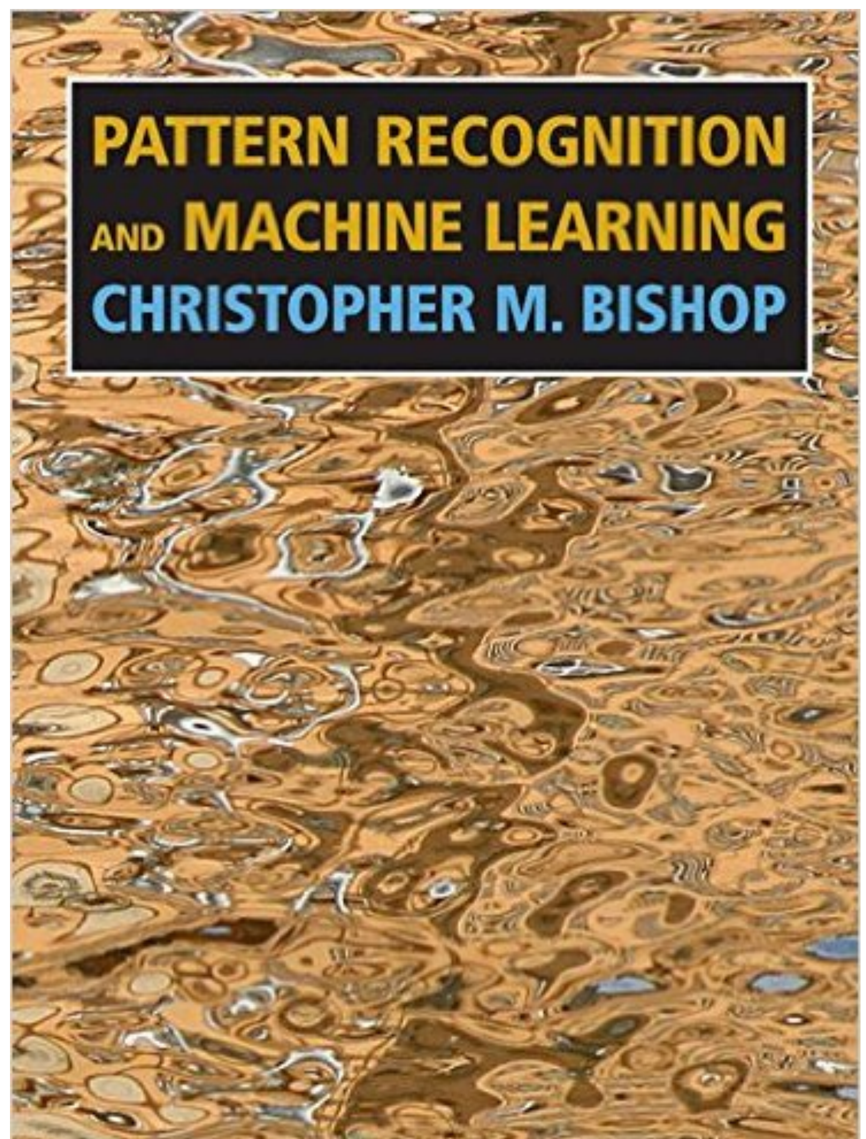
Figure 29.3. A slice through a lake that includes some canyons.

Modern textbooks

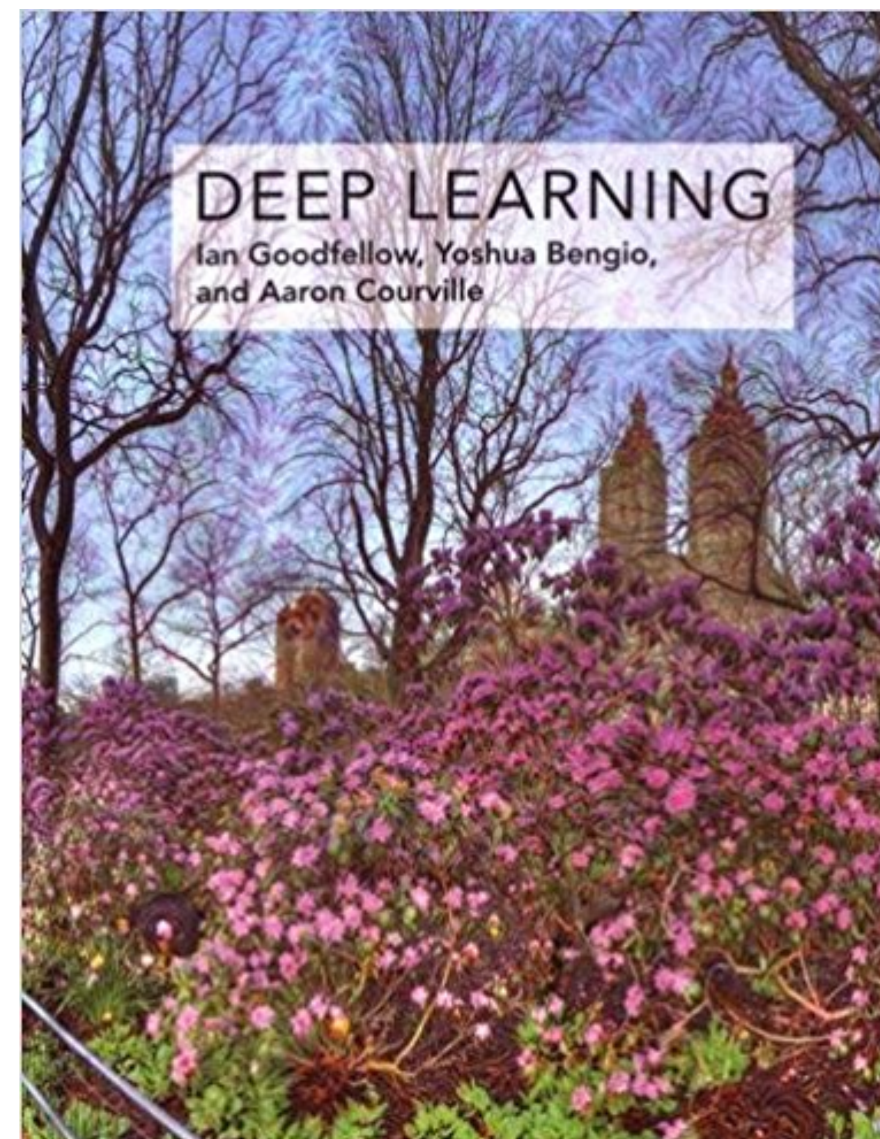
1996



2006



2016



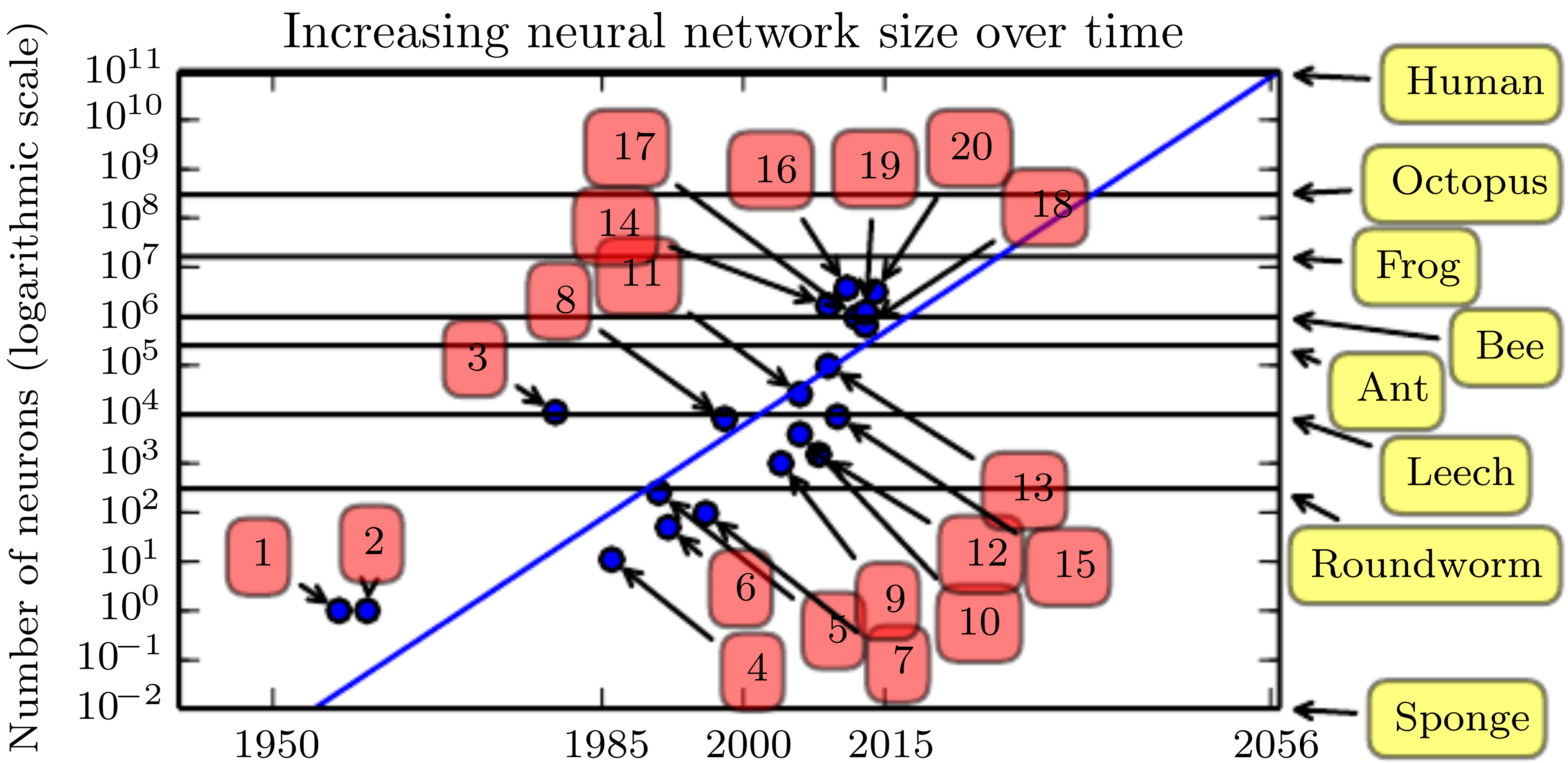


Figure 1.11: Since the introduction of hidden units, artificial neural networks have doubled in size roughly every 2.4 years. Biological neural network sizes from Wikipedia (2015).

1. Perceptron (Rosenblatt, 1958, 1962)
2. Adaptive linear element (Widrow and Hoff, 1960)
3. Neocognitron (Fukushima, 1980)
4. Early back-propagation network (Rumelhart *et al.*, 1986b)
5. Recurrent neural network for speech recognition (Robinson and Fallsi
6. Multilayer perceptron for speech recognition (Bengio *et al.*, 1991)
7. Mean field sigmoid belief network (Saul *et al.*, 1996)
8. LeNet-5 (LeCun *et al.*, 1998b)
9. Echo state network (Jaeger and Haas, 2004)
10. Deep belief network (Hinton *et al.*, 2006)
11. GPU-accelerated convolutional network (Chellapilla *et al.*, 2006)
12. Deep Boltzmann machine (Salakhutdinov and Hinton, 2009a)
13. GPU-accelerated deep belief network (Raina *et al.*, 2009)
14. Unsupervised convolutional network (Jarrett *et al.*, 2009)
15. GPU-accelerated multilayer perceptron (Ciresan *et al.*, 2010)
16. OMP-1 network (Coates and Ng, 2011)
17. Distributed autoencoder (Le *et al.*, 2012)
18. Multi-GPU convolutional network (Krizhevsky *et al.*, 2012)
19. COTS HPC unsupervised convolutional network (Coates *et al.*, 2013)
20. GoogLeNet (Szegedy *et al.*, 2014a)

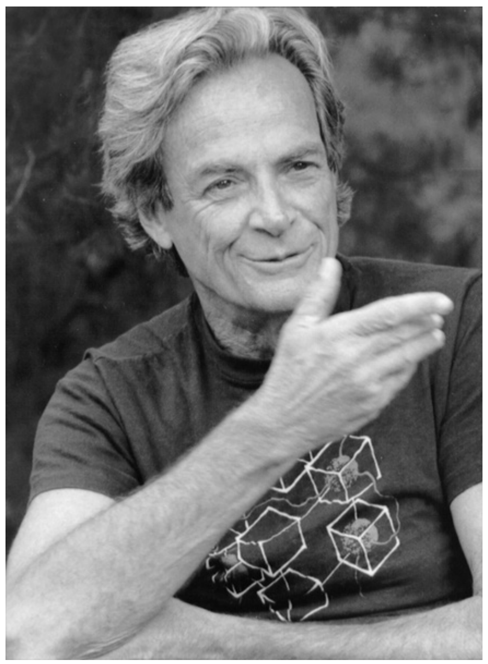
Discriminative vs Generative Learning

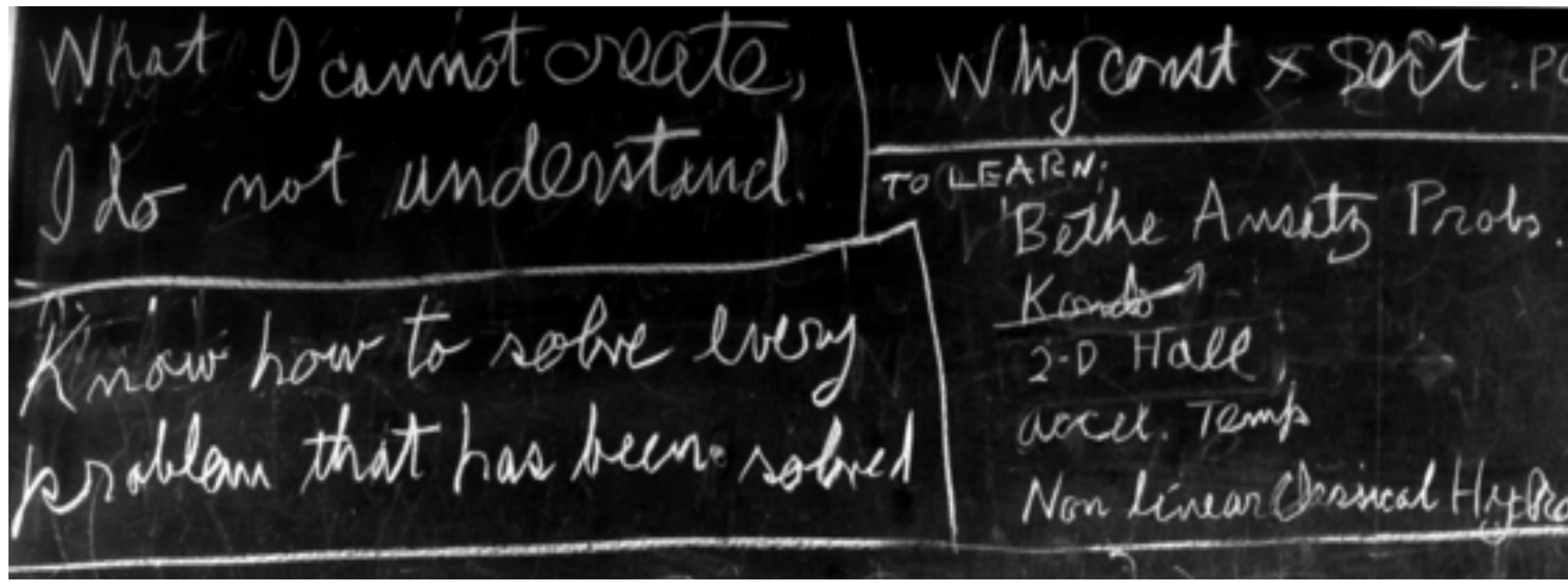
Discriminative vs Generative Learning

read

write

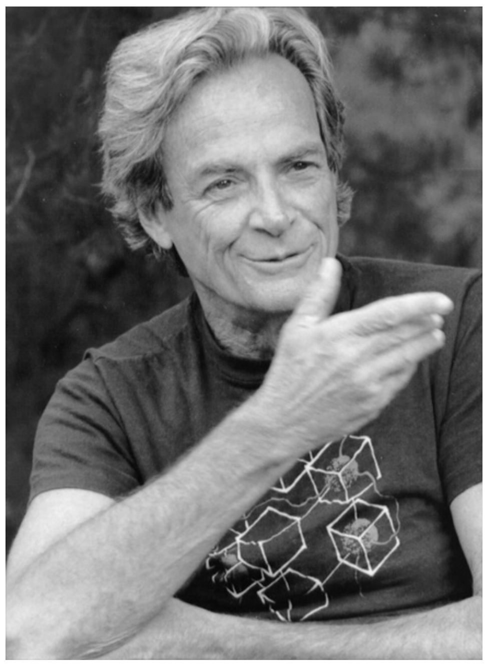
Discriminative vs Generative Learning

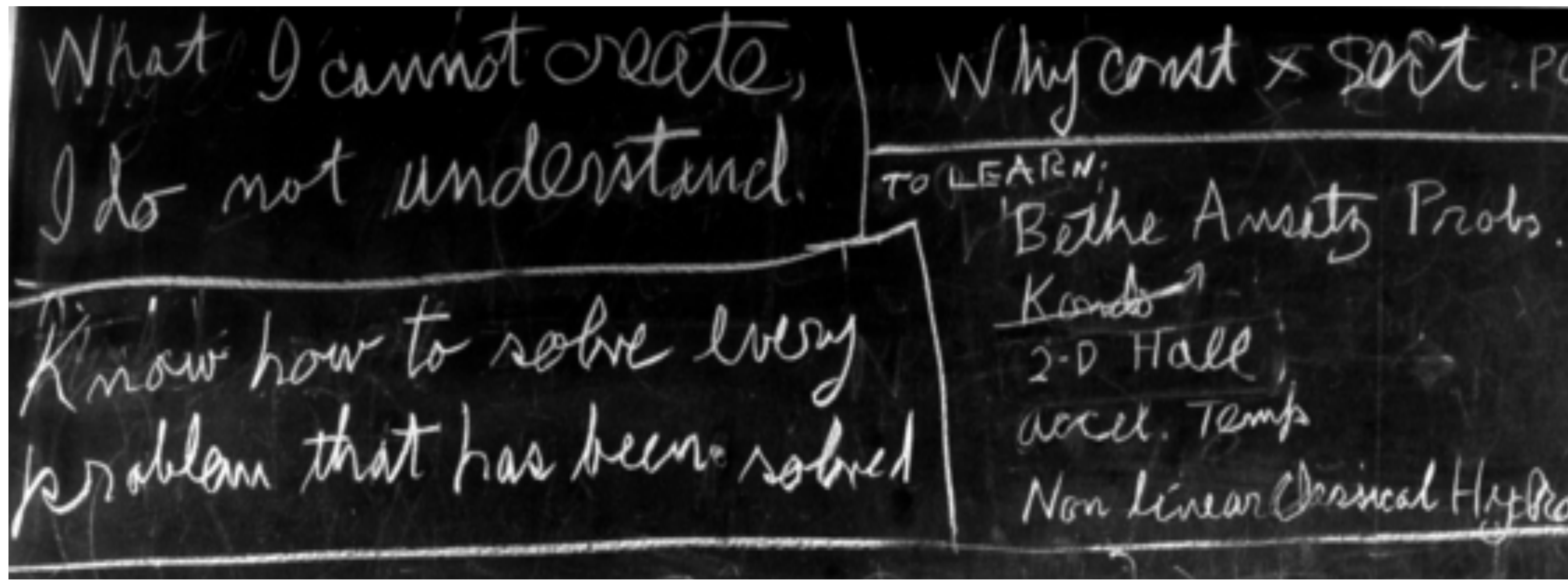




"What I can not create, I do not understand"

Discriminative vs Generative Learning



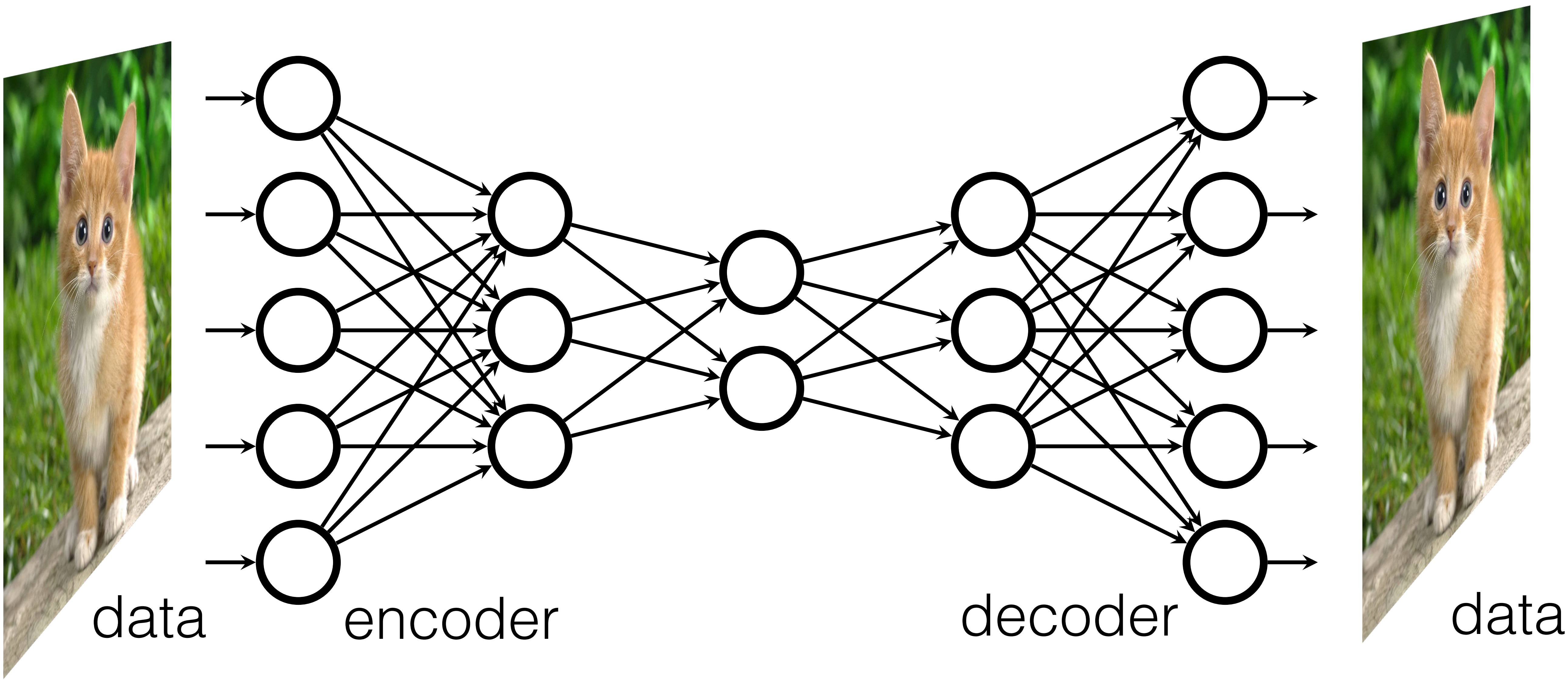


ELSEVIER
Progress in Brain Research
Volume 165, 2007, Pages 535–547
Computational Neuroscience: Theoretical Insights into Brain Function

To recognize shapes, first learn to generate images

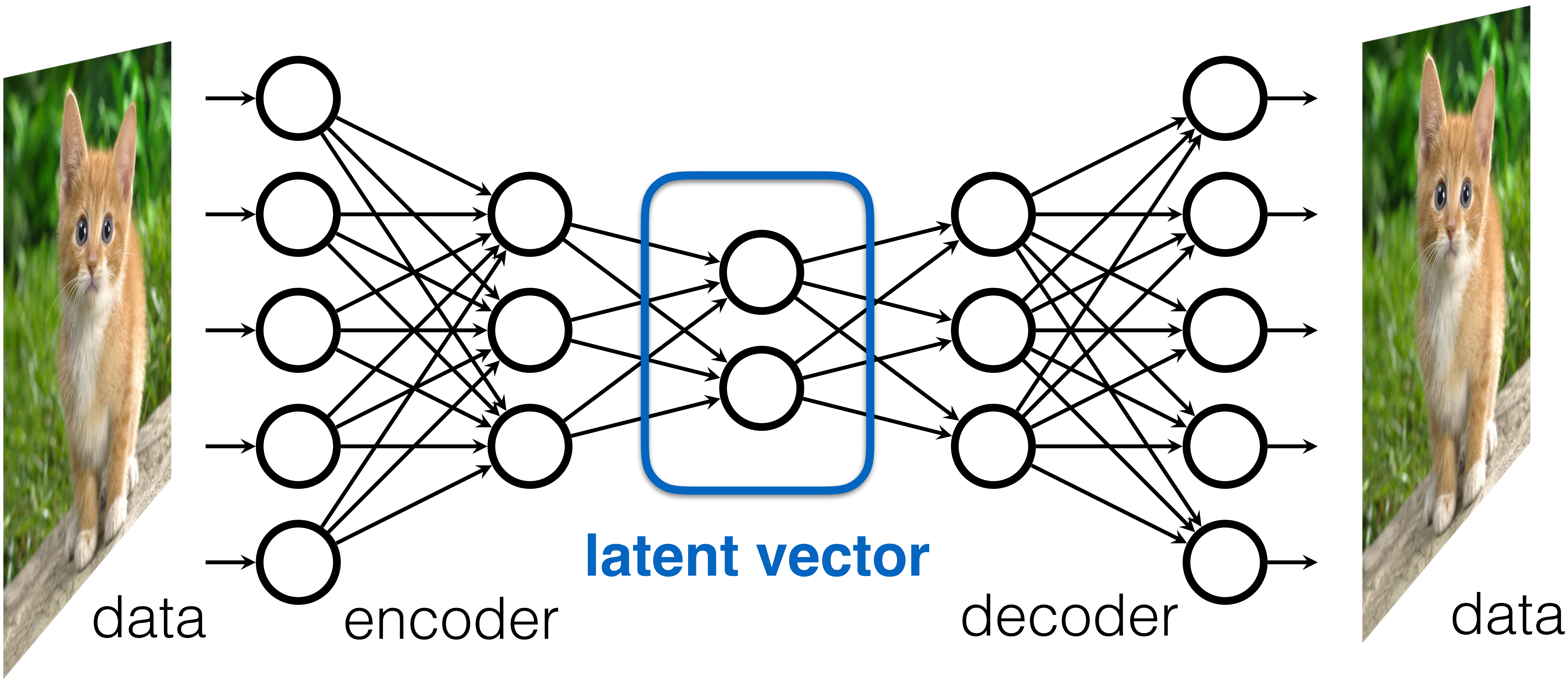
Geoffrey E. Hinton
Department of Computer Science, University of Toronto, 10 Kings College Road, Toronto, M5S 3G4
Canada

Generative Modeling



“Auto-Encoding Variational Bayes", Kingma and Welling,1312.6114

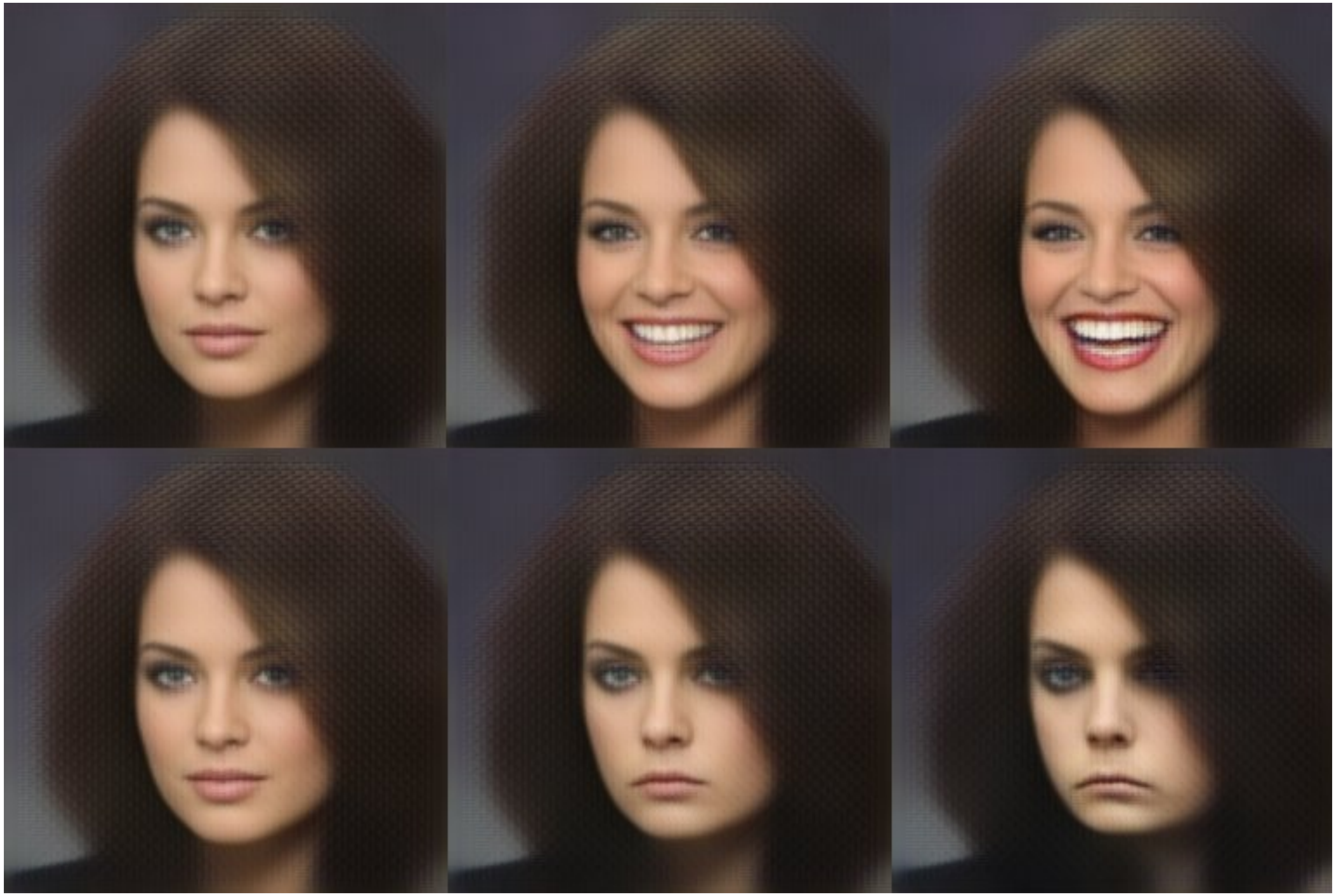
Generative Modeling



“Auto-Encoding Variational Bayes", Kingma and Welling,1312.6114

Latent space interpolation

arithmetics of the "smile vector"

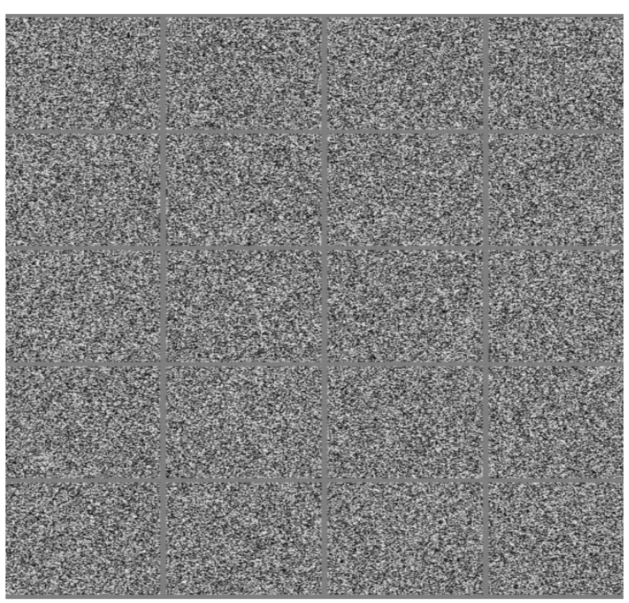


White, 1609.04468

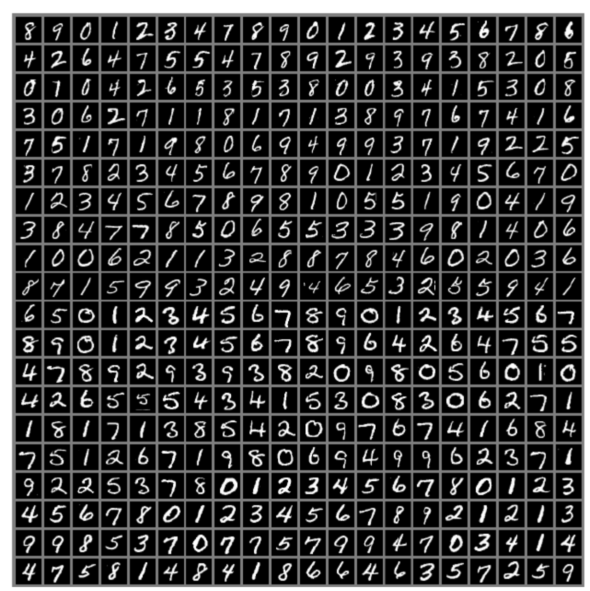
Probabilistic Generative Modeling

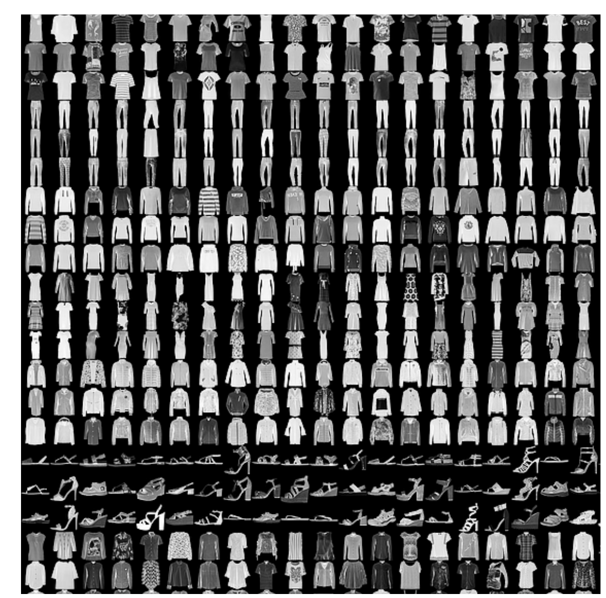
$$p(\mathbf{x})$$

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



“random” images

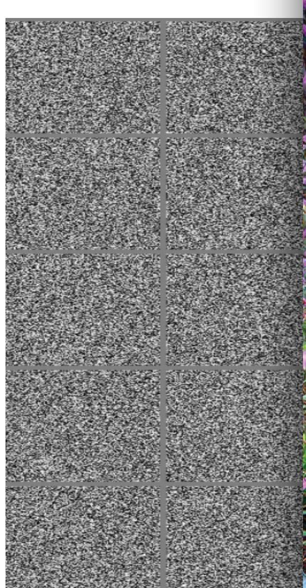


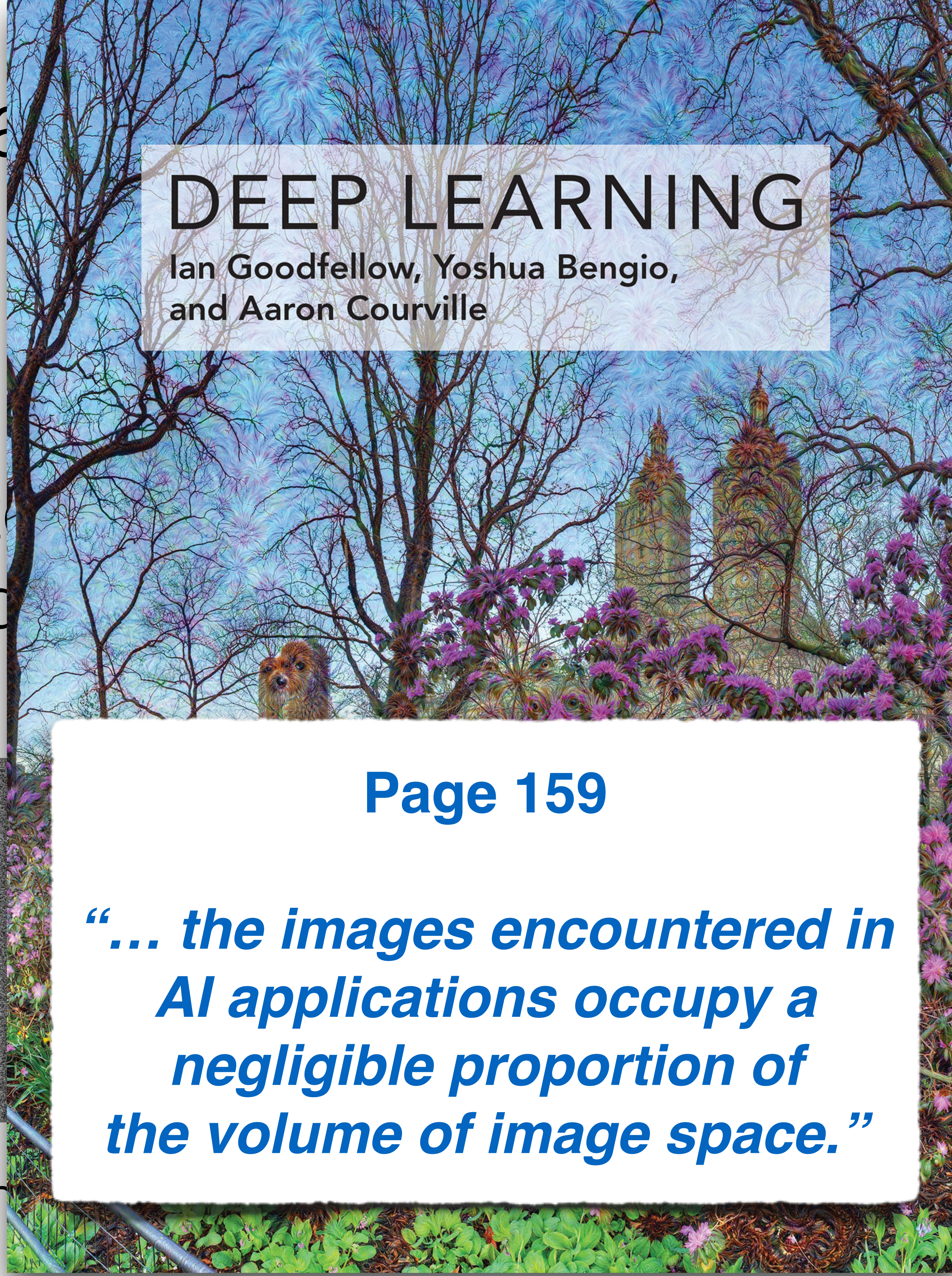


“natural” images

Proba … deling

How t … from a
high c … bution ?





Page 159

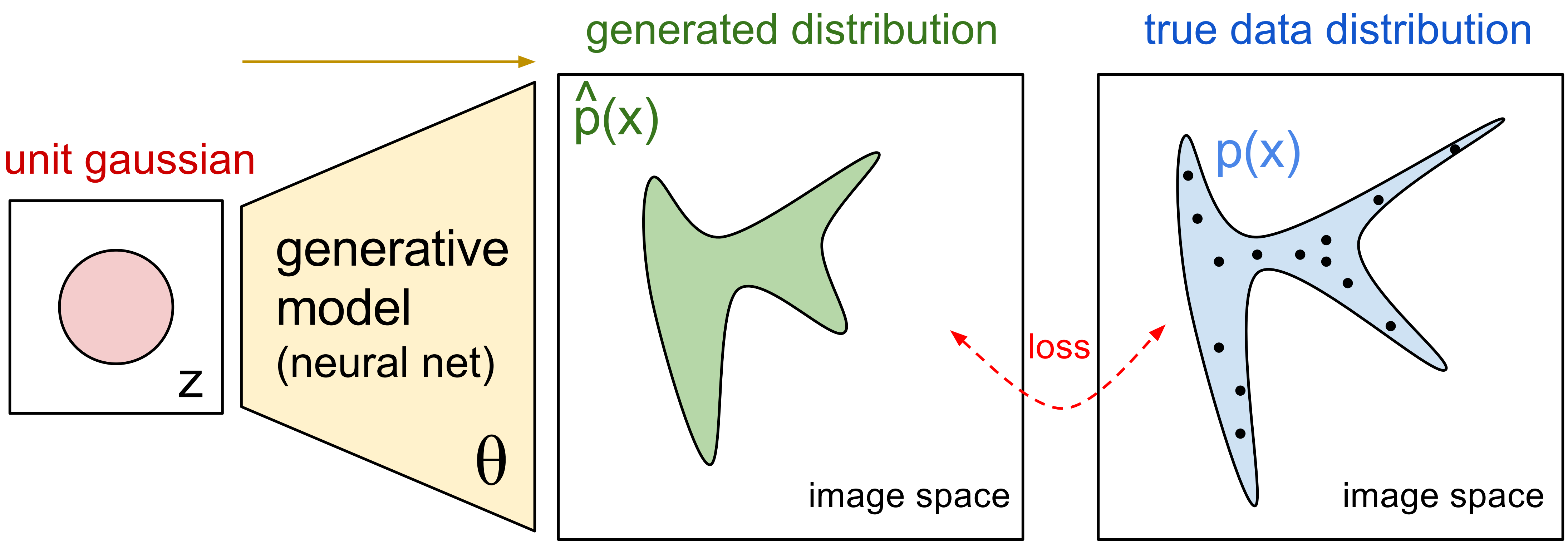
"… the images encountered in AI applications occupy a negligible proportion of the volume of image space."

"random

Probabilistic Generative Modeling

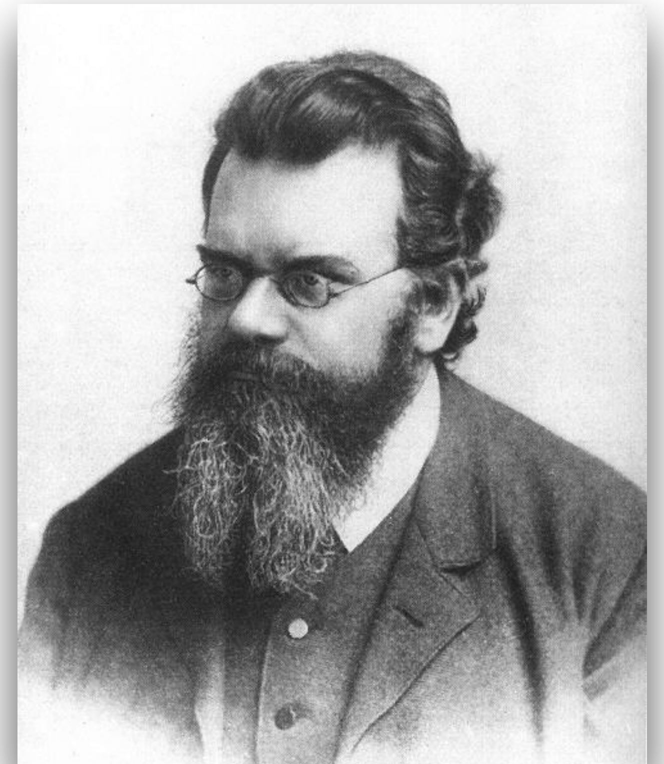
$$p(\mathbf{x})$$

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



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

Generative Modeling and Physics



"Boltzmann" Machines

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

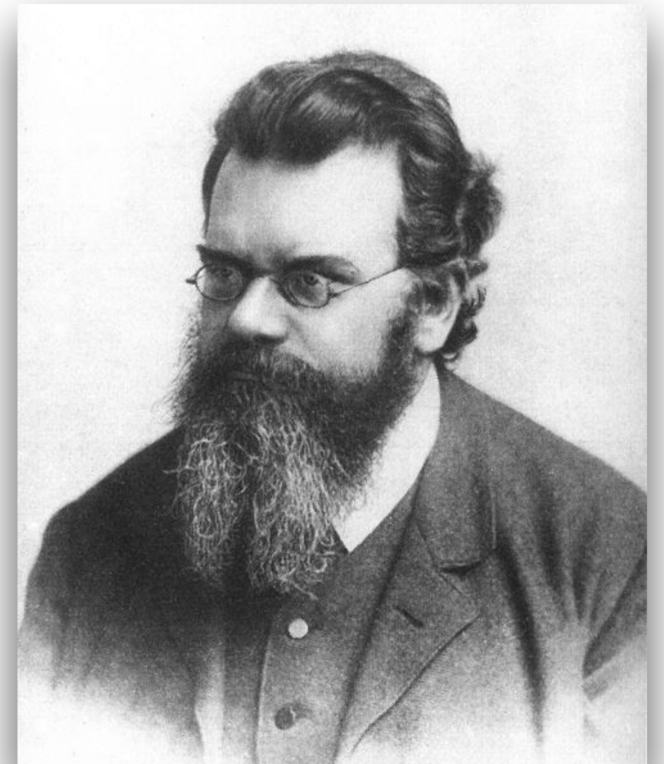
statistical physics

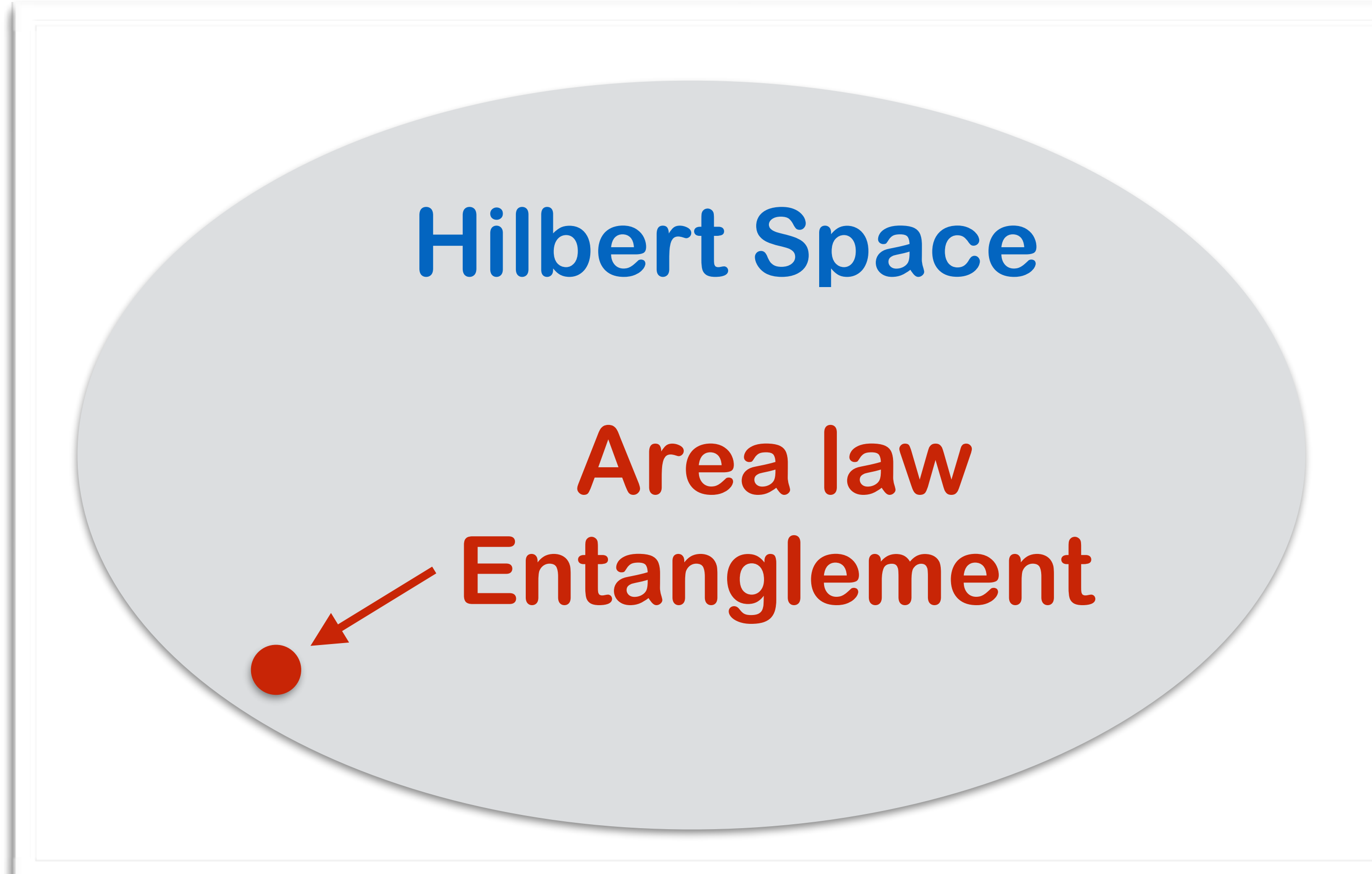
"Born" Machines

$$p(\mathbf{x}) = \frac{|\Psi(\mathbf{x})|^2}{\mathcal{N}}$$

quantum physics

Generative Modeling and Physics





"Boltzmann" Machines

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

statistical physics

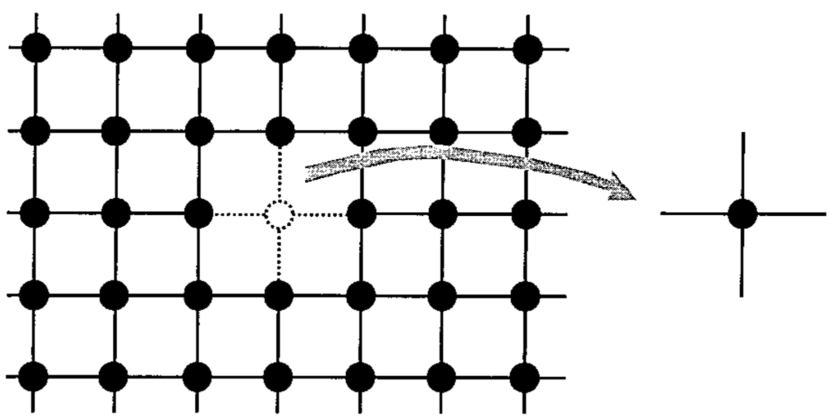
"Born" Machines

$$p(\mathbf{x}) = \frac{|\Psi(\mathbf{x})|^2}{\mathcal{N}}$$

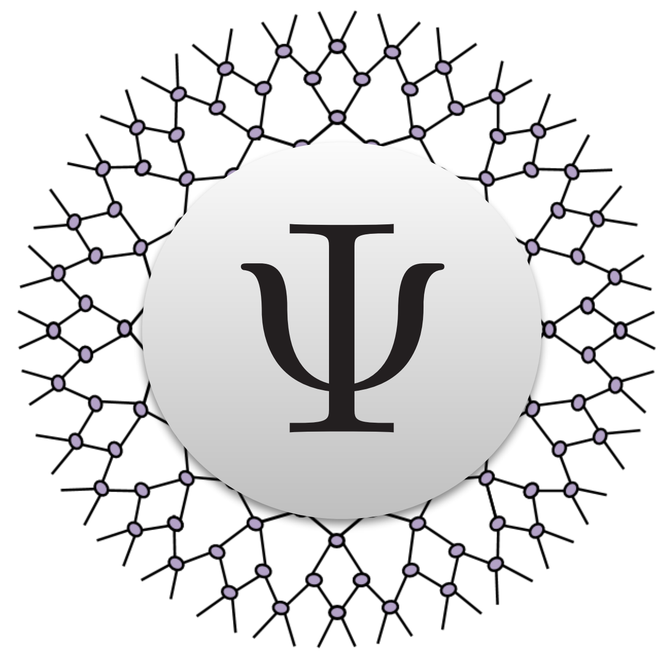
quantum physics

Physicists' gifts to ML

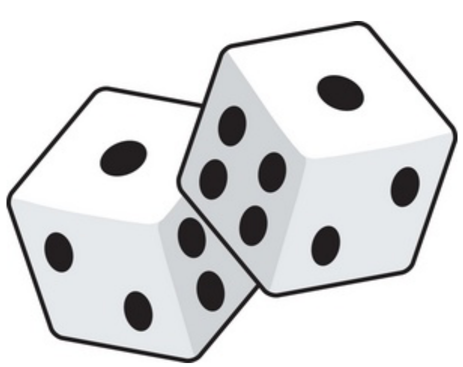
Mean Field Theory



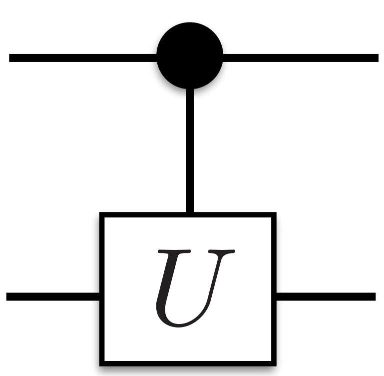
Tensor Networks



Monte Carlo Methods



Quantum Computing



Why machine learning for many-body physics ?

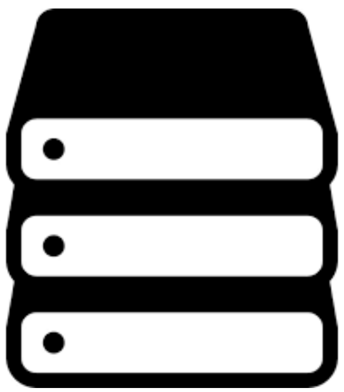
- Conceptual connections: a novel and natural way to think about (quantum) many-body systems

- Data driven approach: making scientific discovery based on big data

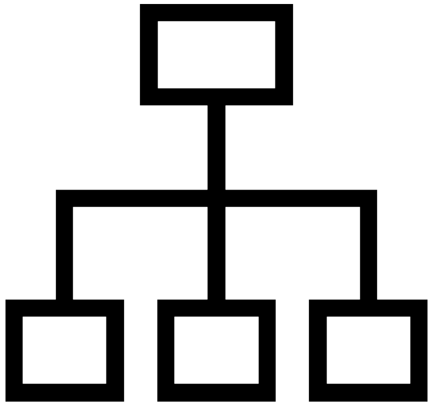
- Techniques: neural networks, kernel methods, pattern recognition, feature extraction, dimensional reduction, clustering analysis, probabilistic modeling, recommender systems, expectation maximization, variational inference, hardware acceleration, software frameworks…

Four Pillars of Machine Learning

Data



Model



Cost function

Optimization

Switch to blackboard

Winter School on Numerical Methods for Strongly Correlated Quantum Systems, February 19-23 2018, Marburg

Deep Learning and Quantum Many-Body Computation

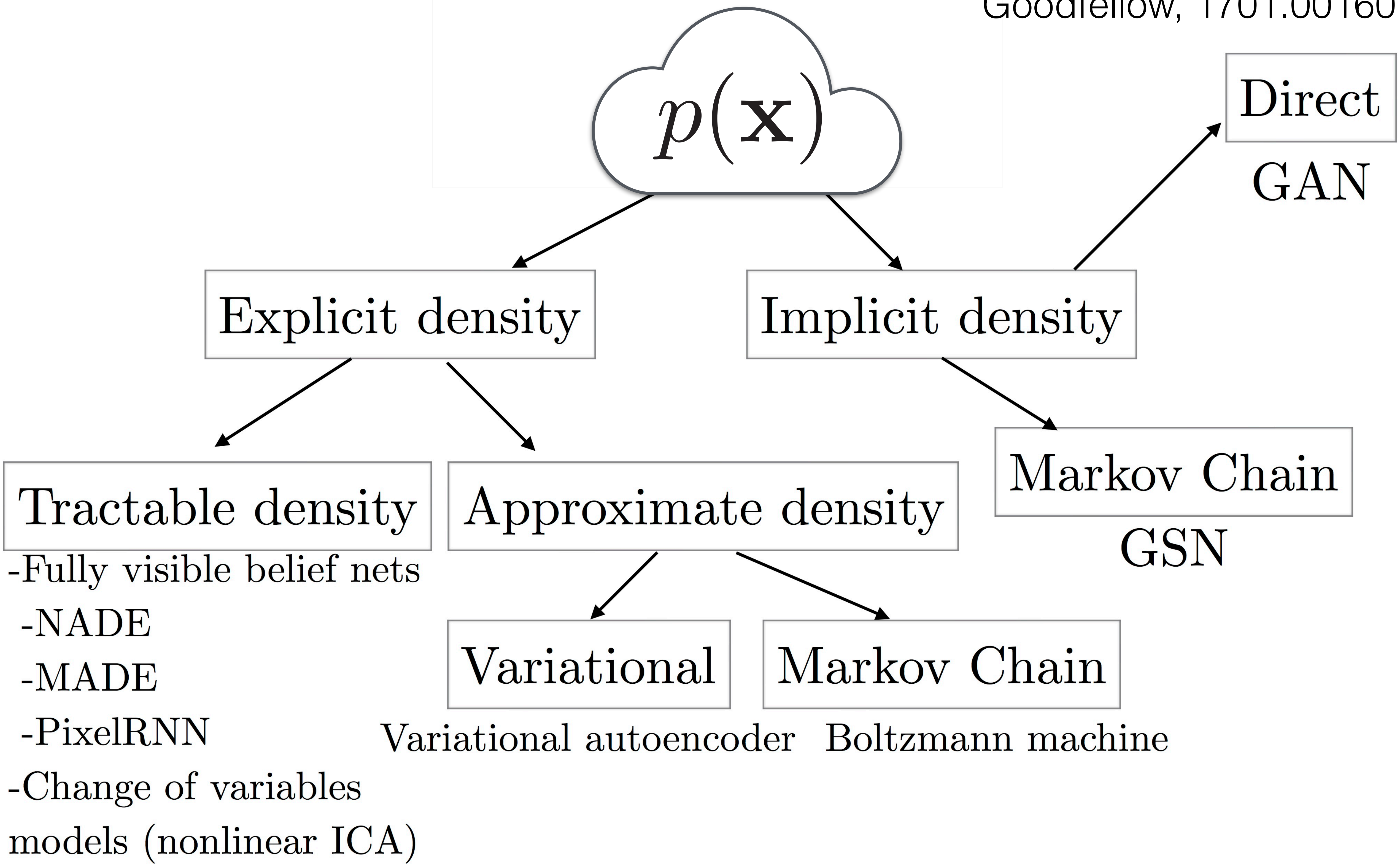
Lei Wang（王磊）

Institute of Physics, CAS

https://wangleiphy.github.io

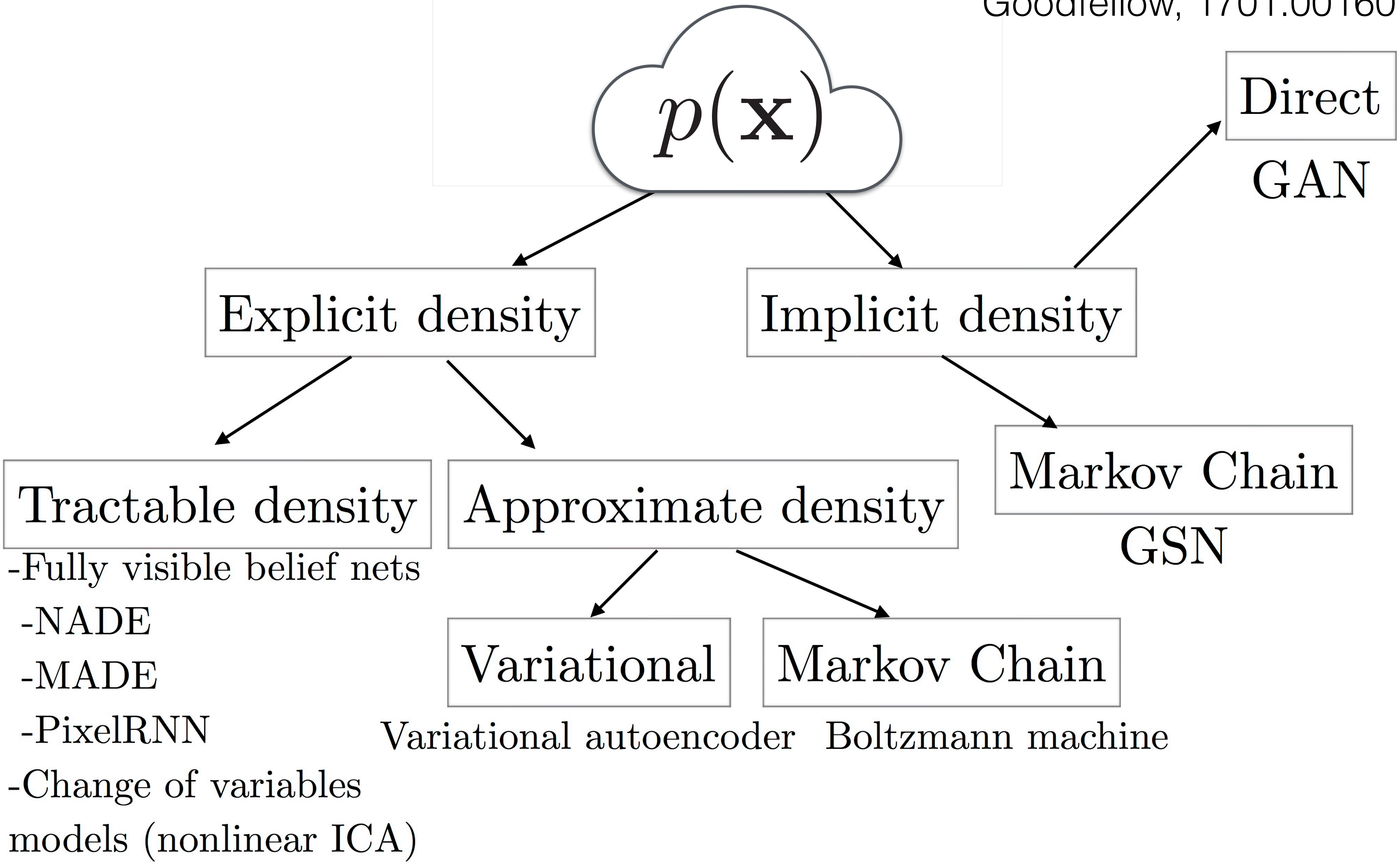
Taxonomy of Generative Models

Goodfellow, 1701.00160



Taxonomy of Generative Models

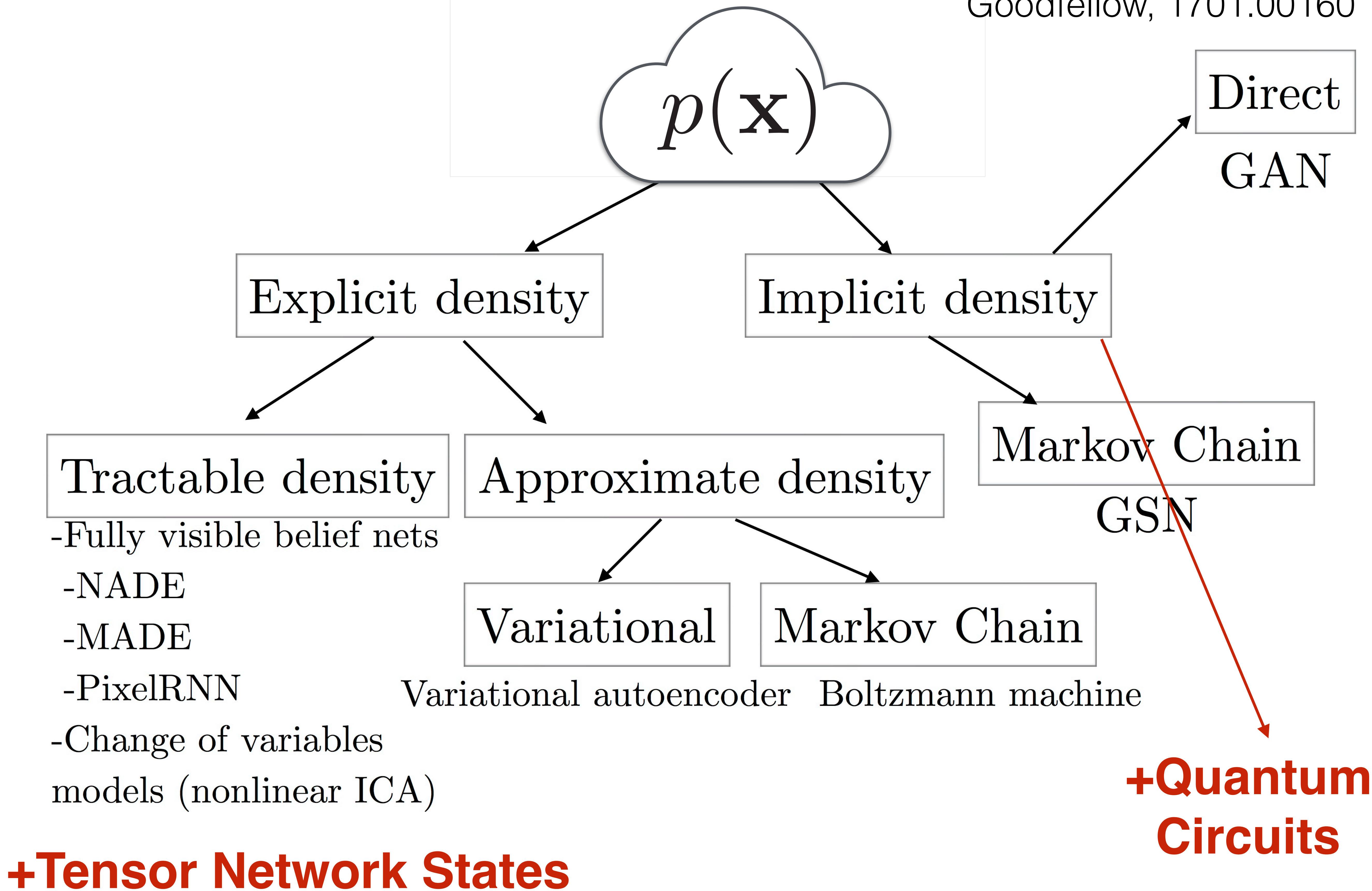
Goodfellow, 1701.00160



+Tensor Network States

Taxonomy of Generative Models

Goodfellow, 1701.00160



Lecture note http://wangleiphy.github.io/lectures/DL.pdf

Table 2: A summary of generative models and their salient features. Question marks mean generalizations are possible, but nontrivial.

Name	Training Cost	Data Space	Latent Space	Architecture	Sampling	Likelihood	Expressibility	Difficulty (Learn/Sample)
RBM	Log-likelihood	Arbitrary	Arbitrary	Bipartite	MCMC	Intractable partition function	★	☠☠/☠☠☠
DBM	ELBO	Arbitrary	Arbitrary	Bipartite	MCMC	Intractable partition function & posterior	★★★	☠☠/☠☠☠
Autoregressive Model	Log-likelihood	Arbitrary	None	Ordering	Sequential	Tractable	★★	☠/☠☠
Normalizing Flow	Log-likelihood	Continuous	Continuous, Same dimension as data	Bijector	Parallel	Tractable	★★	☠/☠
VAE	ELBO	Arbitrary	Continuous	Arbitrary?	Parallel	Intractable posterior	★★★	☠/☠
MPS/TTN	Log-likelihood	Arbitrary?	None or tree tensor	No loop	Sequential	Tractable	★★★	☠☠/☠☠
GAN	Adversarial	Continuous	Arbitrary?	Arbitrary	Parallel	Implicit	★★★★	☠☠☠/☠

WARNING

The following content may contain spoilers

They may spoil your fun of imagination & creation

Proceed with caution!!!

Material and Chemical Discovery

Material Discovery

Microscopic Composition

Data Generation

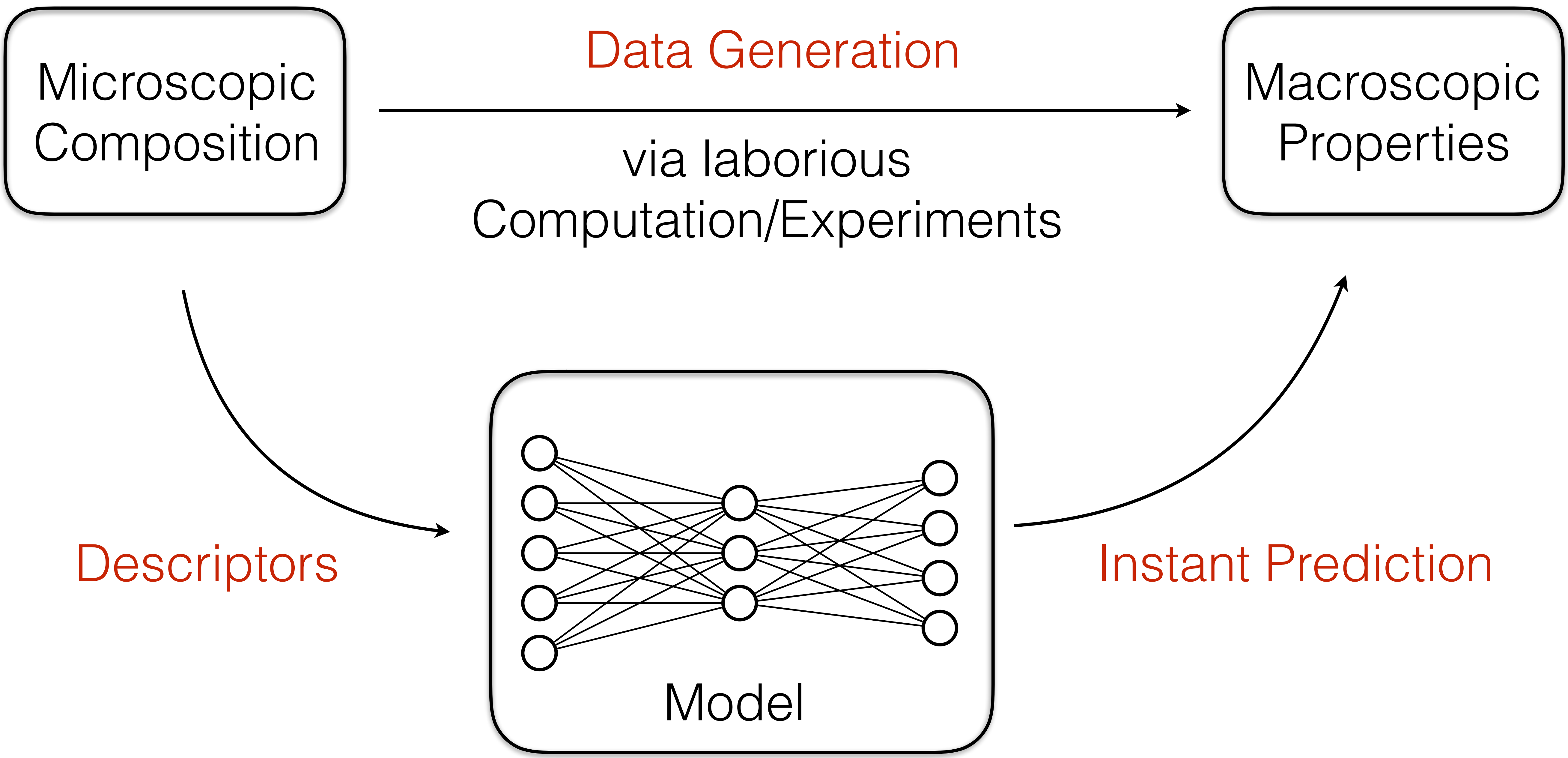
→

via laborious Computation/Experiments

Macroscopic Properties

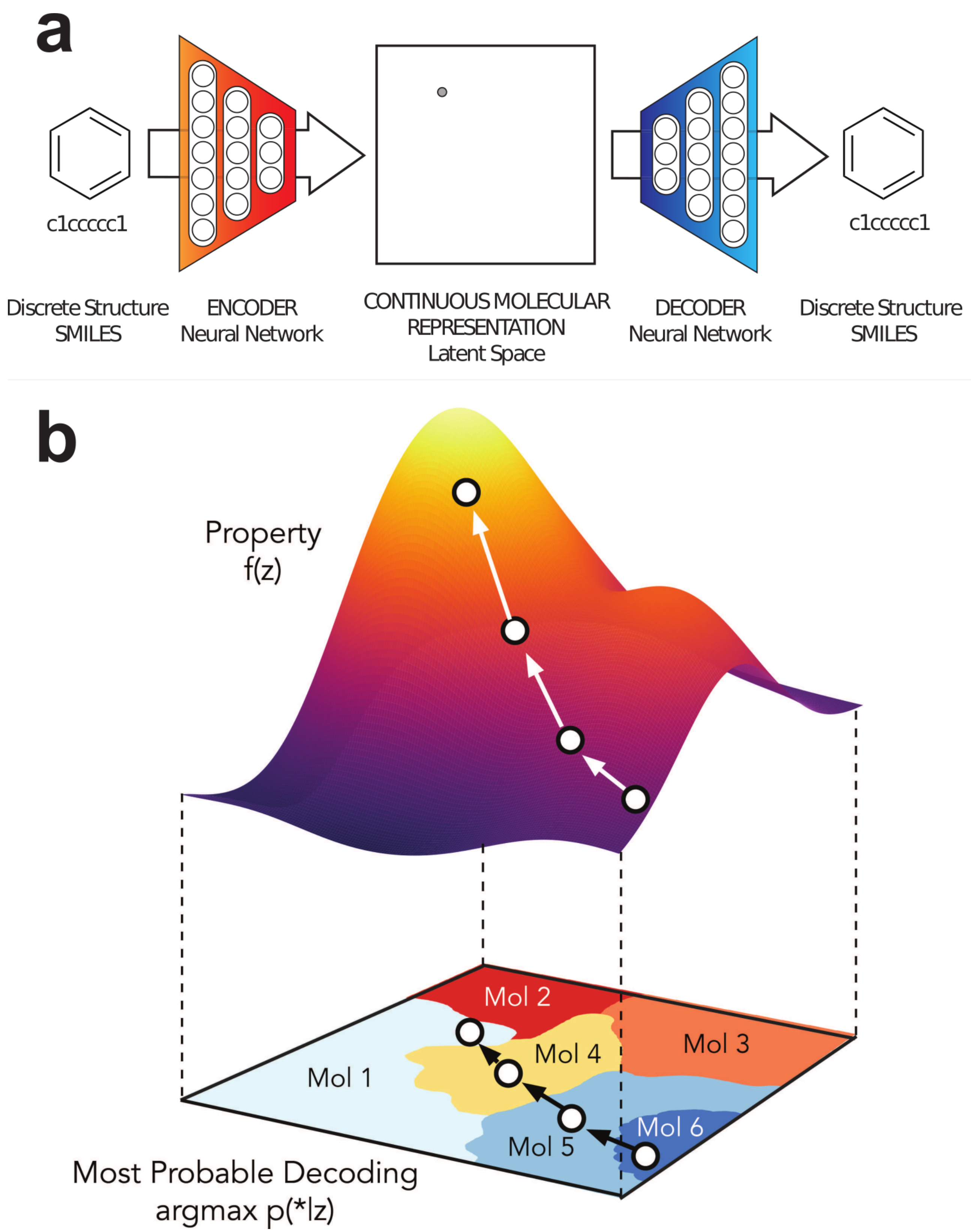
“Machine Learning in Materials Science: Recent Progress and Critical Next Steps”
Rampi Ramprasad @ IPAM program on Understanding Many-Particle Systems with ML, 2016

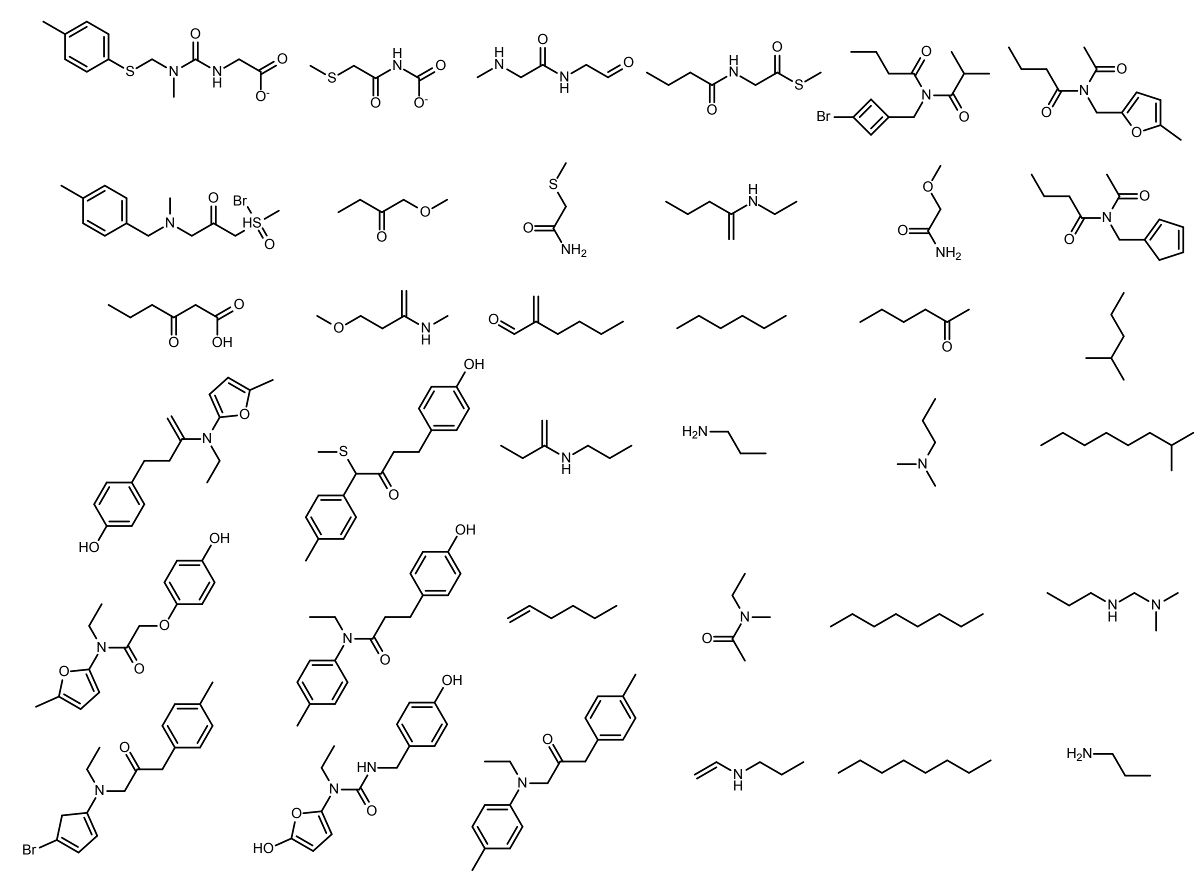
Material Discovery



"Machine Learning in Materials Science: Recent Progress and Critical Next Steps"
Rampi Ramprasad @ IPAM program on Understanding Many-Particle Systems with ML, 2016

Chemical design using VAE





interpolation of the molecules at the four corners

Gomez-Bombarelli et al, 1610.02415

Density Functionals

Finding the density-functional

Table 1. *Physical Review* Articles with more than 1000 Citations Through June 2003

Publication	# cites	Av. age	Title	Author(s)
PR **140**, A1133 (1965)	3227	26.7	Self-Consistent Equations Including Exchange and Correlation Effects	W. Kohn, L. J. Sham
PR **136**, B864 (1964)	2460	28.7	Inhomogeneous Electron Gas	P. Hohenberg, W. Kohn
PRB **23**, 5048 (1981)	2079	14.4	Self-Interaction Correction to Density-Functional Approximations for Many-Electron Systems	J. P. Perdew, A. Zunger
PRL **45**, 566 (1980)	1781	15.4	Ground State of the Electron Gas by a Stochastic Method	D. M. Ceperley, B. J. Alder
PR **108**, 1175 (1957)	1364	20.2	Theory of Superconductivity	J. Bardeen, L. N. Cooper, J. R. Schrieffer
PRL **19**, 1264 (1967)	1306	15.5	A Model of Leptons	S. Weinberg
PRB **12**, 3060 (1975)	1259	18.4	Linear Methods in Band Theory	O. K. Anderson
PR **124**, 1866 (1961)	1178	28.0	Effects of Configuration Interaction of Intensities and Phase Shifts	U. Fano
RMP **57**, 287 (1985)	1055	9.2	Disordered Electronic Systems	P. A. Lee, T. V. Ramakrishnan
RMP **54**, 437 (1982)	1045	10.8	Electronic Properties of Two-Dimensional Systems	T. Ando, A. B. Fowler, F. Stern
PRB **13**, 5188 (1976)	1023	20.8	Special Points for Brillouin-Zone Integrations	H. J. Monkhorst, J. D. Pack

PR, Physical Review; PRB, Physical Review B; PRL, Physical Review Letters; RMP, Reviews of Modern Physics.

Top four most-cited PR papers are all DFT-foundational papers!

Finding the density-functional

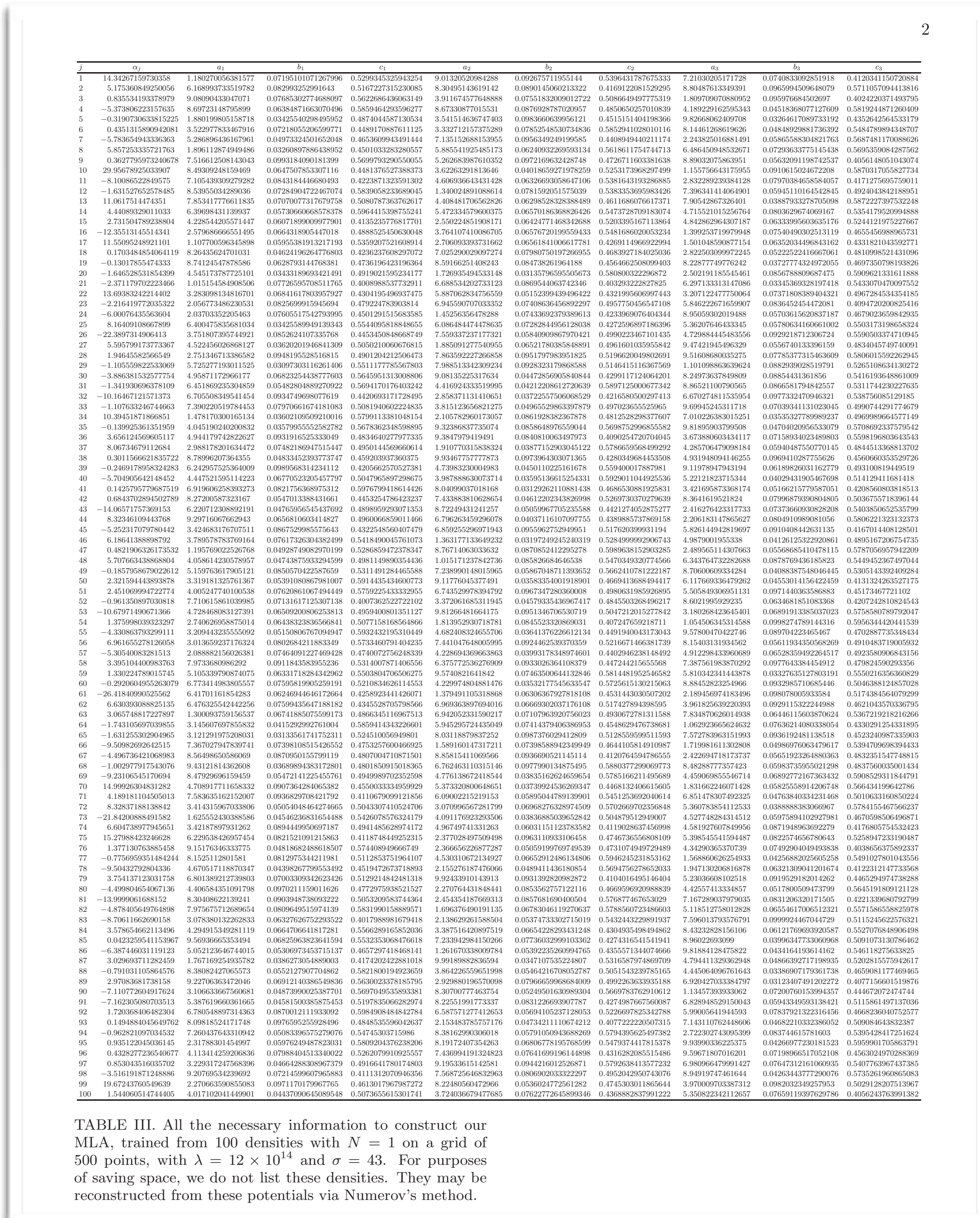
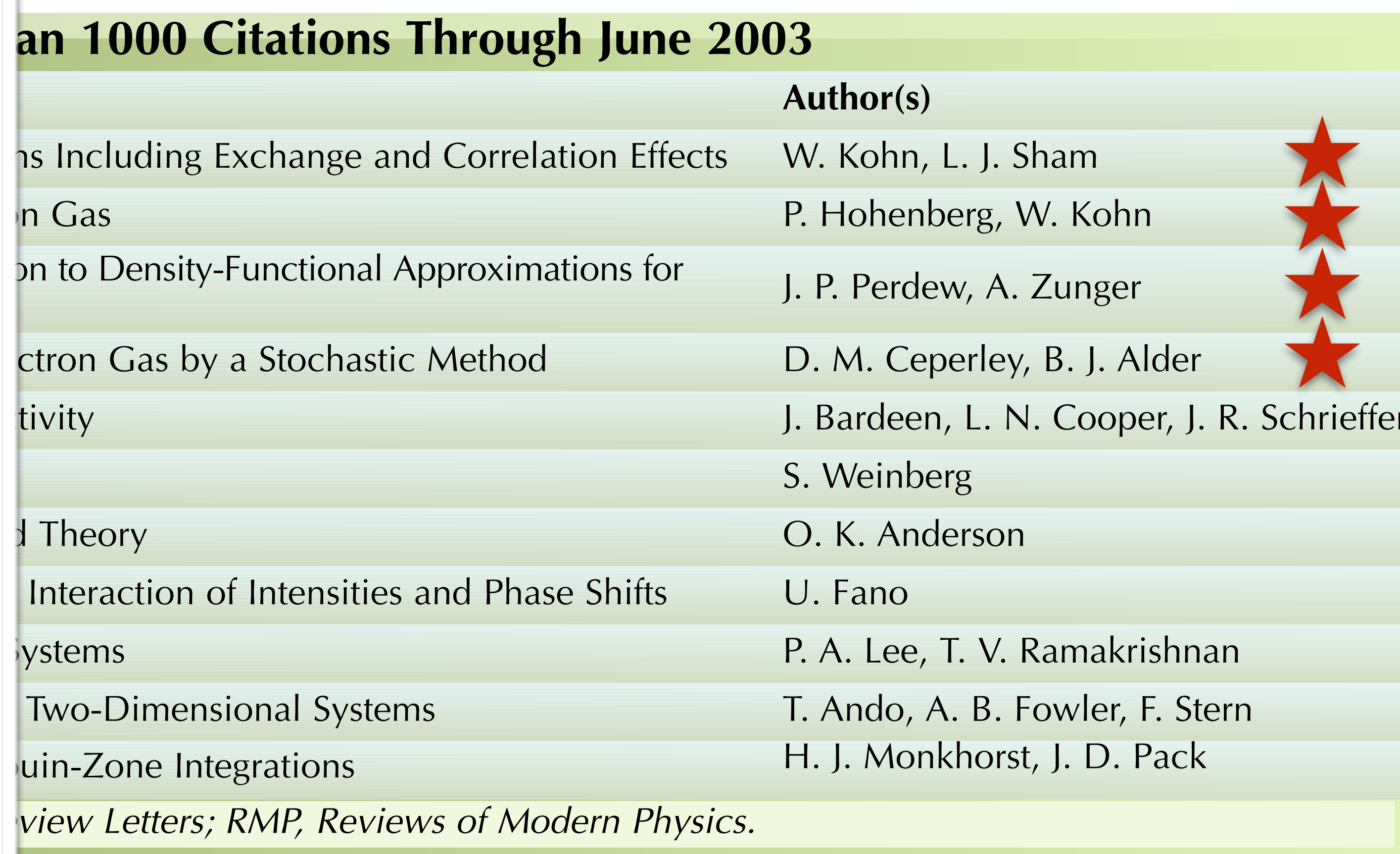

2

TABLE III. All the necessary information to construct our MLA, trained from 100 densities with $N = 1$ on a grid of 500 points, with $\lambda = 12 \times 10^{14}$ and $\sigma = 43$. For purposes of saving space, we do not list these densities. They may be reconstructed from these potentials via Numerov's method.


an 1000 Citations Through June 2003

	Author(s)
ns Including Exchange and Correlation Effects	W. Kohn, L. J. Sham
n Gas	P. Hohenberg, W. Kohn
on to Density-Functional Approximations for	J. P. Perdew, A. Zunger
ctron Gas by a Stochastic Method	D. M. Ceperley, B. J. Alder
tivity	J. Bardeen, L. N. Cooper, J. R. Schrieffer
	S. Weinberg
d Theory	O. K. Anderson
Interaction of Intensities and Phase Shifts	U. Fano
ystems	P. A. Lee, T. V. Ramakrishnan
Two-Dimensional Systems	T. Ando, A. B. Fowler, F. Stern
uin-Zone Integrations	H. J. Monkhorst, J. D. Pack

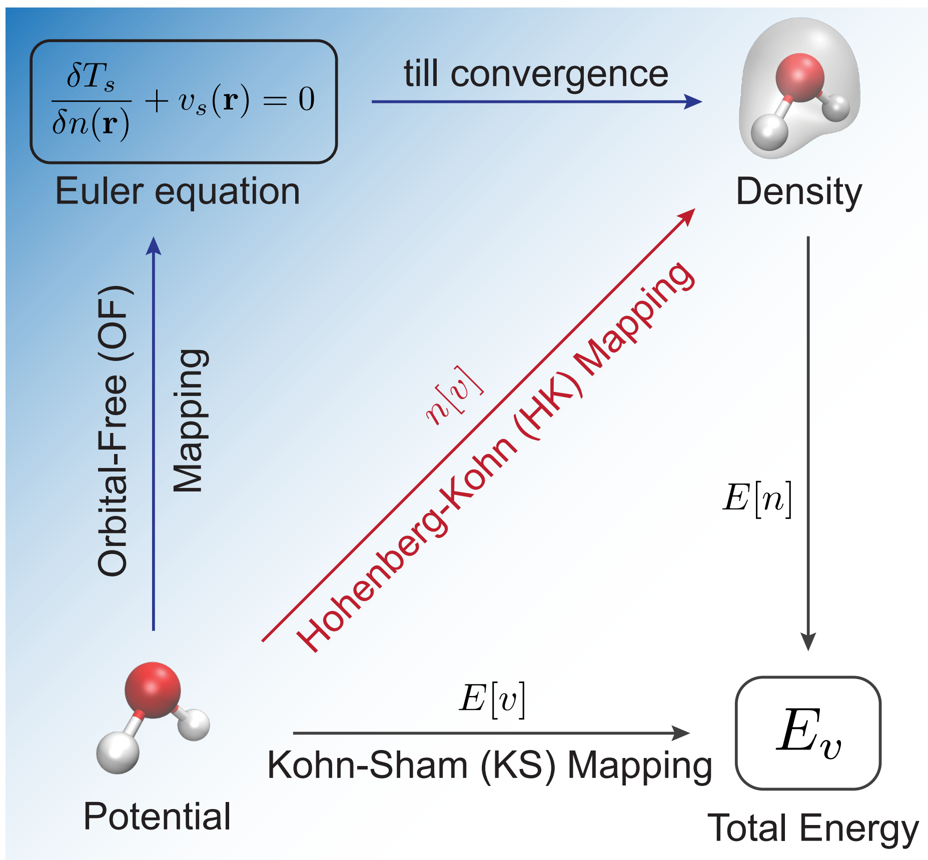
view Letters; RMP, Reviews of Modern Physics.

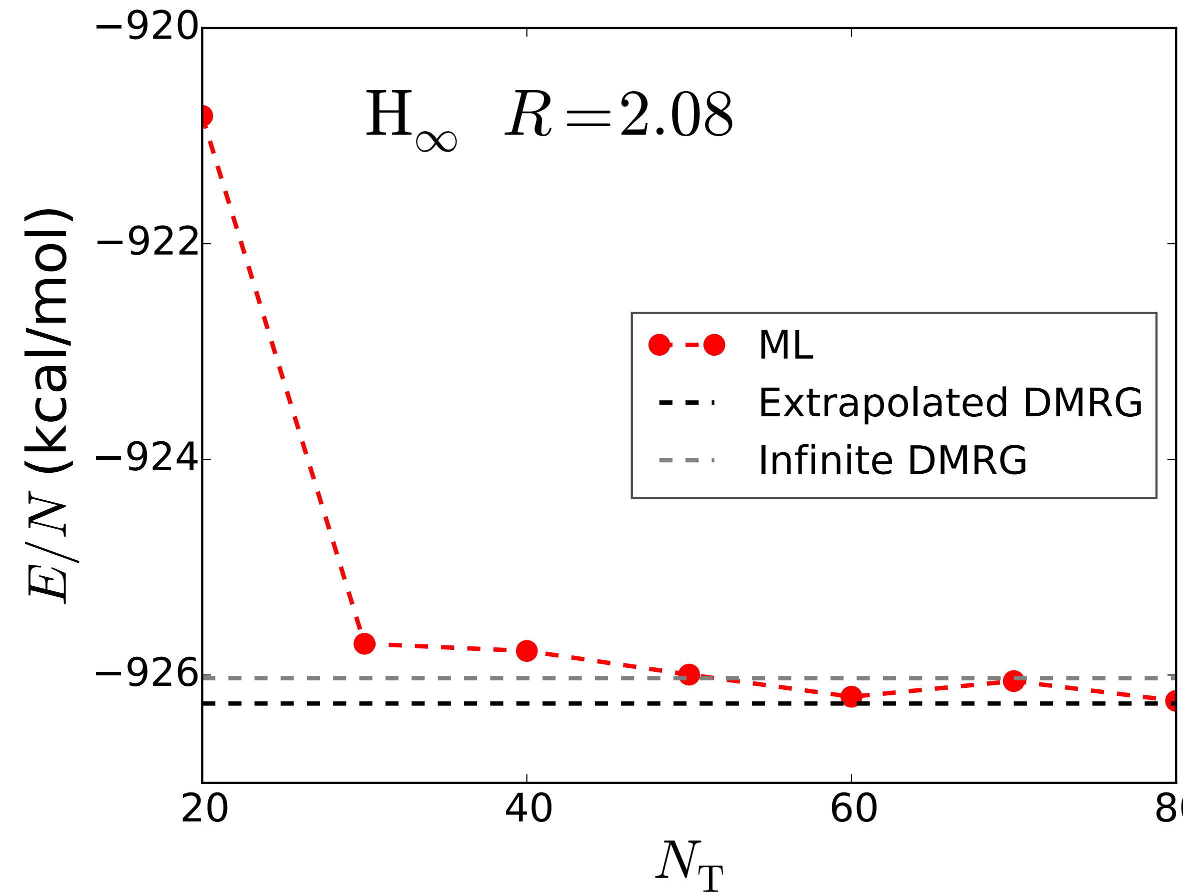
ers are all DFT-foundational papers!

machine learned one dimensional kinetic-energy functional with O(10^3) parameters

Snyder et al, PRL 108, 253002 (2012)

Recent Progresses





Avoiding Euler equation by directly learning n[v] map

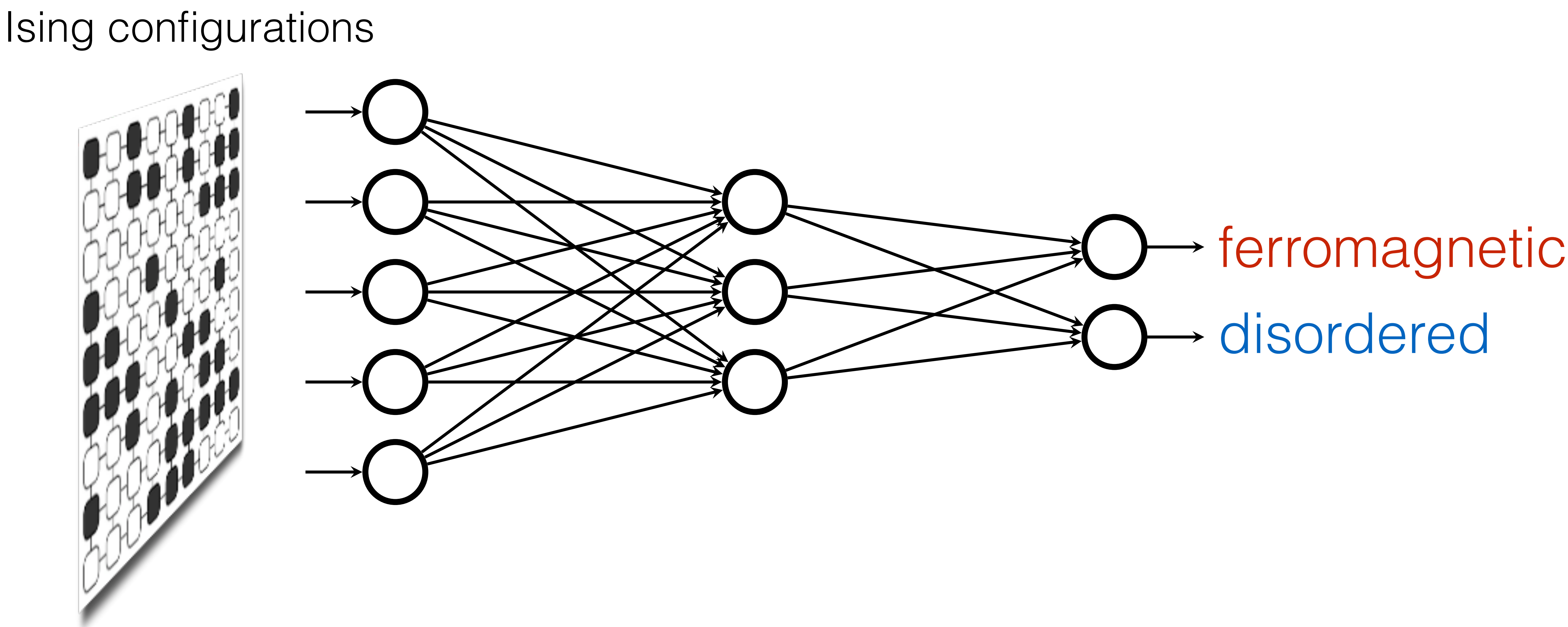
Beockherde, Vogt, Li, Tuckerman, Burke, Meuller 1609.02815

Learning the xc-functional from DMRG calculation in 1D continuous space

Li, Baker, White, Burke, 1609.03705

“Phase” Recognition

Supervised Approach



data

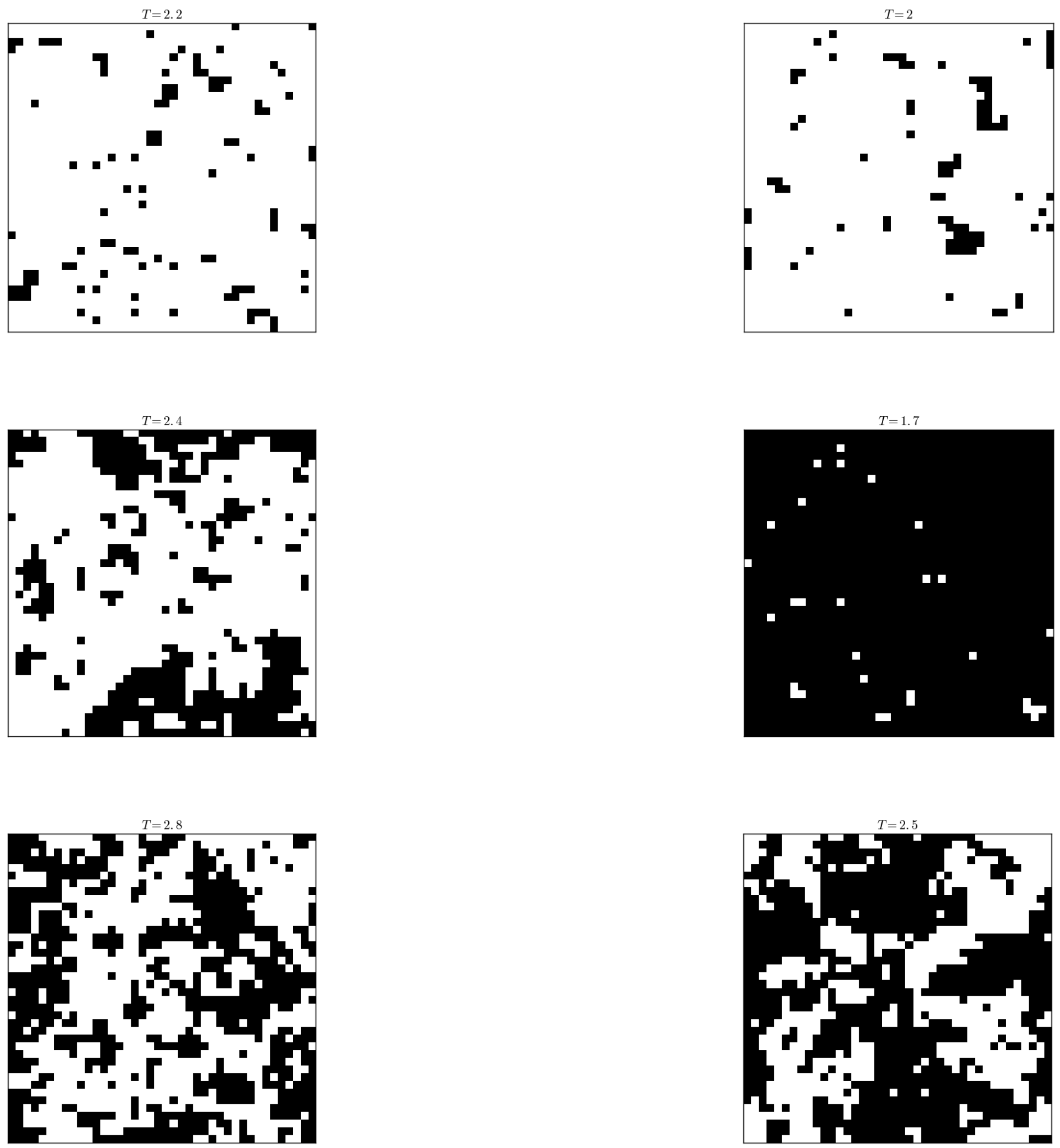
"Machine Learning Phase of Matter"

label

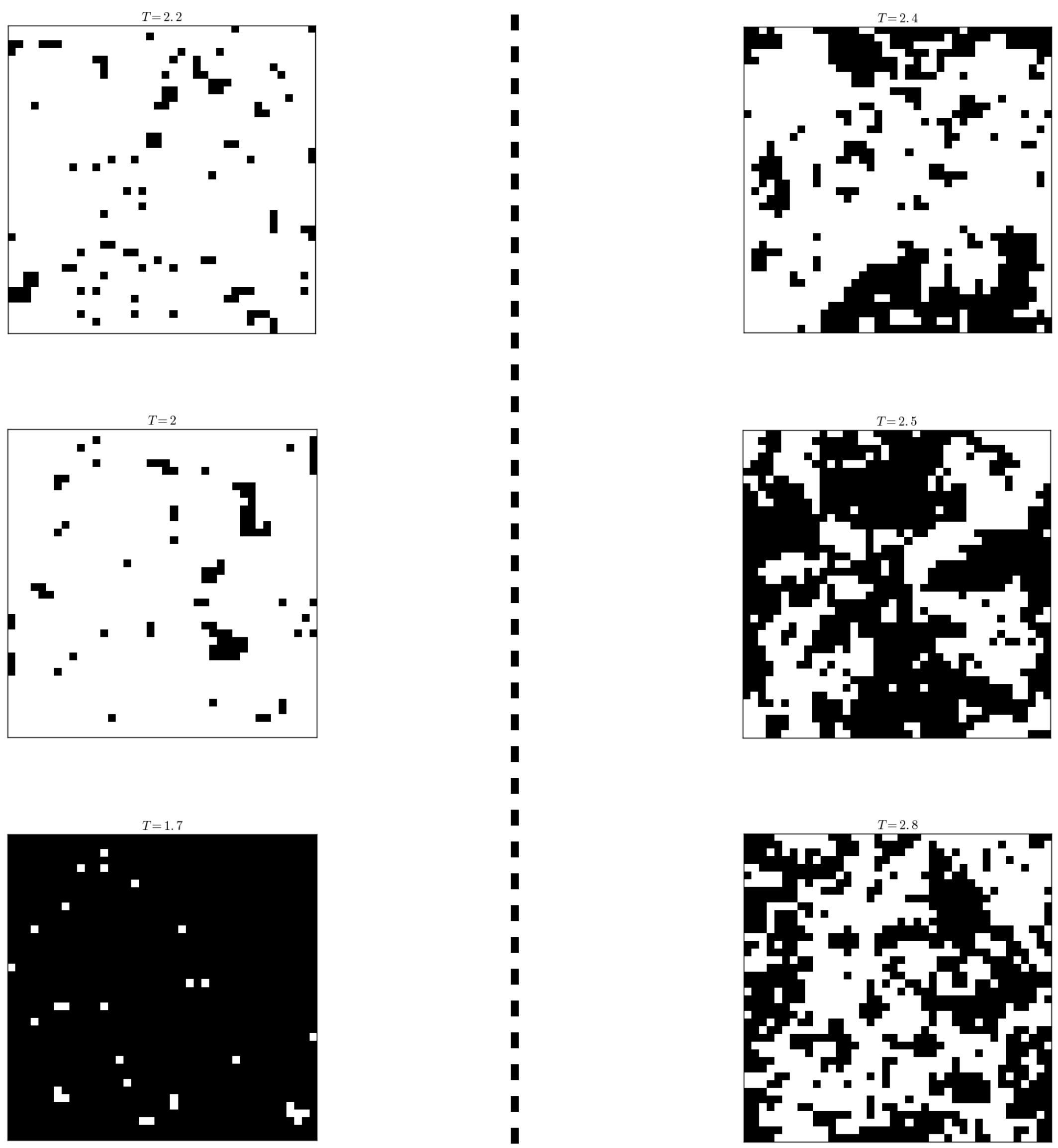
Carrasquilla and Melko, 1605.01735

+ many more on quantum spins, fermions, disordered systems, topological models …

Unsupervised Approach



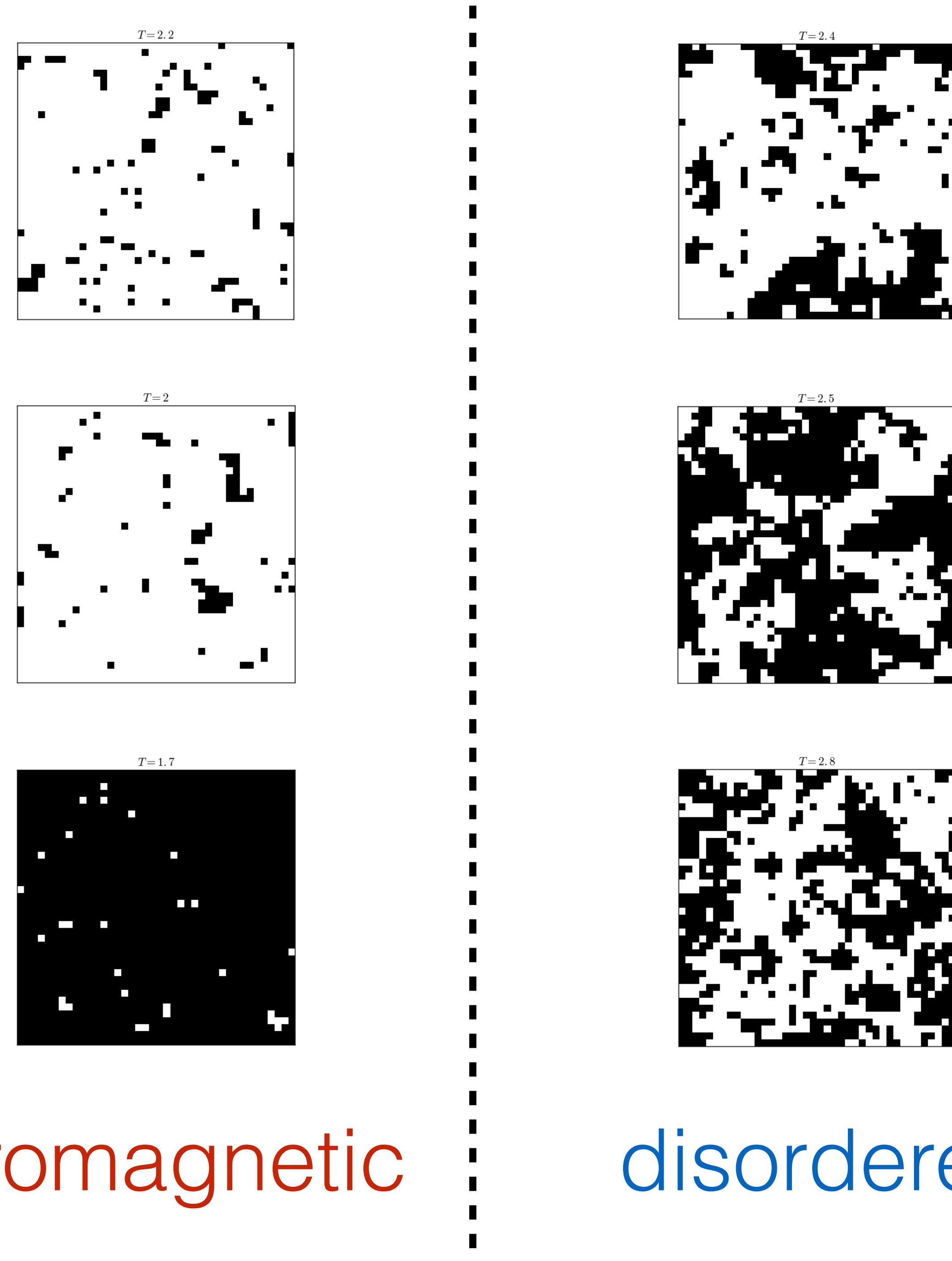
Unsupervised Approach



ferromagnetic

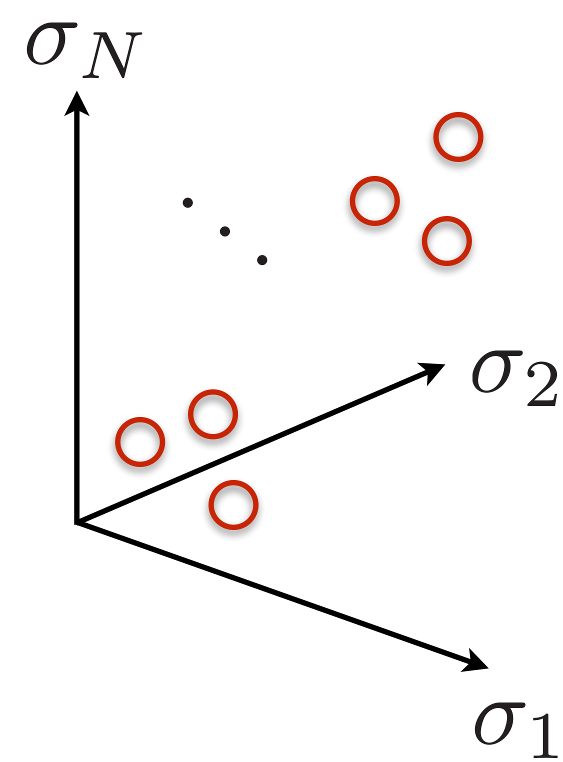
disordered

Unsupervised Approach



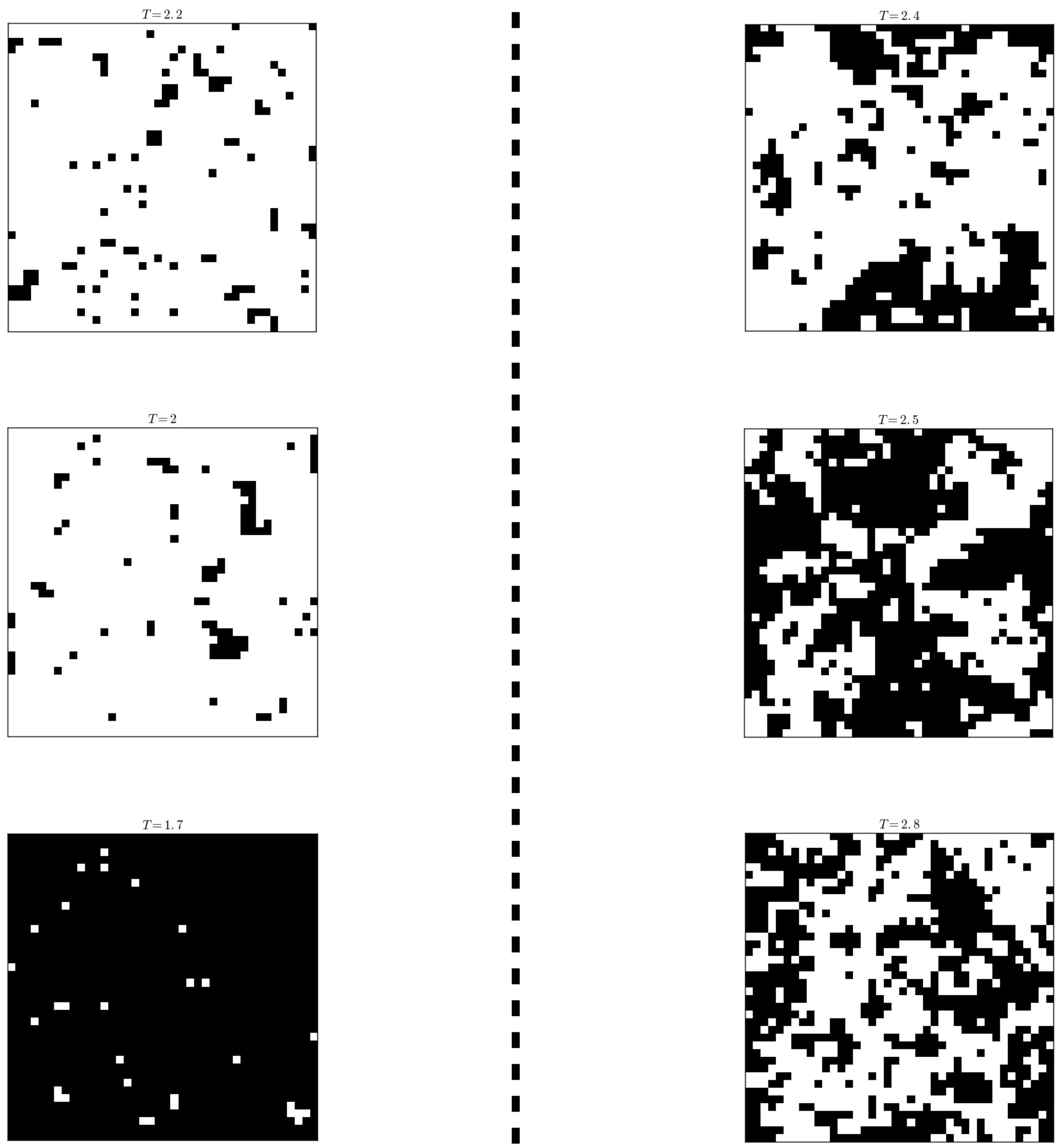
ferromagnetic

disordered



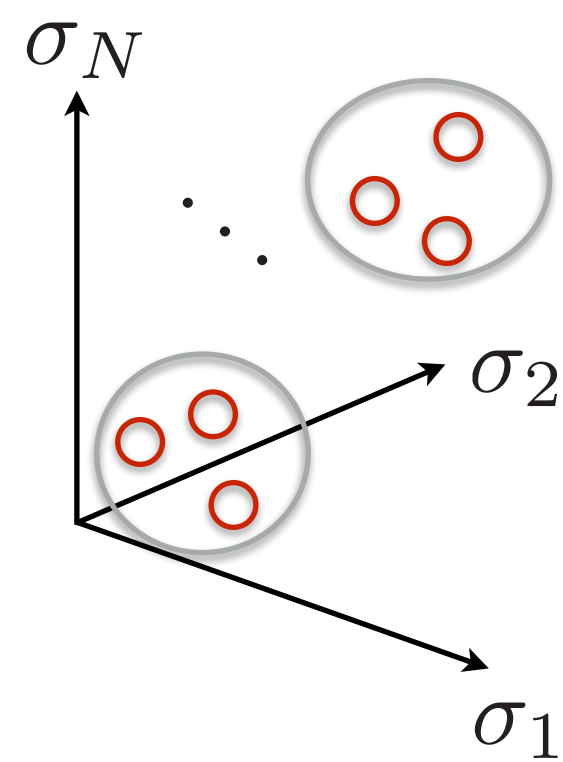
only data, no label

Unsupervised Approach



ferromagnetic | disordered

LW, 1606.00318
Discovering phase transition with dimensional reduction and clustering analysis



only data, no label

Different Schemes

Nieuwenburg, Liu, Huber, 1610.02048
Liu, Nieuwenburg, 1706.08111
Broecker, Assaad, Trebst, 1707.00663

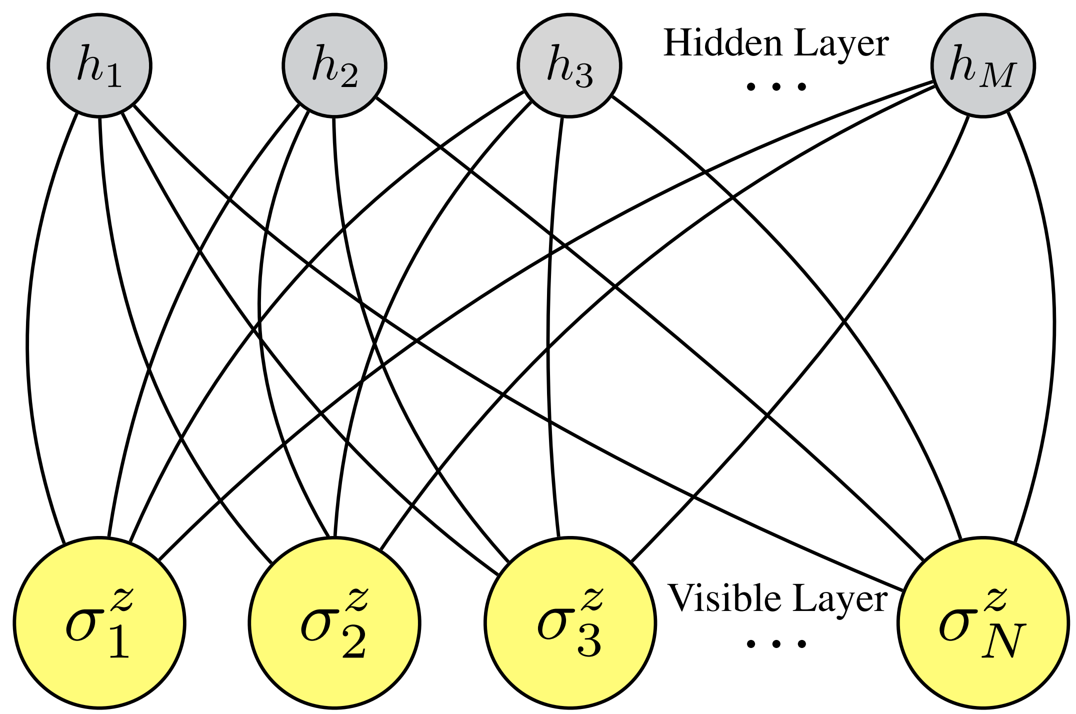
Extensions

Wetzel, 1703.02435
Hu, Singh, Scalettar, 1704.00080
Wetzel, Scherzer, 1705.05582
Wang and Zhai, 1706.07977

Variational Ansatz

RBM as a variational ansatz

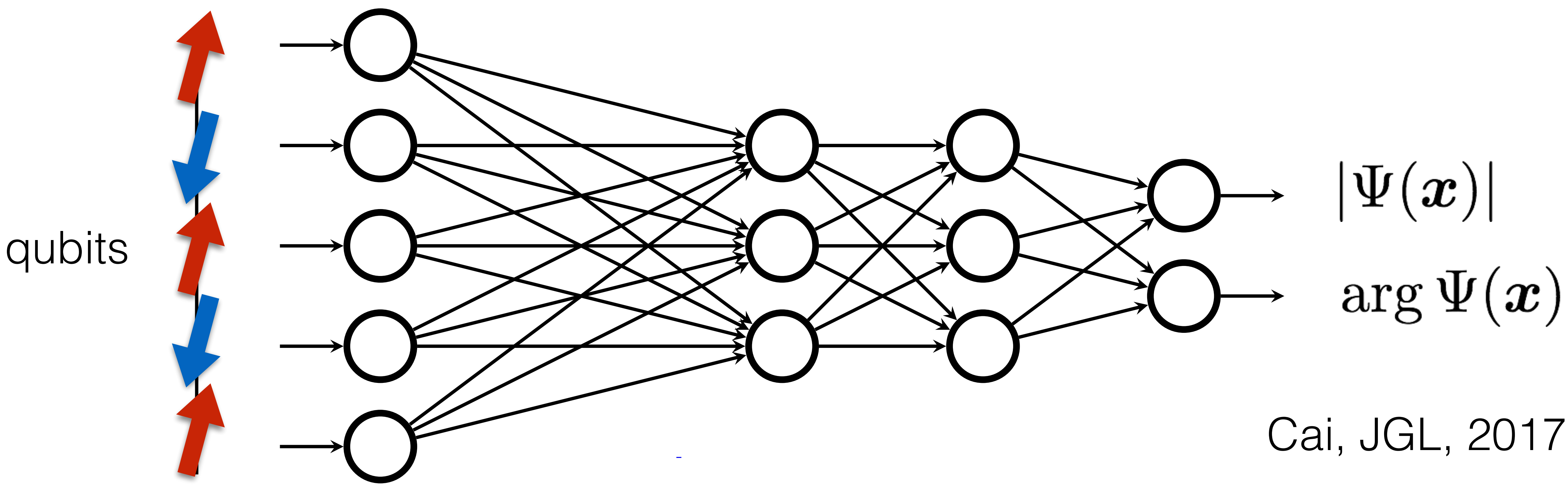
Carleo and Troyer, 1606.02318



- Exact construction for 1d SPT, 2d toric code state etc
- Related to tensor network, string-bond, correlator product states
- Killer app ? Long-range, volume law entanglement, chiral state, improved Jastrow

Deng, Li, Gao, Chen, Cheng, Xiang, LW, Clark, Glasser, Carl Budich, Imada…

Deep neural net as a variational ansatz



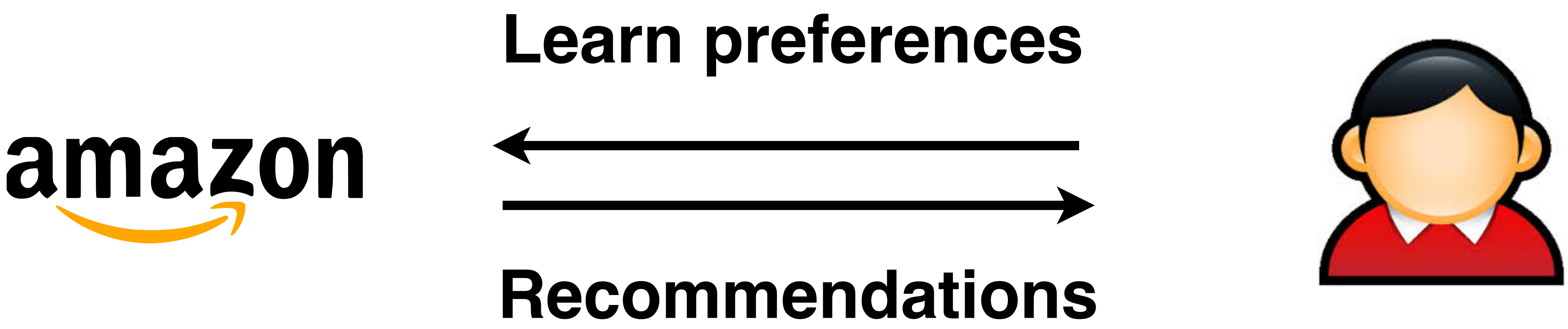
- Train the deep neural net ansatz using Backprop
- Feature discovery and abstraction power of the deep hierarchical structure
- Bottleneck appears to be the stochastic optimization (VMC)

Monte Carlo Update Proposals

A Video from Google DeepMind

http://www.nature.com/nature/journal/v518/n7540/fig_tab/nature14236_SV2.html

Proposals from Boltzmann Machine

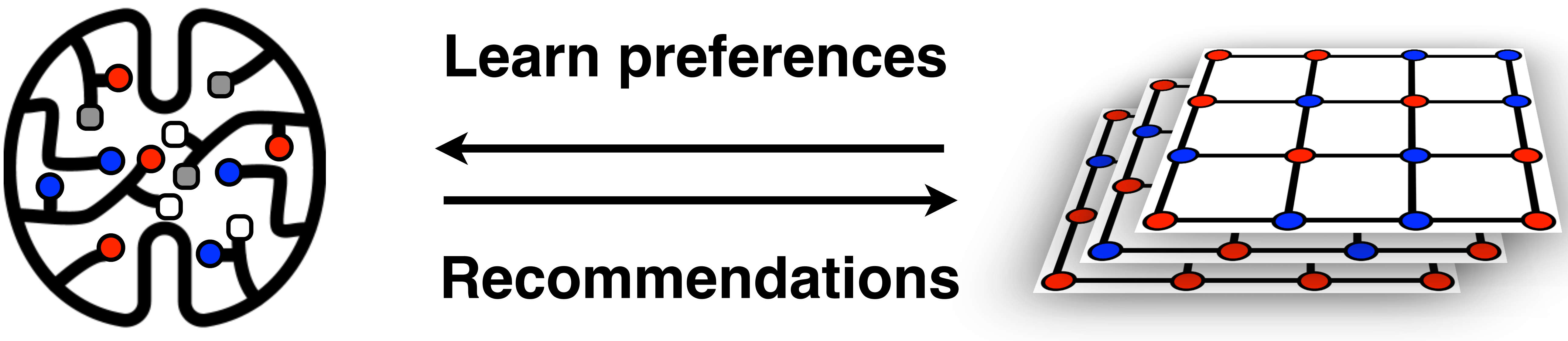


- Use Boltzmann Machines as recommender systems for Monte Carlo simulation of physical systems

Li Huang and LW, 1610.02746

Liu, Qi, Meng, Fu, 1610.03137

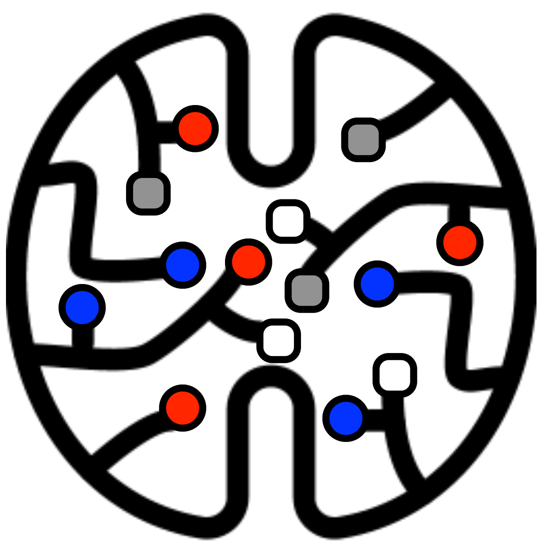
Proposals from Boltzmann Machine



- Use Boltzmann Machines as recommender systems for Monte Carlo simulation of physical systems

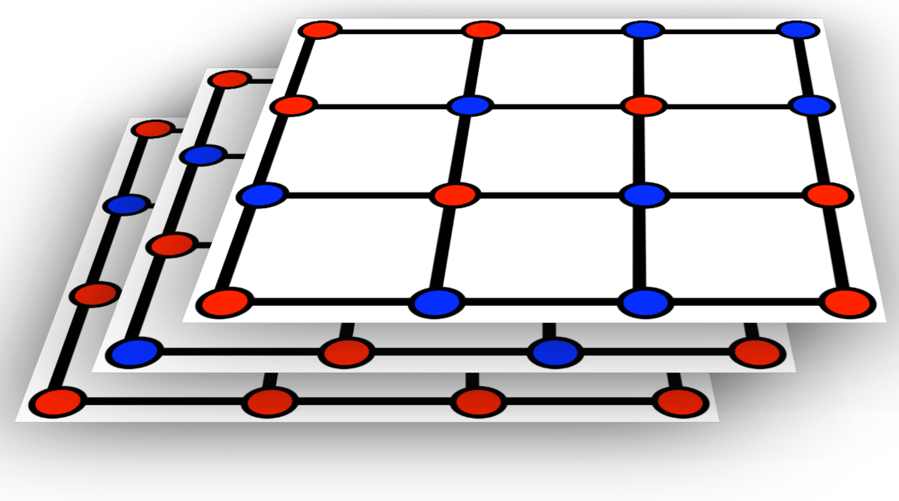
Li Huang and LW, 1610.02746
Liu, Qi, Meng, Fu, 1610.03137

Proposals from Boltzmann Machine



Learn preferences

Recommendations



- Use Boltzmann Machines as recommender systems for Monte Carlo simulation of physical systems

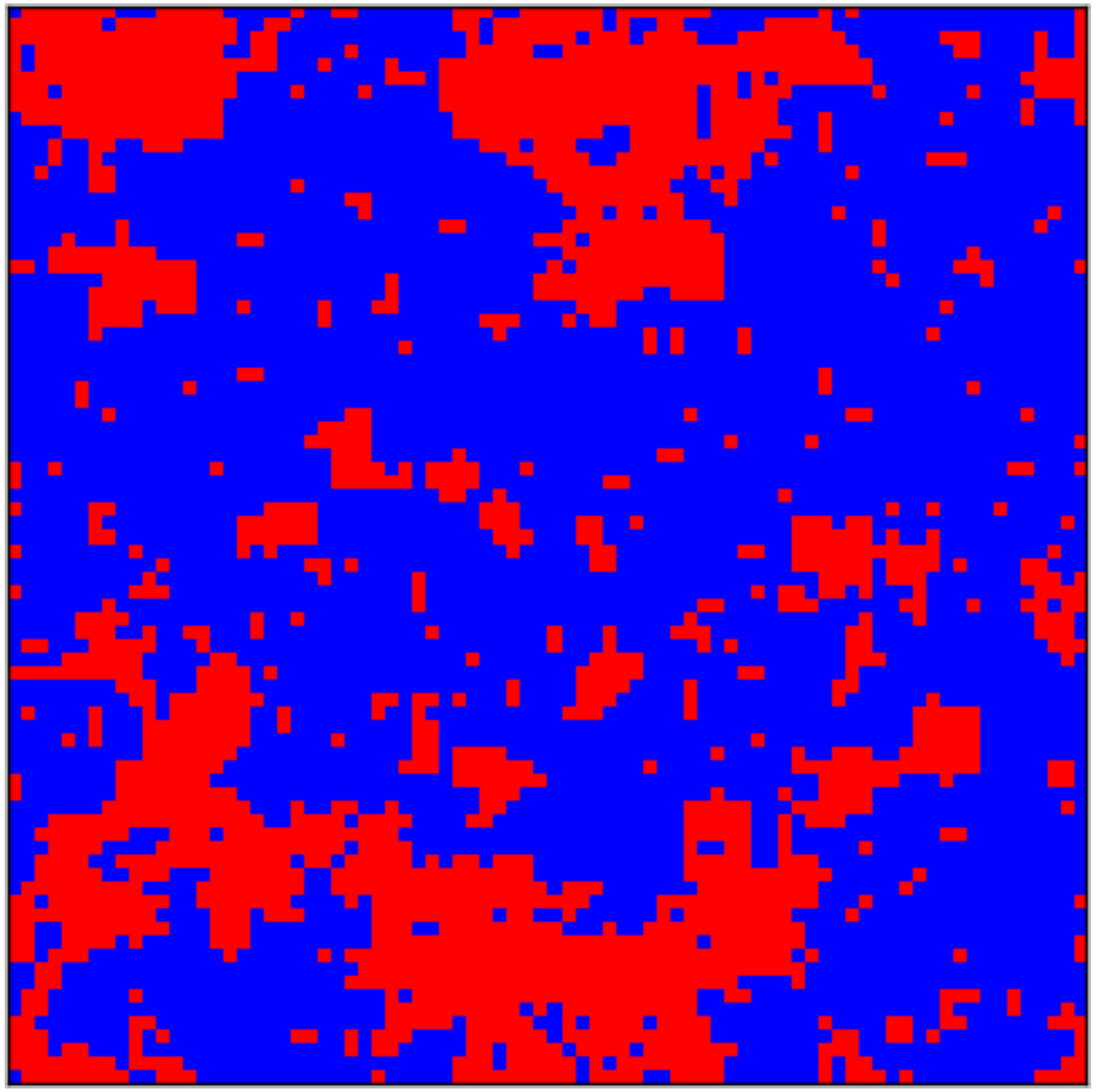
Li Huang and LW, 1610.02746

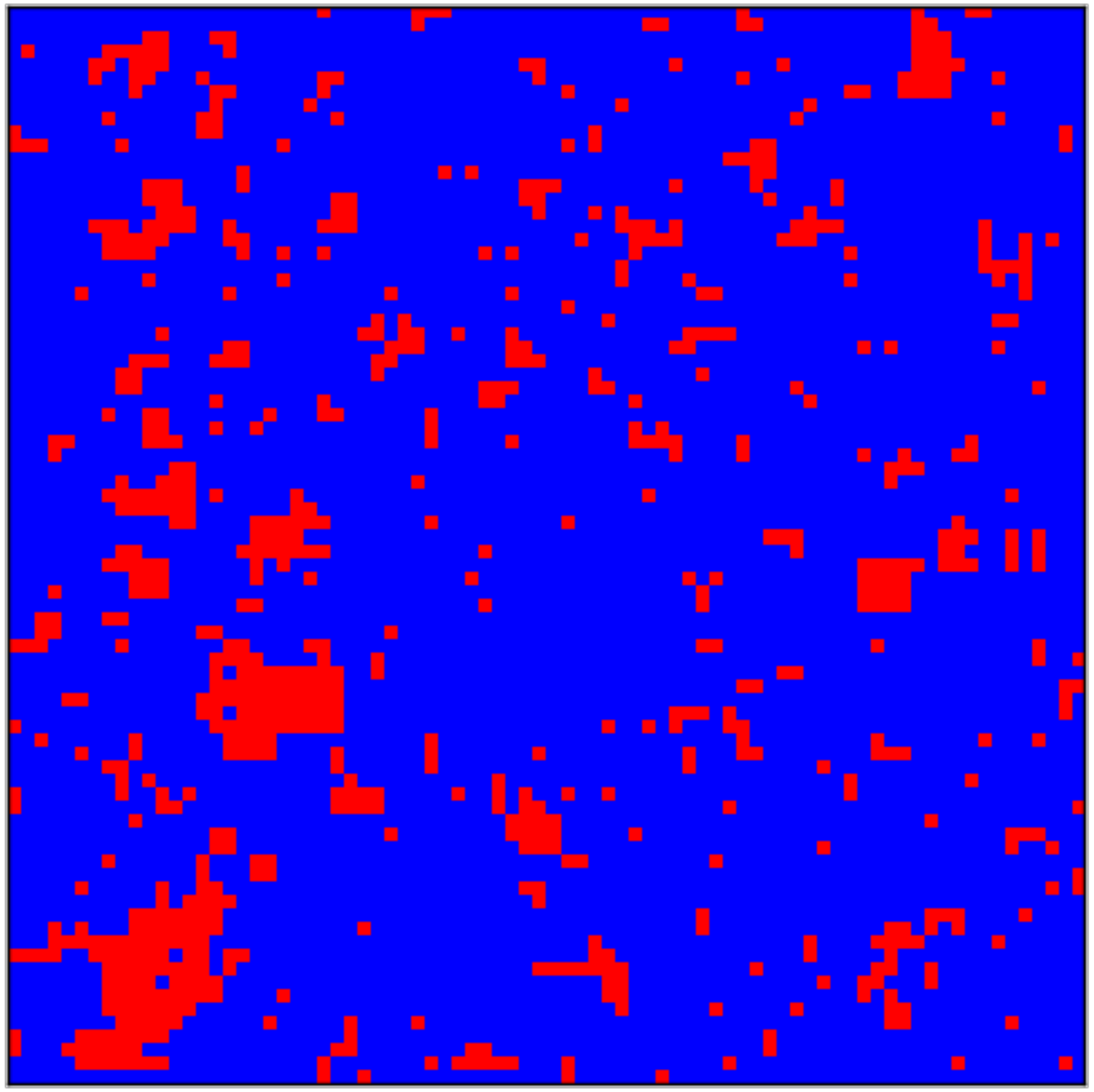
Liu, Qi, Meng, Fu, 1610.03137

- Moreover, BM parametrizes Monte Carlo policies and explores novel algorithms!

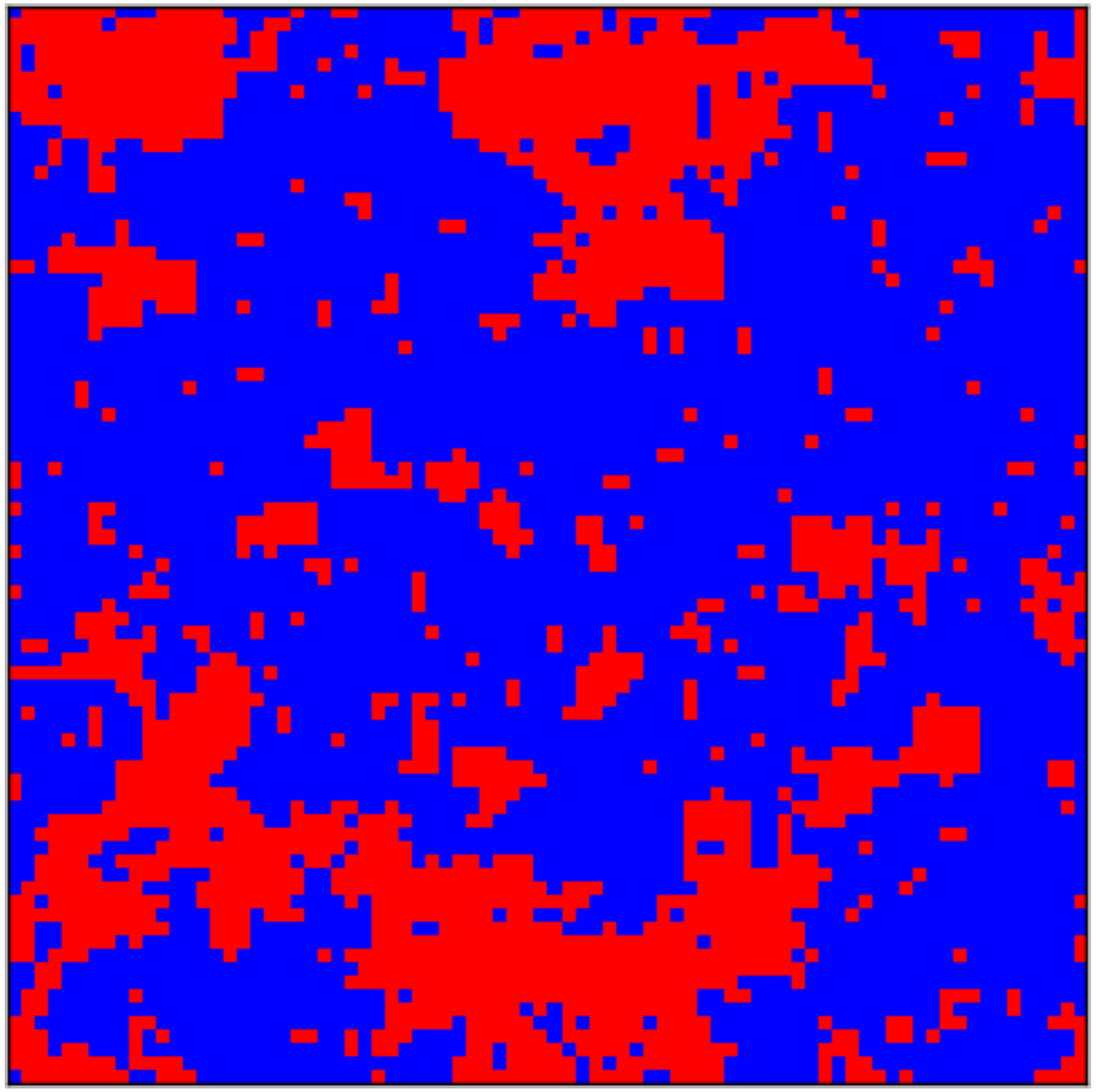
LW, 1702.08586

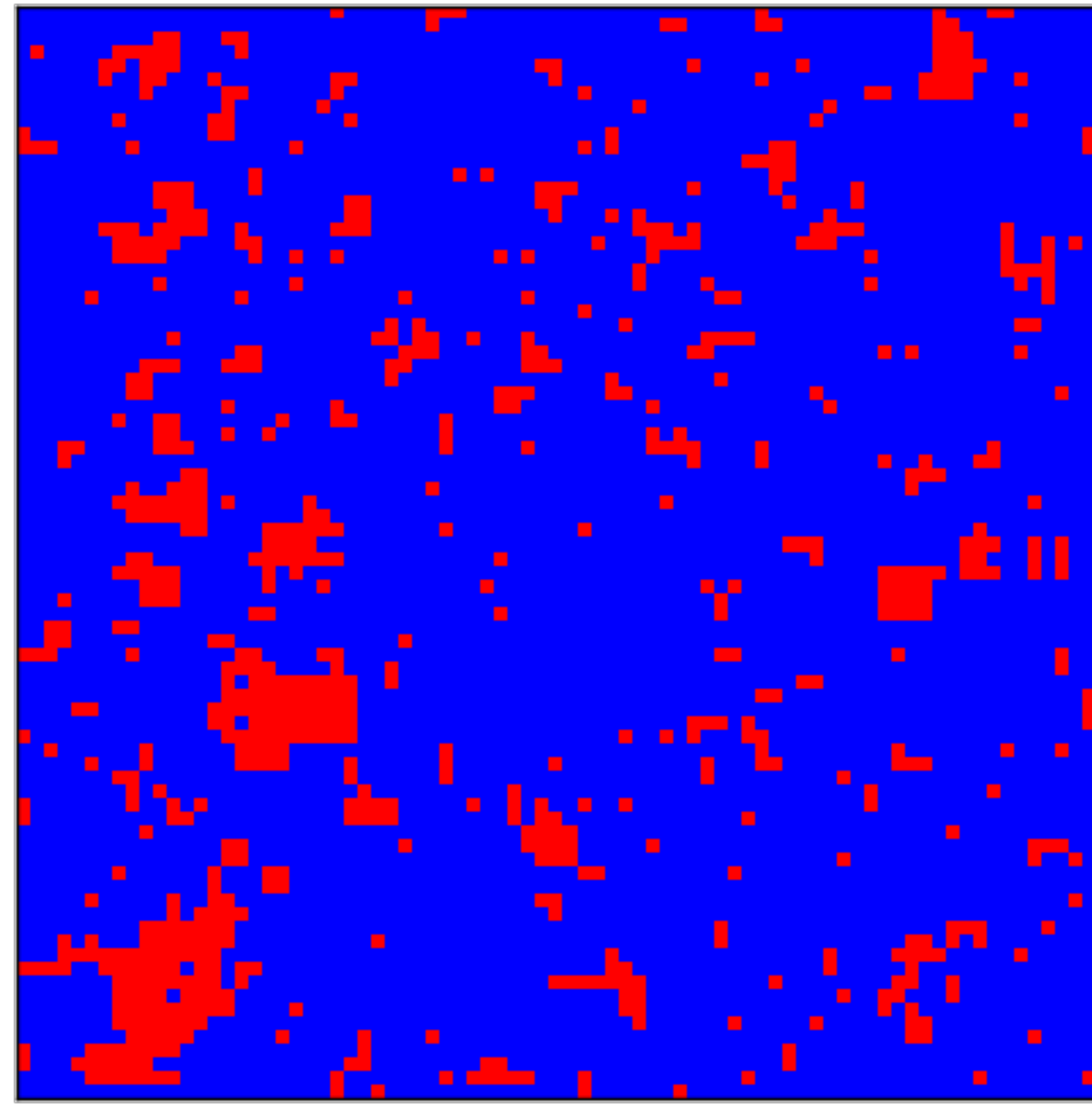
Local vs Cluster algorithms





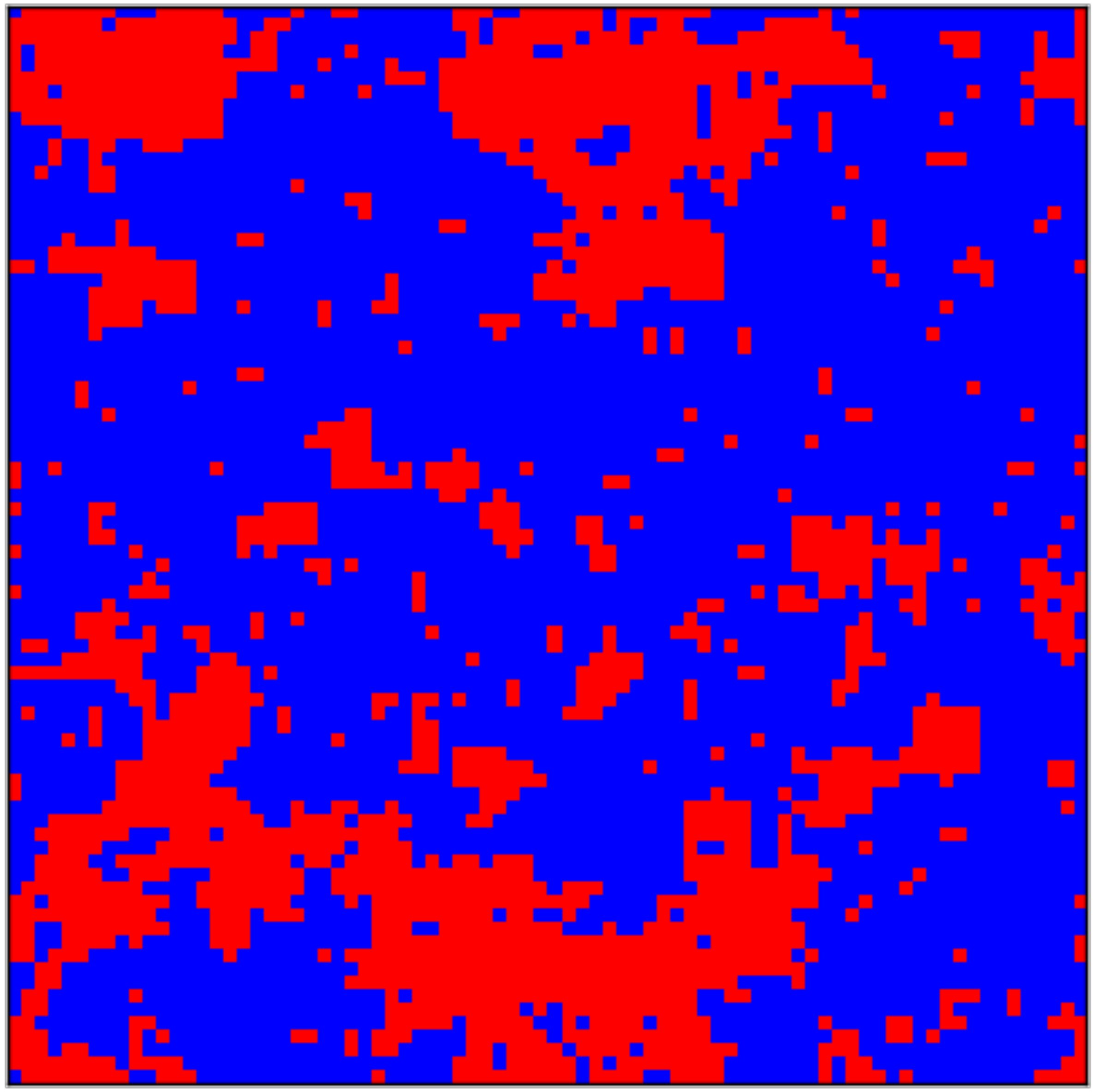
Local vs Cluster algorithms

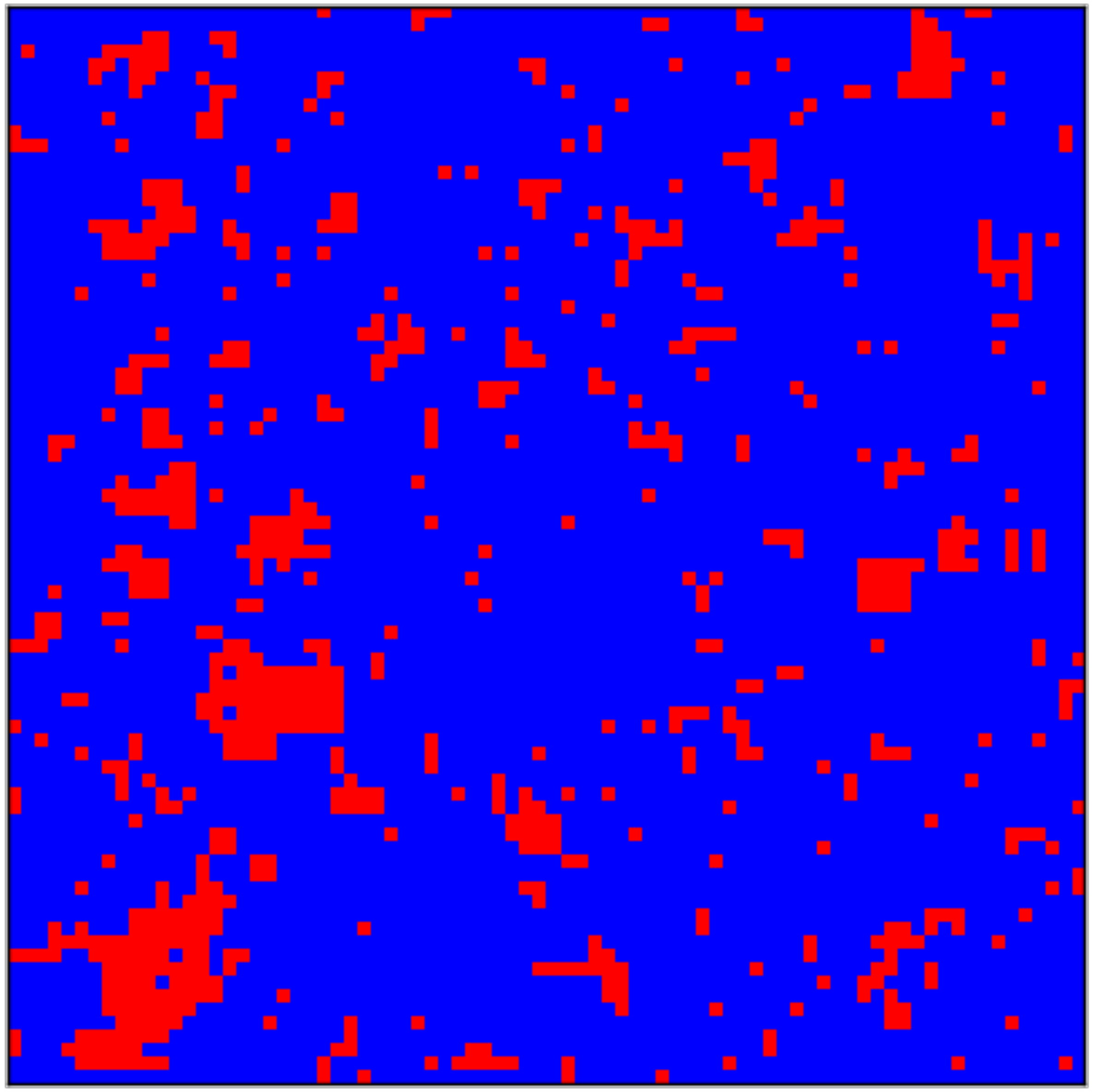




is slower than

Local vs Cluster algorithms





Algorithmic innovation outperforms Moore's law!

Deep learning the MC proposal

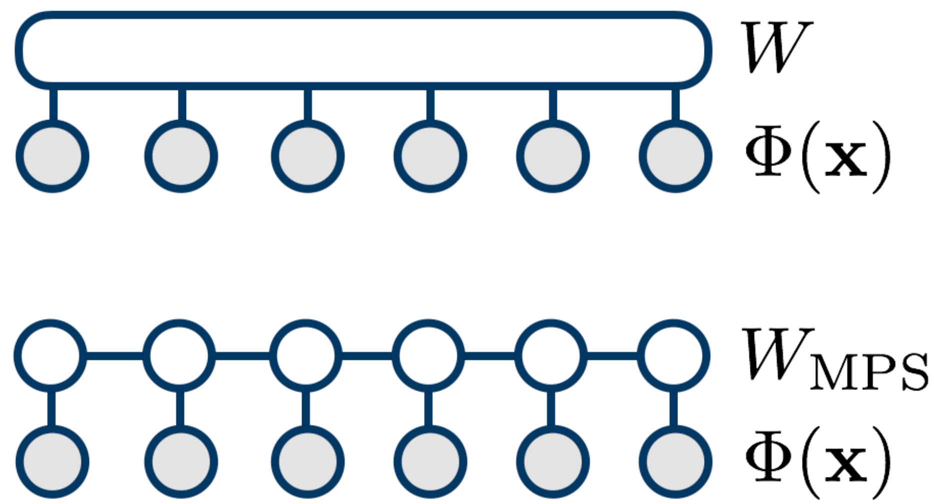
$$A(\boldsymbol{x} \to \boldsymbol{x}') = \min\left[1, \underbrace{\boxed{\frac{q(\boldsymbol{x}' \to \boldsymbol{x})}{q(\boldsymbol{x} \to \boldsymbol{x}')}}}_{\text{Policy}} \cdot \underbrace{\boxed{\frac{\pi(\boldsymbol{x}')}{\pi(\boldsymbol{x})}}}_{\text{Fixed by the Physics}}\right]$$

- A-NICE-MC 1706.07561
- Generalize hybrid MC using neural networks 1711.09268
- Probabilistic programs as proposals 1801.03612

Tensor Networks

MPS for pattern recognition

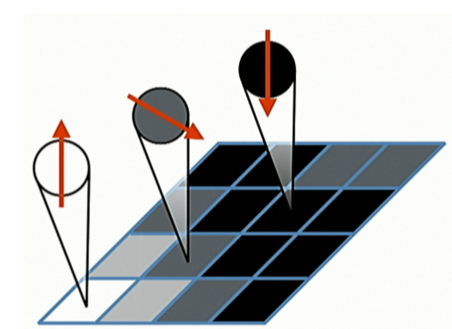
$f(\mathbf{x}) =$



$\approx$

$$f(\mathbf{x}) = \sum_{\{s\}} W_{s_1 s_2 \cdots s_N} \phi^{s_1}(x_1) \otimes \phi^{s_2}(x_2) \otimes \cdots \phi^{s_N}(x_N)$$

$$\phi^{s_j}(x_j) = \left[\cos\left(\frac{\pi}{2}x_j\right),\ \sin\left(\frac{\pi}{2}x_j\right)\right]$$



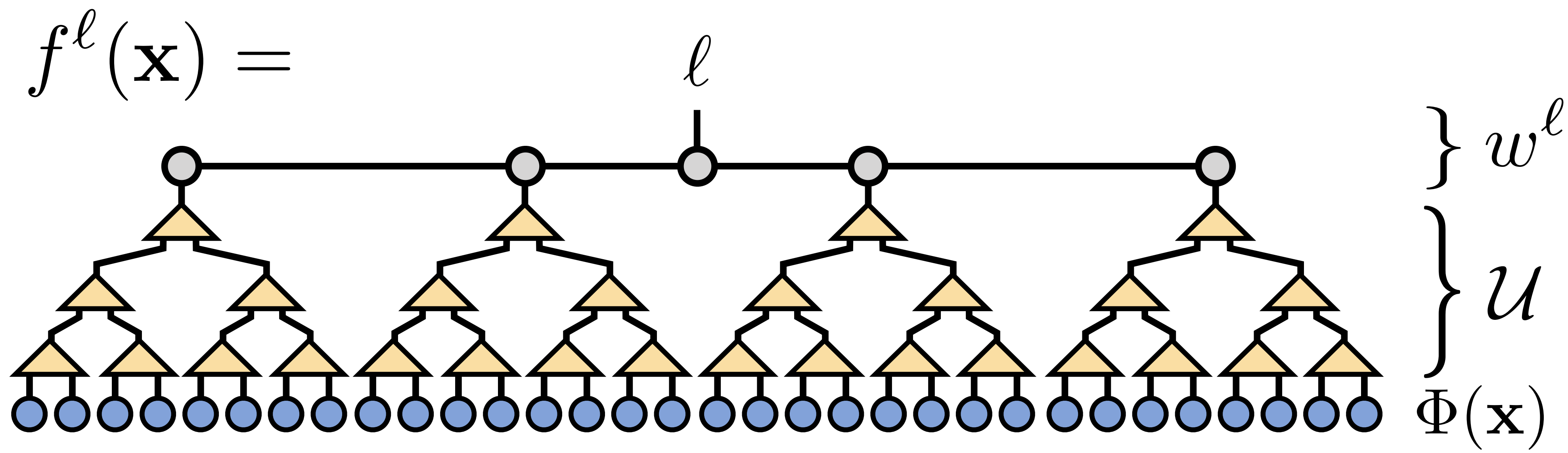
%99.03 accuracy on MNIST dataset*

* bond dimension 120
images scaled to 14*14

Stoudenmire and Schwab, 1605.05775

cf Liu et al, 1710.04833
Novikov et al,1605.03795

Unsupervised petraining with a TTN

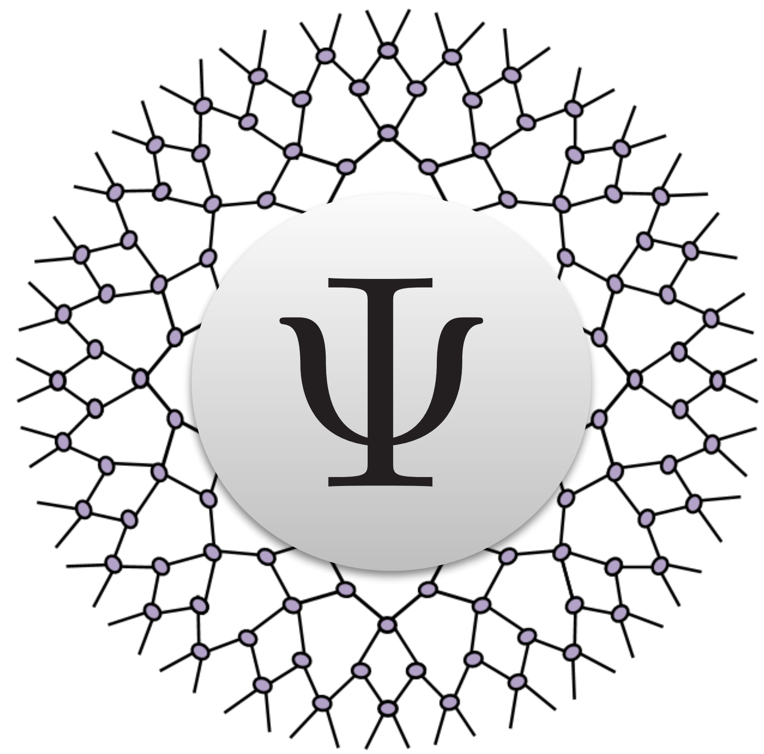


%88.97 accuracy on fashion MNIST

Comparable accuracy to ConvNets

Stoudenmire, 1801.00315

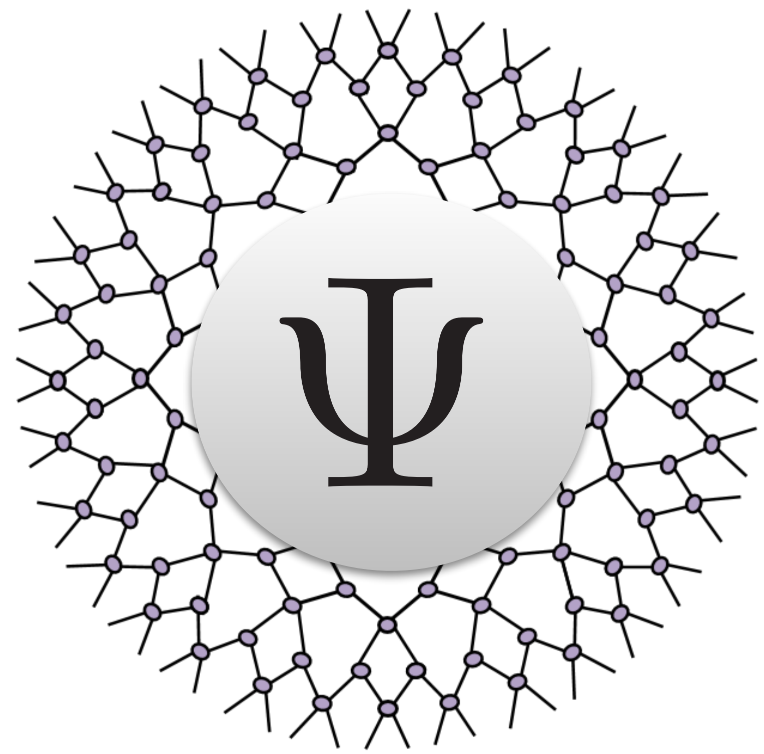
Quantum inspired generative modeling

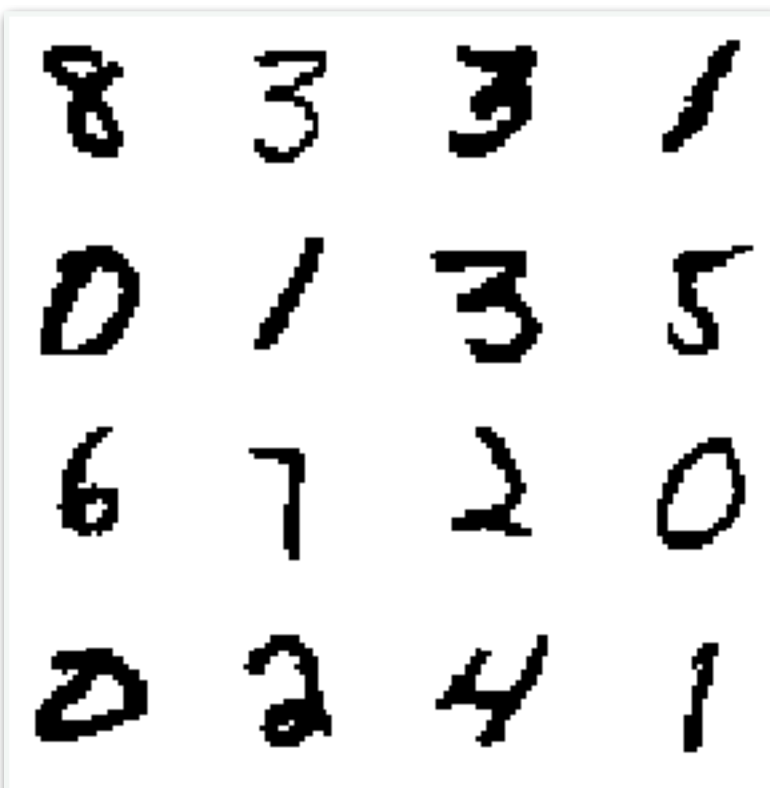


Han, Wang, Fan, LW, Zhang, 1709.01662

Quantum inspired generative modeling

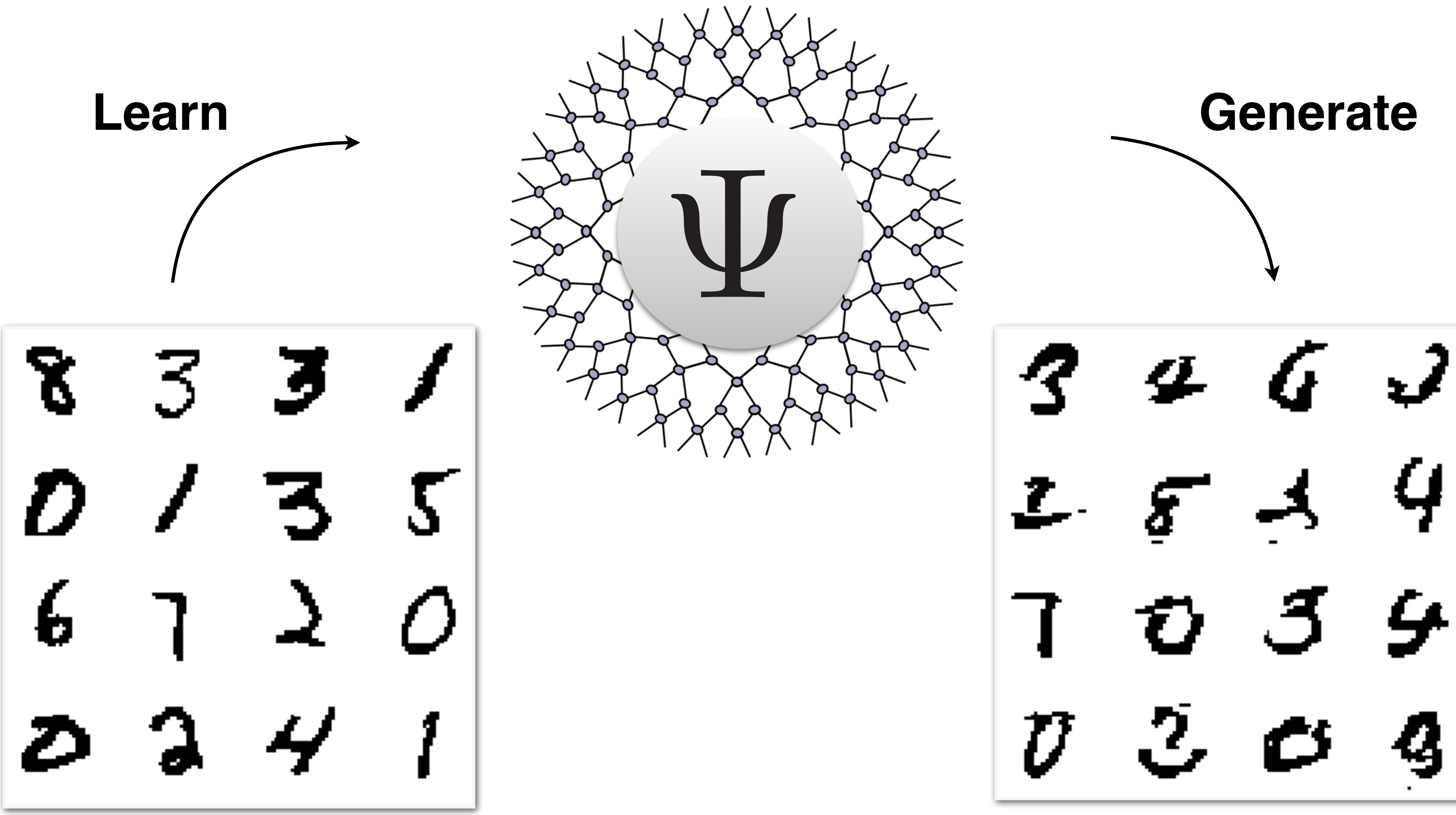
Learn





Han, Wang, Fan, LW, Zhang, 1709.01662

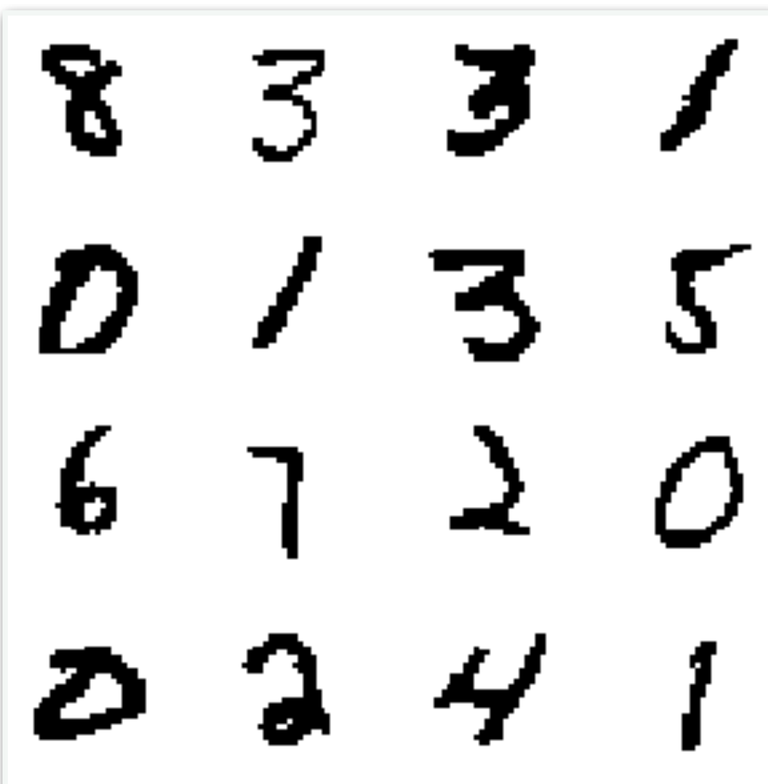
Quantum inspired generative modeling

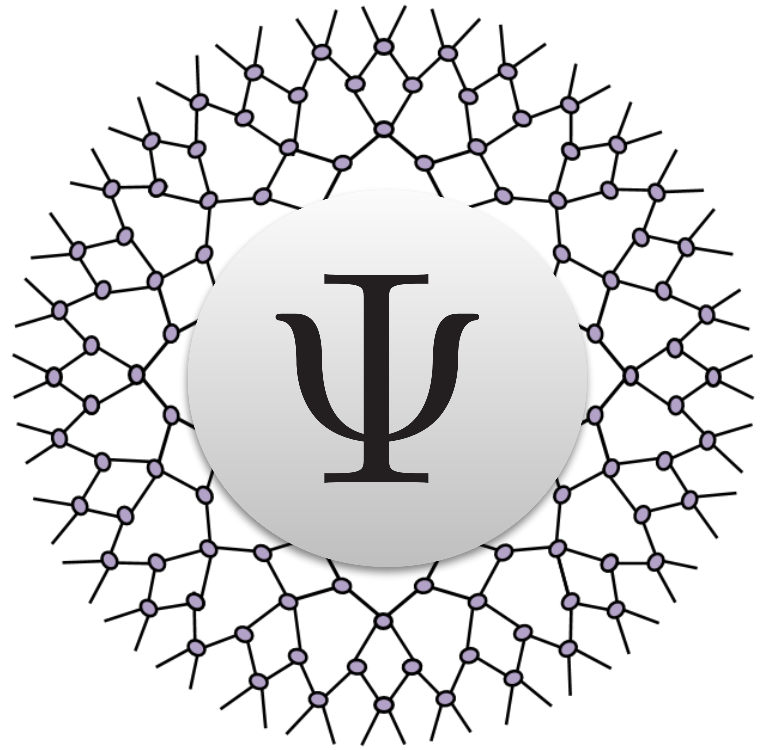


Han, Wang, Fan, LW, Zhang, 1709.01662

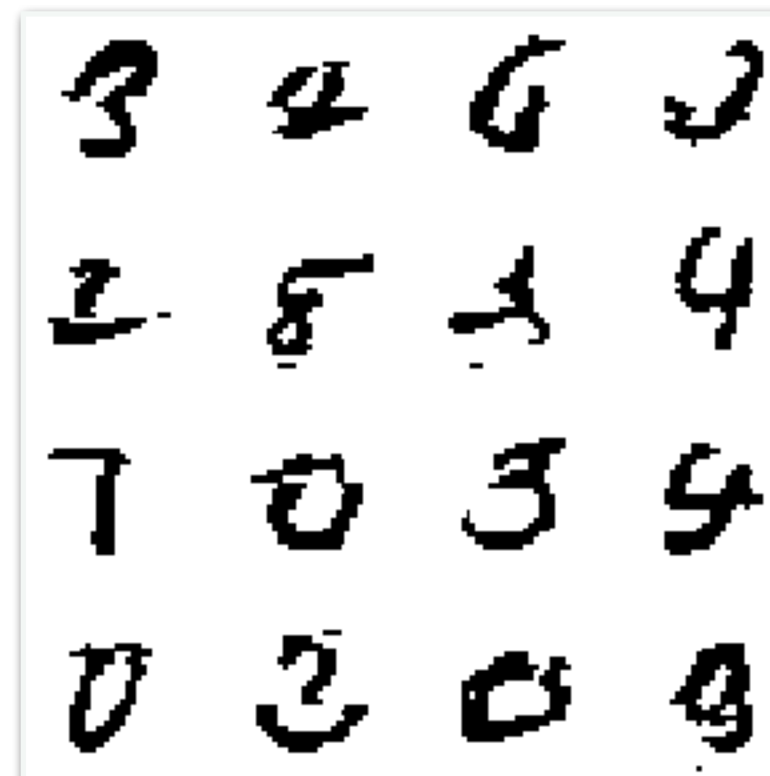
Quantum inspired generative modeling

Learn





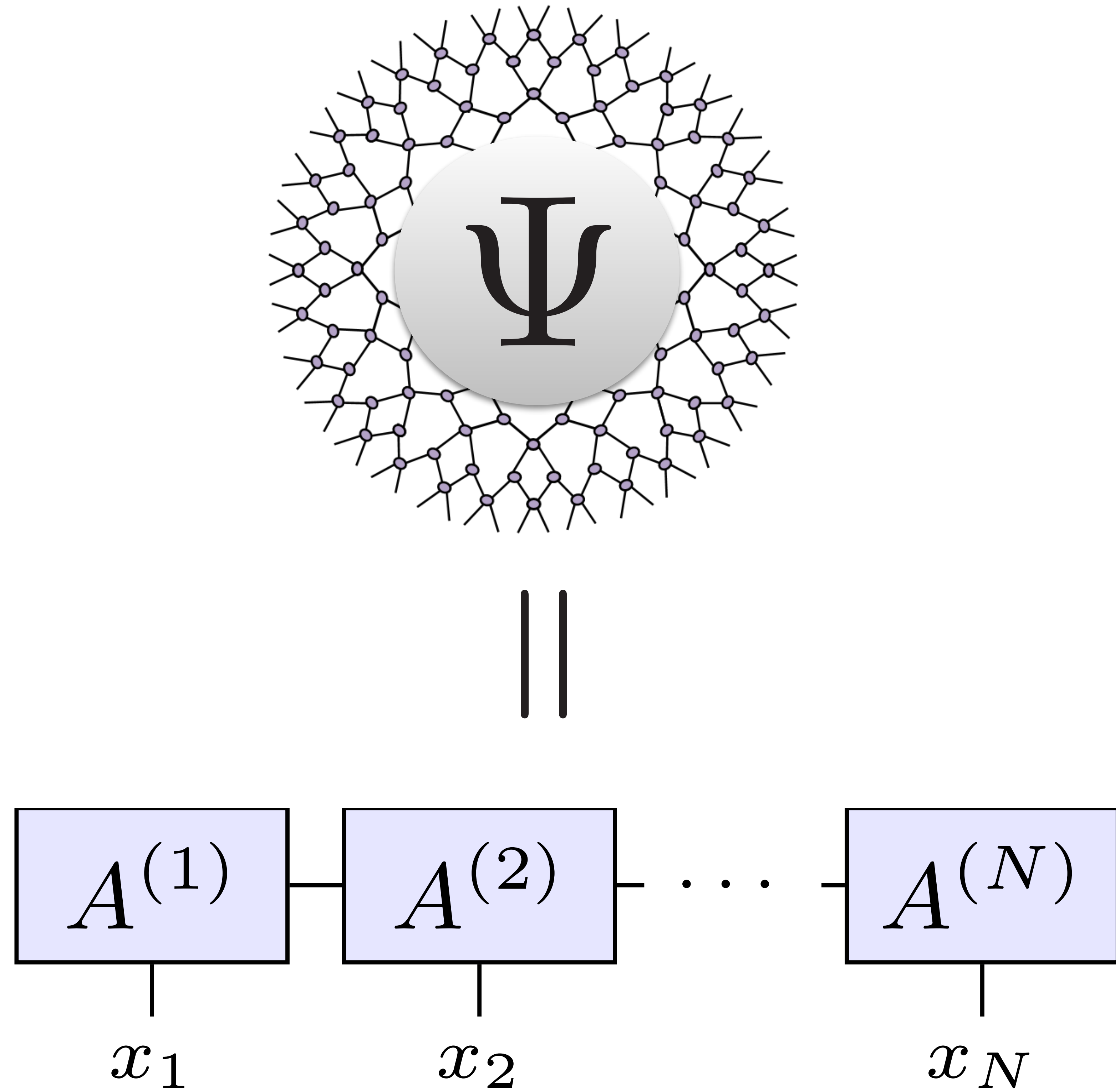
Generate



"Teach a quantum state to write digits"

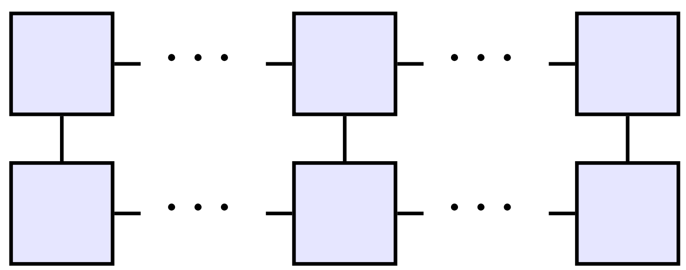
Han, Wang, Fan, LW, Zhang, 1709.01662

Generative modeling using Matrix Product States

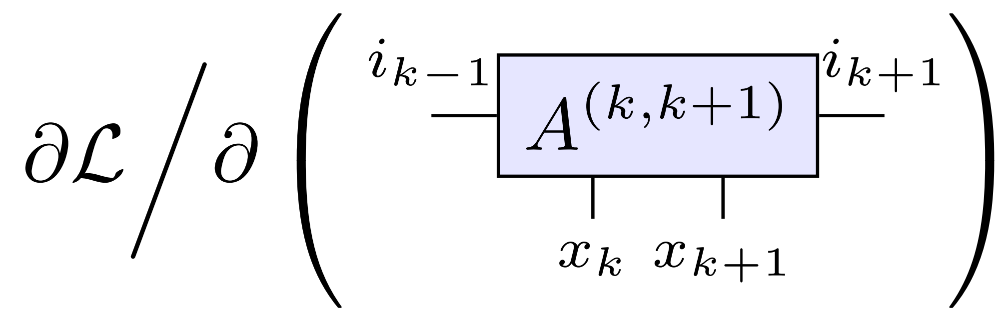


Connections to Boltzmann Machine, Chen, Cheng, Xie, LW, Xiang 1701.04831

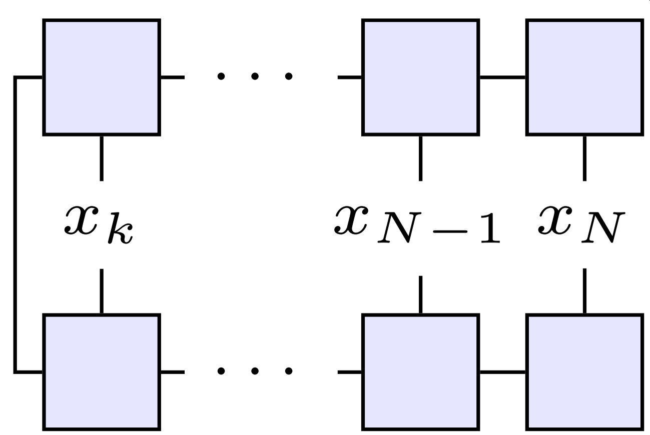
Generative modeling using Matrix Product States

$\mathcal{N} =$ 

Tractable partition function via MPS contraction



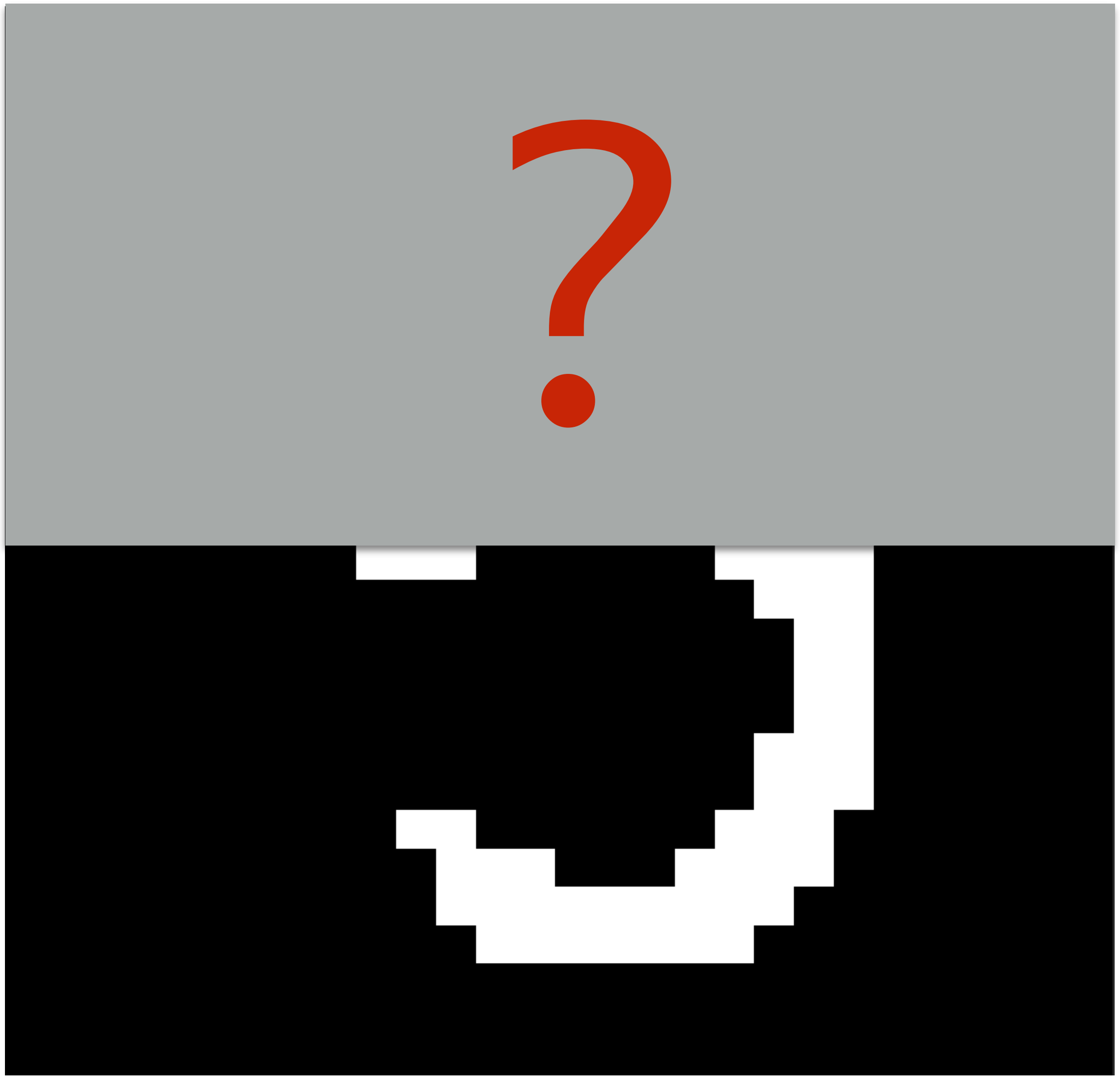
Efficient and adaptive learning via DMRG



Direct sampling using “Zipper”

Quantum Perspective on Deep Learning

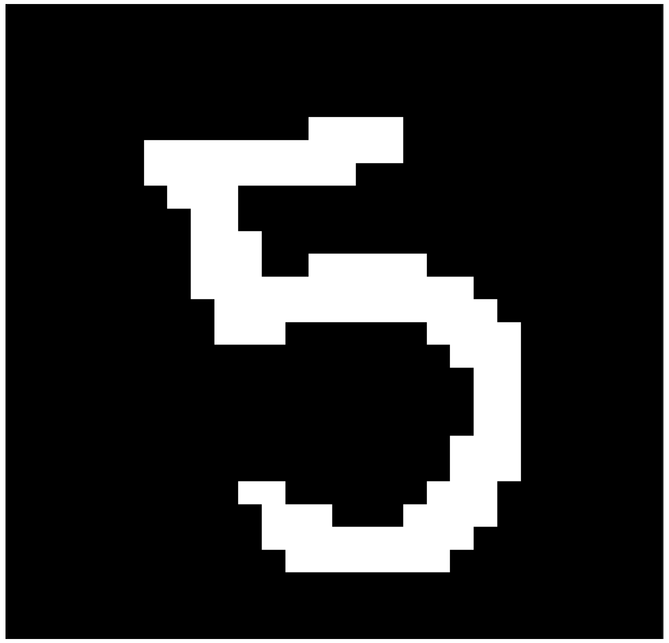
Quantum Perspective on Deep Learning

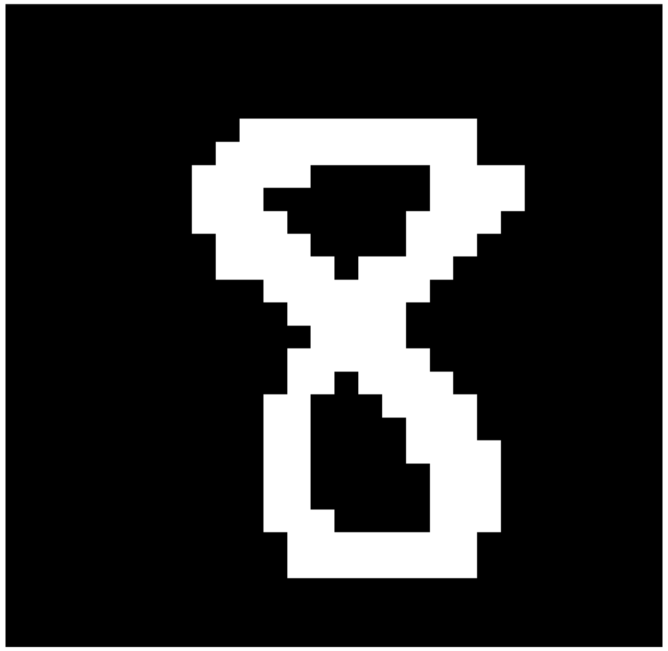


Quantum Perspective on Deep Learning

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

A: Information pattern of the probability functions

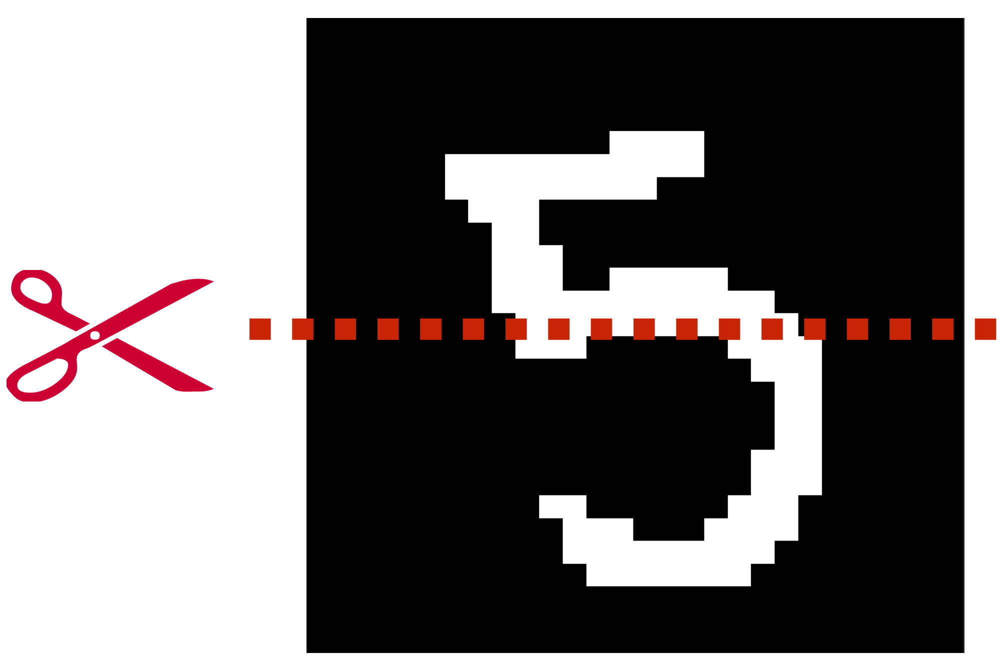


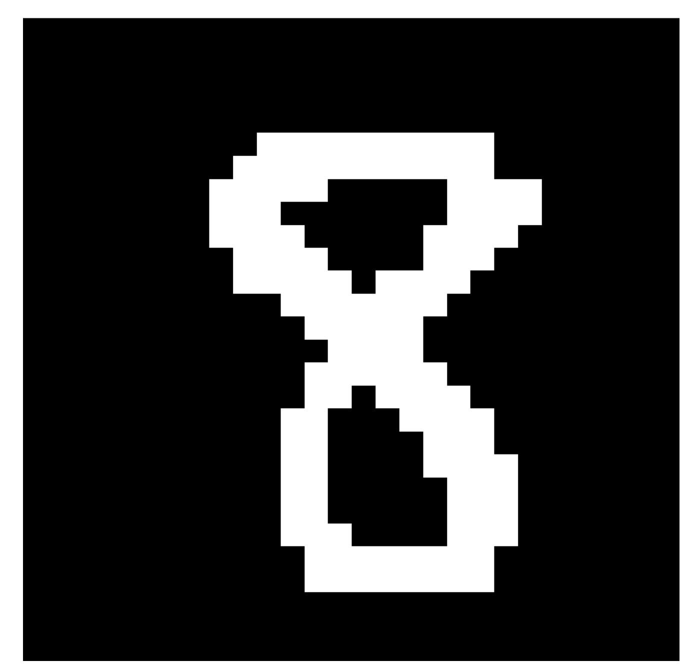


Quantum Perspective on Deep Learning

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

A: Information pattern of the probability functions

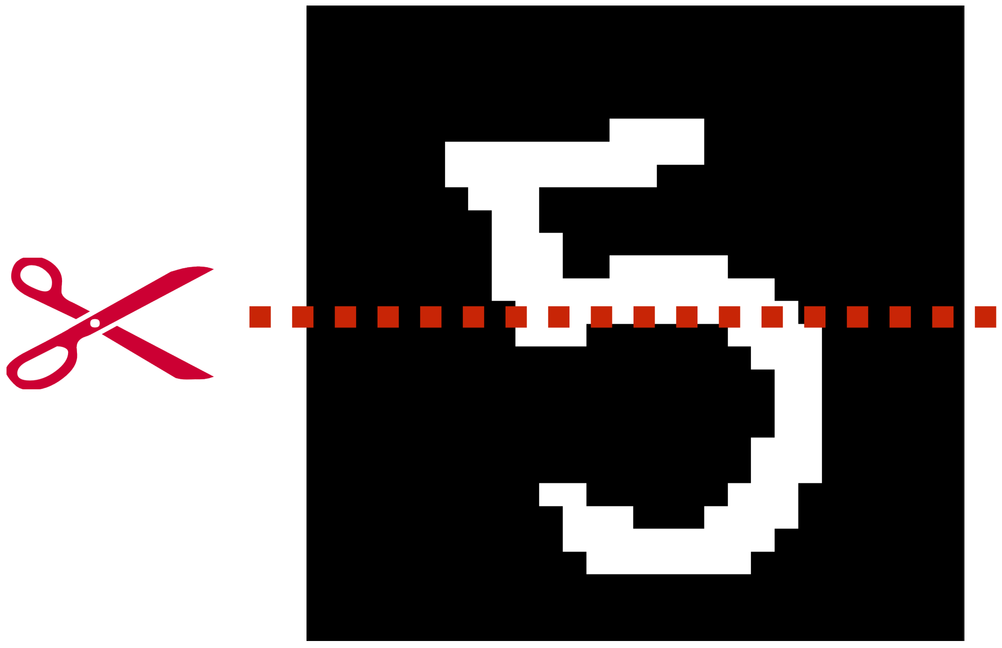


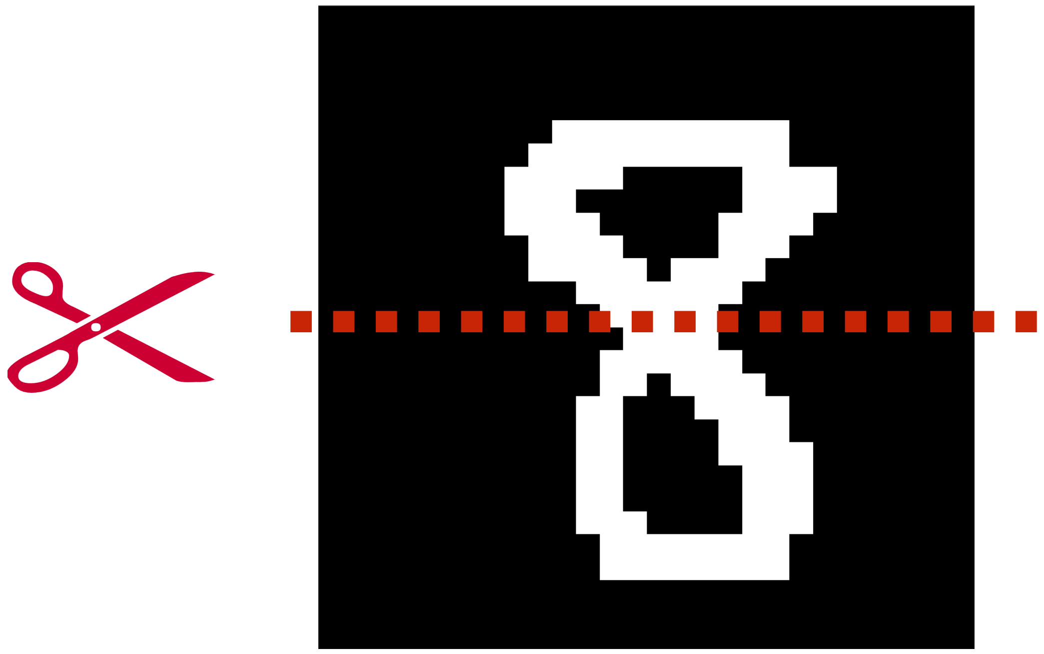


Quantum Perspective on Deep Learning

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

A: Information pattern of the probability functions

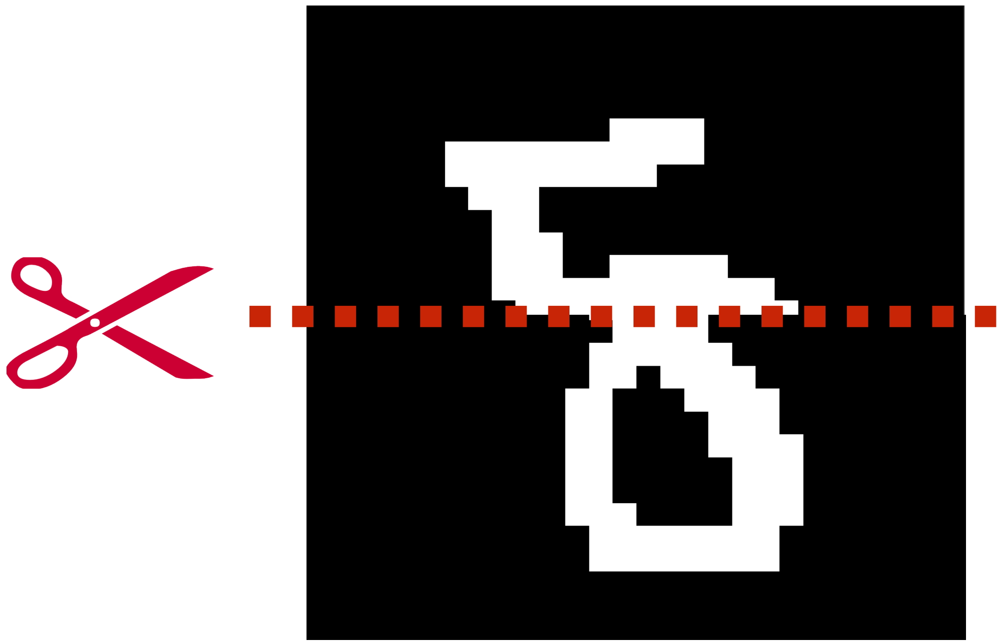


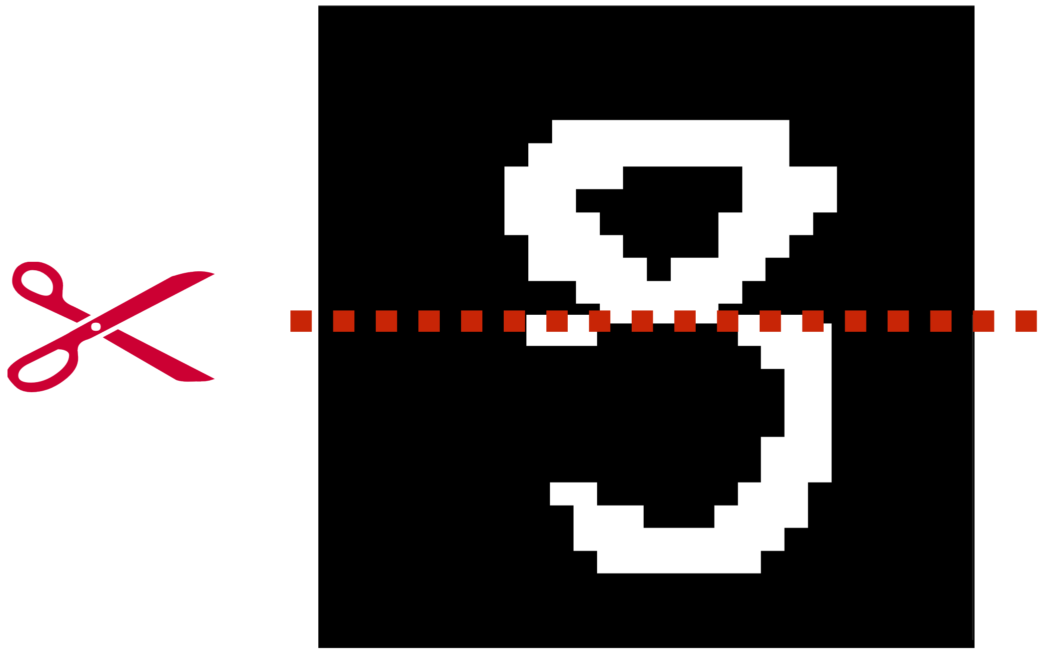


Quantum Perspective on Deep Learning

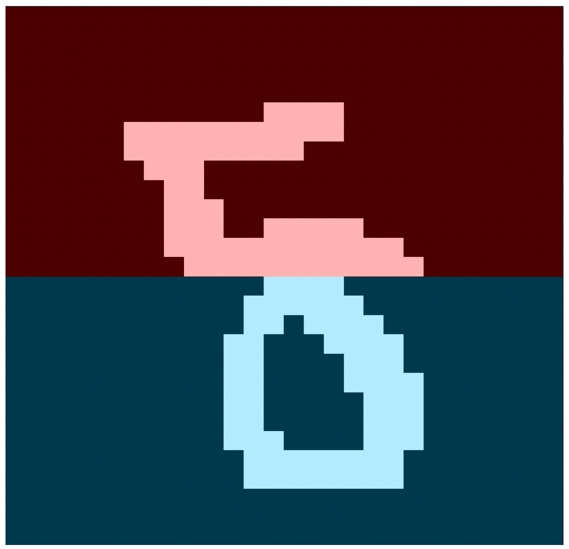
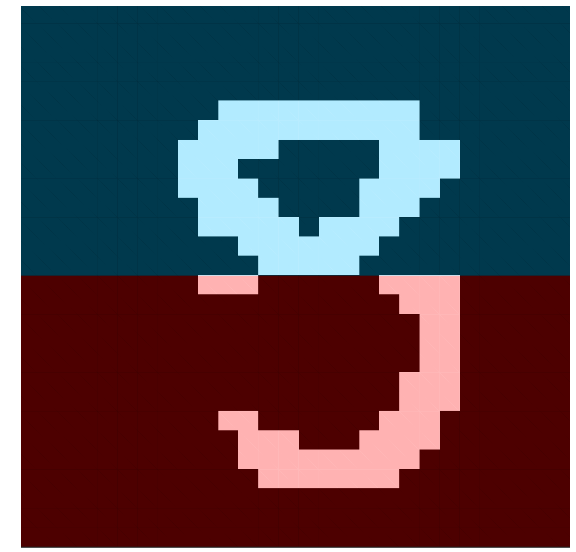
Q: How to quantify our prior knowledge on the data distribution ?

A: Information pattern of the probability functions

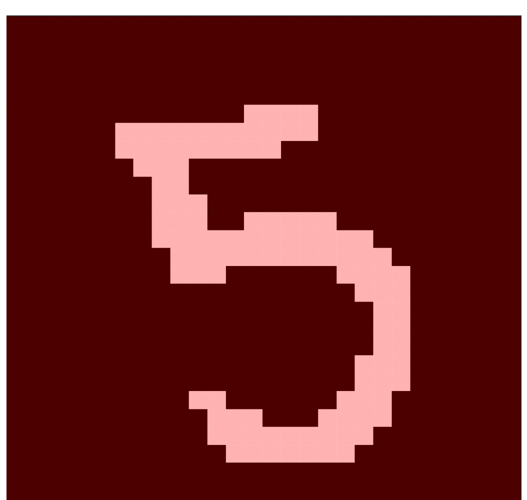
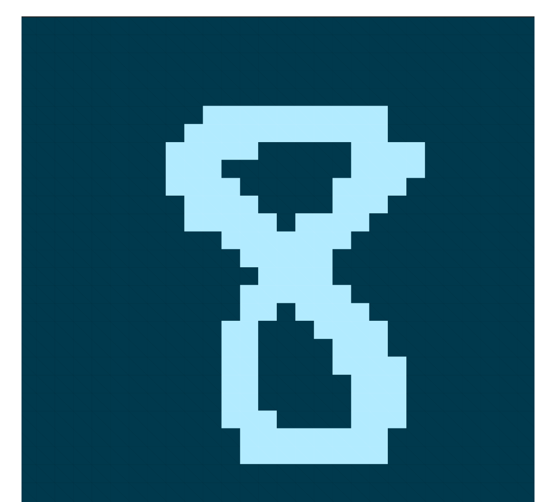




Quantum Perspective on Deep Learning

$p\Big($  $\Big) \times p\Big($  $\Big)$

divided by

$p\Big($  $\Big) \times p\Big($  $\Big)$

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.

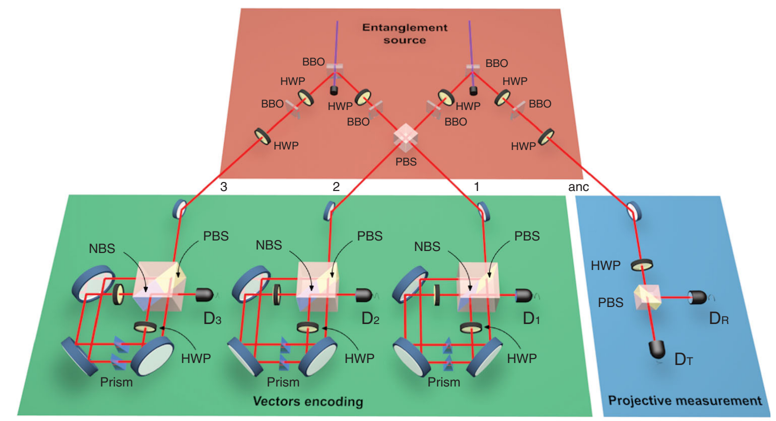
Deep Learning and Quantum Entanglement: Fundamental Connections with Implications to Network Design

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
The Hebrew University of Jerusalem

Quantum Machine Leanring

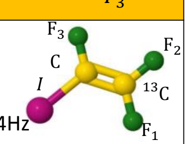
Quantum Machine Learning

- Search
- Sampling
- Clustering
- Optimization
- Linear system solver
- Support vector machines
- Principal component analysis



Cai et al, PRL **114**, 110504 (2015)

	^{13}C	F_1	F_2	F_3
^{13}C	15479.9Hz			
F_1	-297.7Hz	-33130.1Hz		
F_2	-275.7Hz	64.6Hz	-42681.4Hz	
F_3	39.1Hz	51.5Hz	-129.0Hz	-56443.5Hz
T_2^*	1.22s	0.66s	0.63s	0.61s
T_2	7.9s	4.4s	6.8s	4.8s



Li et al, PRL **114**, 140504 (2015)

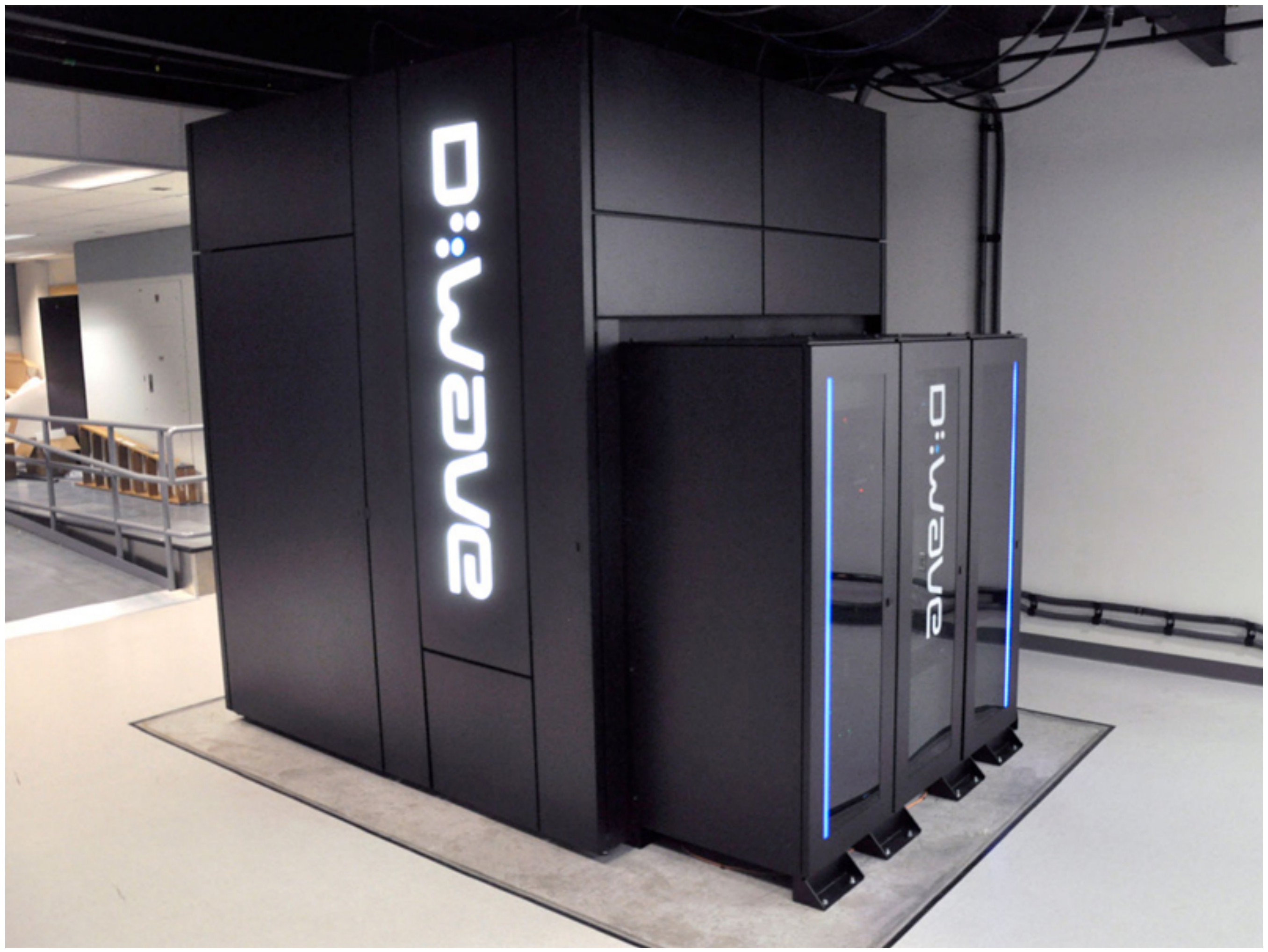
few qubits demo

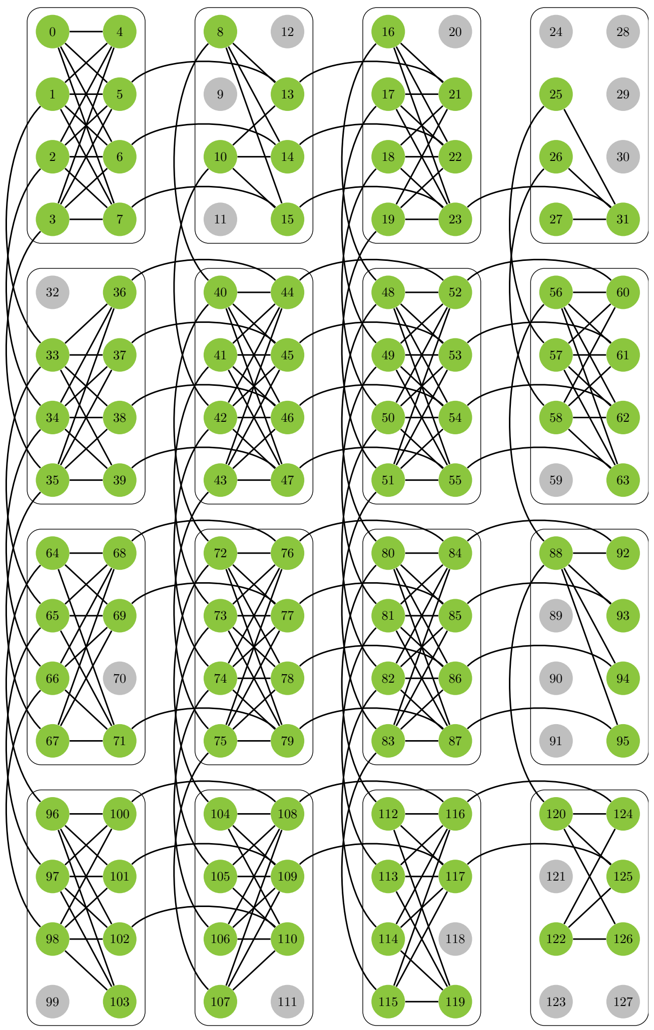
"Use a quantum computer to speed up ML subroutines"

Review "Quantum machine learning", Biamonte et al, Nature 2017

Quantum Boltzmann Machines

$15 million "analog quantum device"





~2000 "qubits"

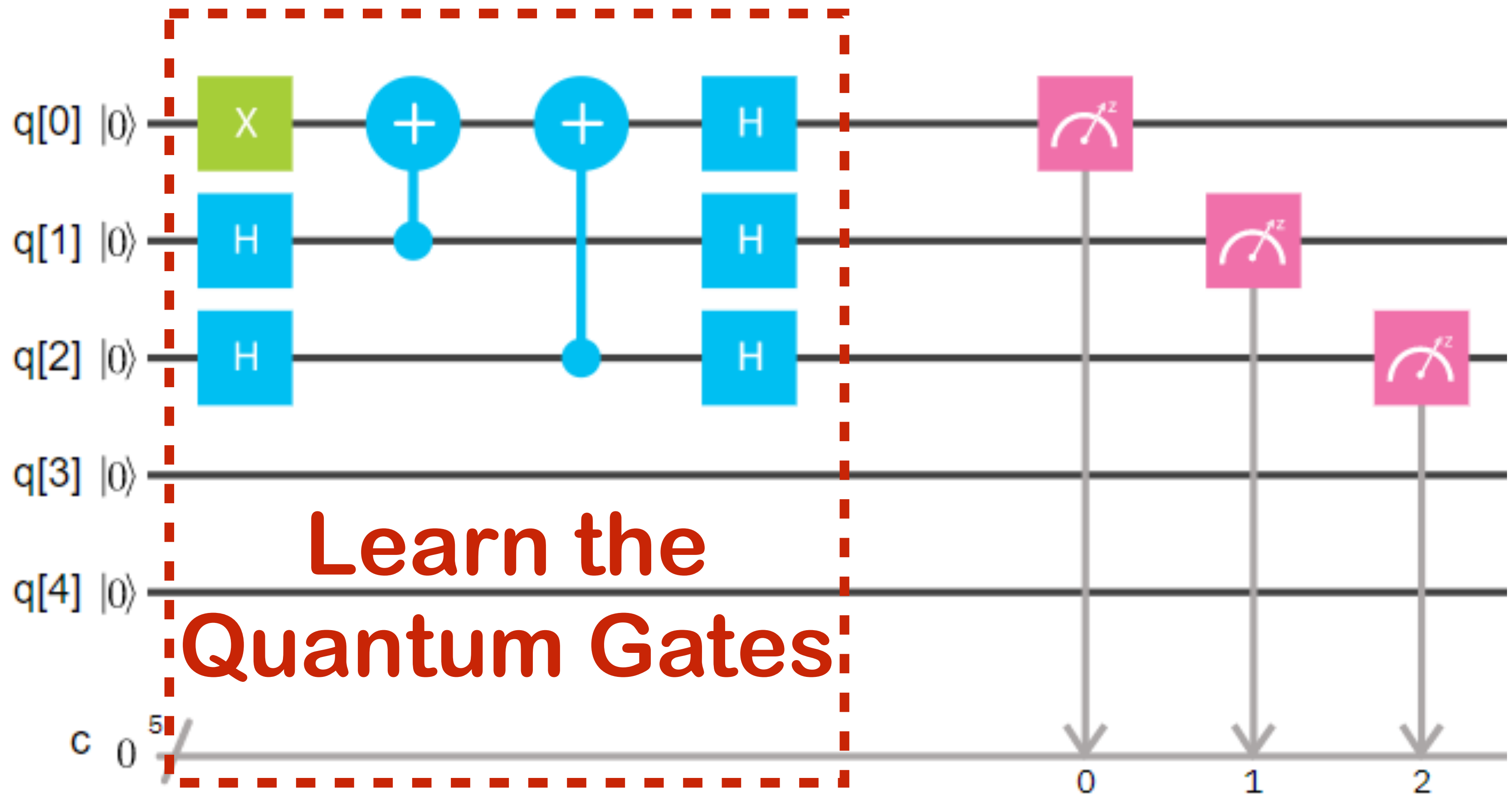
Is there any advantage of this quantum architecture?

Amin et al, 1601.02036   Perdomo-Ortiz et al, 1708.09757

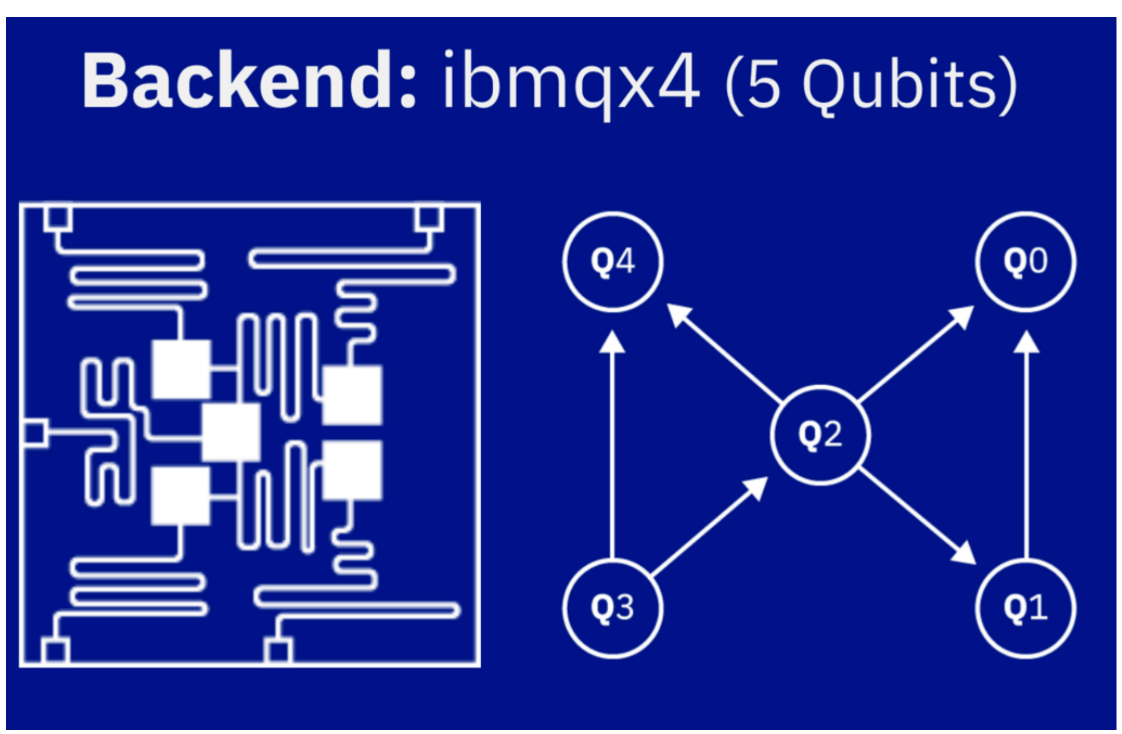
Quantum VAE 1802.05779

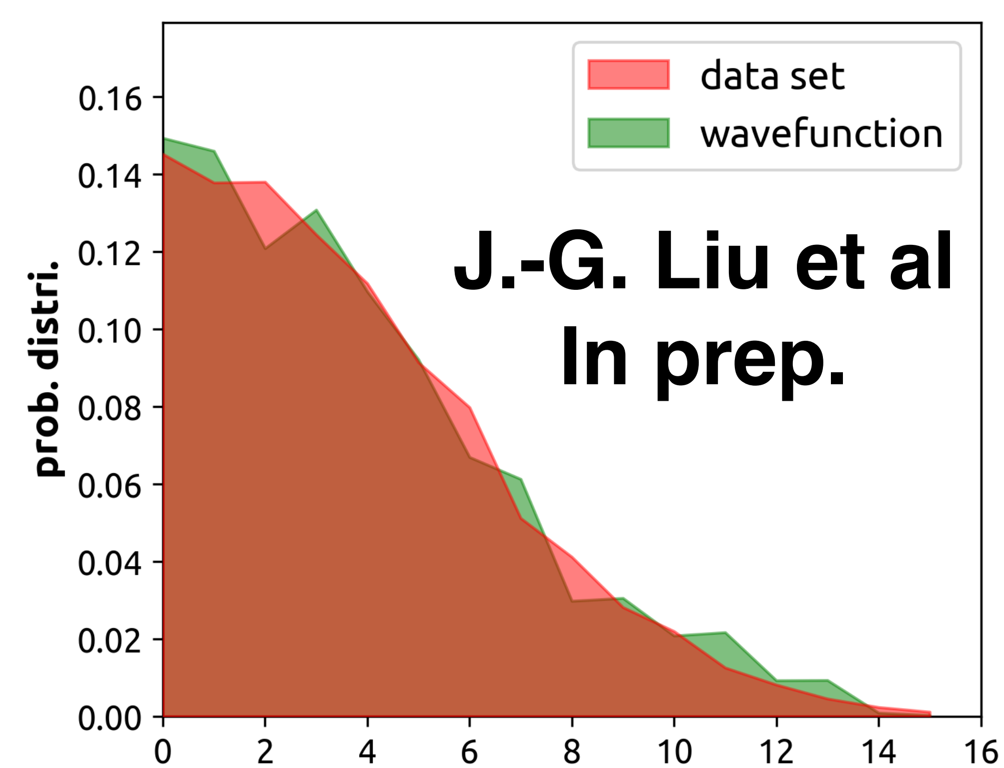
Born Machine in the Cloud

Initial State



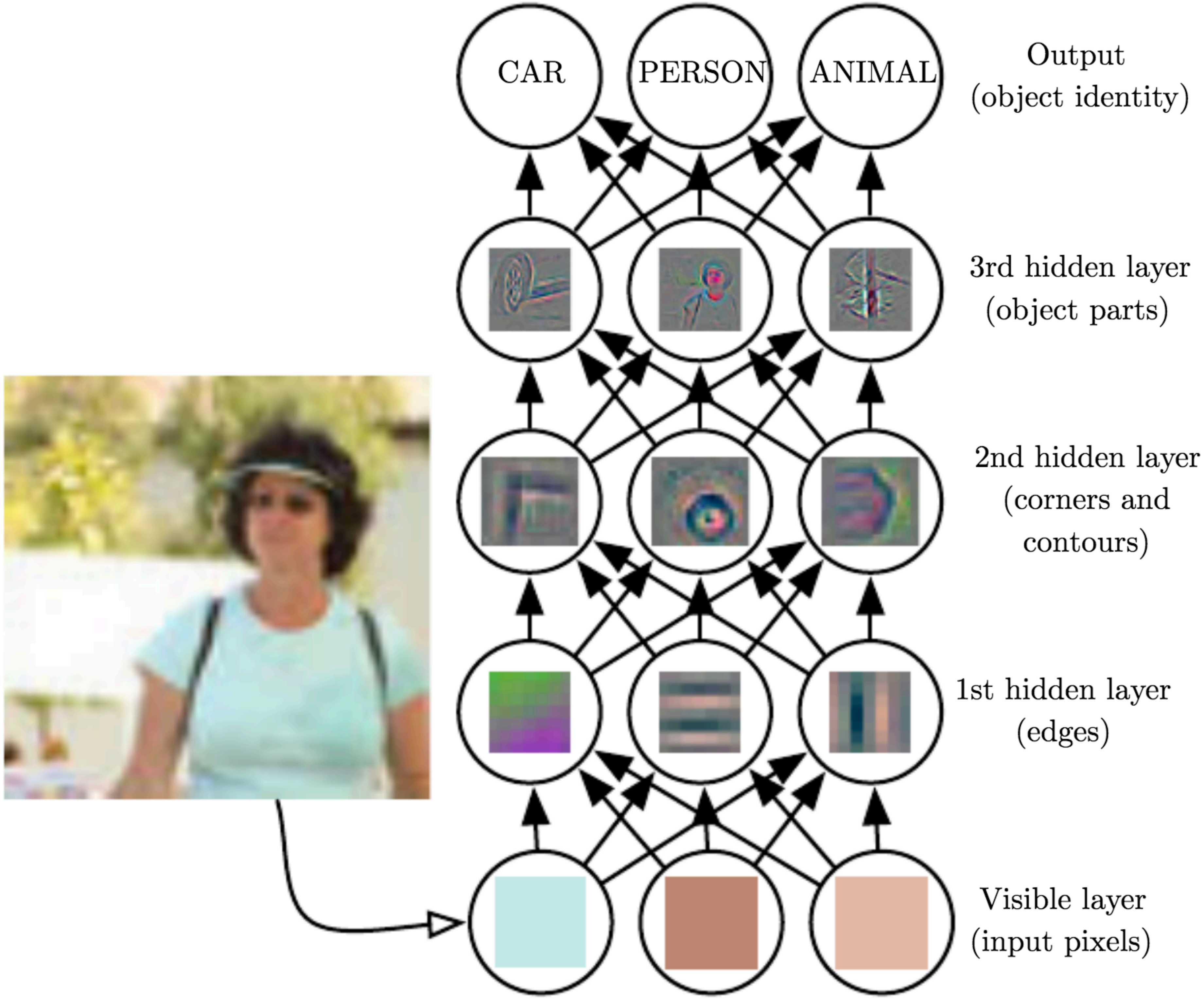
Collapse to a bit-string by measurement



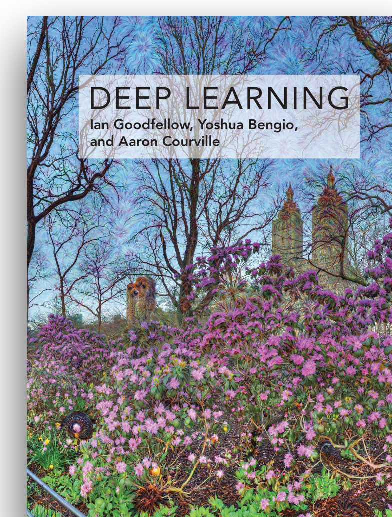


Renormalization Group

Deep Neural Network and RG



Page 6
Figure 1.2



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

OpenReview.net

Search ICLR 2013 conference

Login

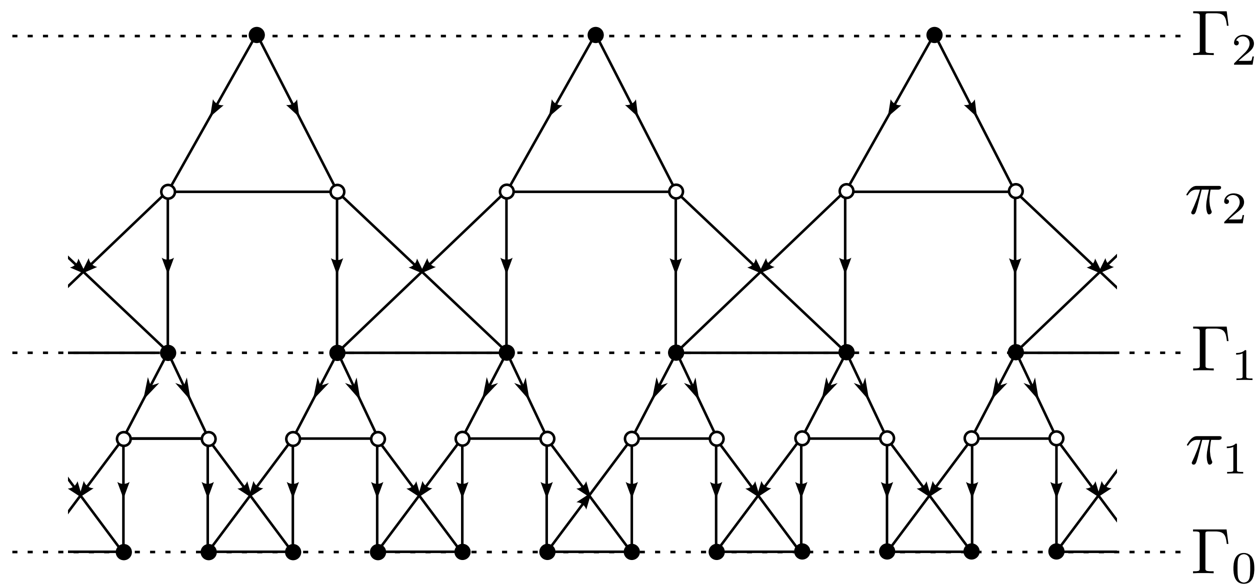
Deep learning and the renormalization group

Cédric Bény

15 Jan 2013 ICLR 2013 conference submission readers: everyone

Decision: reject

Abstract: Renormalization group methods, which analyze the way in which the effective behavior of a system depends on the scale at which it is observed, are key to modern condensed-matter theory and particle physics. The aim of this paper is to compare and contrast the ideas behind the renormalization group (RG) on the one hand and deep machine learning on the other, where depth and scale play a similar role. In order to illustrate this connection, we review a recent numerical method based on the RG---the multiscale entanglement renormalization ansatz (MERA)---and show how it can be converted into a learning algorithm based on a generative hierarchical Bayesian network model. Under the assumption---common in physics---that the distribution to be learned is fully characterized by local correlations, this algorithm involves only explicit evaluation of probabilities, hence doing away with sampling.



arxiv:1301.3124

Deep learning and the renormalization group

Cédric Bény

15 Jan 2013 ICLR 2013 conference submission readers: everyone

Decision: reject

Yann LeCun

05 Apr 2013 ICLR 2013 submission review readers: everyone

Review: It seems to me like there could be an interesting connection between approximate inference in graphical models and the renormalization methods.

There is in fact a long history of interactions between condensed matter physics and graphical models. For example, it is well known that the loopy belief propagation algorithm for inference minimizes the Bethe free energy (an approximation of the free energy in which only pairwise interactions are taken into account and high-order interactions are ignored). More generally, variational methods inspired by statistical physics have been a very popular topic in graphical model inference.

The renormalization methods could be relevant to deep architectures in the sense that the grouping of random variable resulting from a change of scale could be be made analogous with the pooling and subsampling operations often used in deep models.

It's an interesting idea, but it will probably take more work (and more tutorial expositions of RG) to catch the attention of this community.

Deep learning and the renormalization group

Cédric Bény

15 Jan 2013 ICLR 2013 conference submission readers: everyone

Decision: reject

Yann LeCun

05 Apr 2013 ICLR 2013 submission review readers: everyone

Review: It seems to me like there could be an interesting connection between approximate inference in graphical models and the renormalization methods.

There is in fact a long history of interactions between condensed matter physics and graphical models. For example, it is well known that the loopy belief propagation algorithm for inference minimizes the Bethe free energy (an approximation of the free energy in which only pairwise interactions are taken into account and high-order interactions are ignored). More generally, variational methods inspired by statistical physics have been a very popular topic in graphical model inference.

The renormalization methods could be relevant to deep architectures in the sense that the grouping of random variable resulting from a change of scale could be be made analogous with the pooling and subsampling operations often used in deep models.

It's an interesting idea, but it will probably take more work (and more tutorial expositions of RG) to catch the attention of this community.

QUANTA MAGAZINE illuminating science

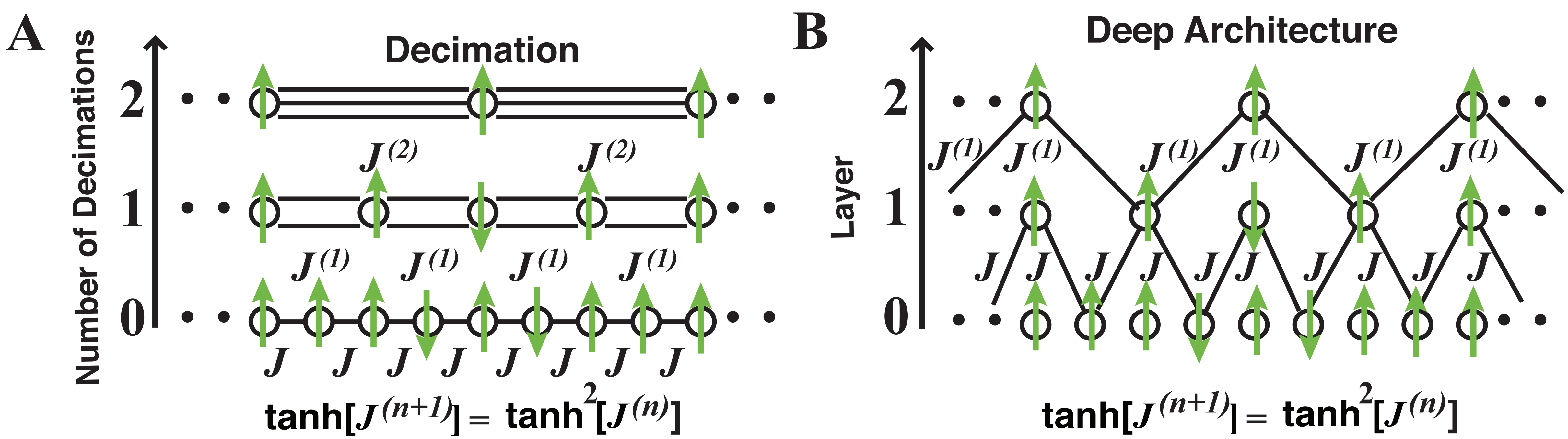
A Common Logic to Seeing Cats and Cosmos

Olena Shmahalo / Quanta Magazine

There may be a universal logic to how physicists, computers and brains tease out important features from among other irrelevant bits of data.

"An exact mapping between the Variational Renormalization Group and Deep Learning", Mehta and Schwab, 1410.3831

“Exact Mapping”



$$e^{-H(\boldsymbol{h})} = \sum_{\boldsymbol{x}} e^{T(\boldsymbol{x},\boldsymbol{h}) - H(\boldsymbol{x})}$$

RG Transformation

$$e^{-E(\boldsymbol{h})} = \sum_{\boldsymbol{x}} e^{-E(\boldsymbol{x},\boldsymbol{h})}$$

Boltzmann Machine

Harsh comments below the Quanta Magazine article

Noah says:
December 26, 2014 at 9:54 am

I just spend an hour reading Mehta-Schwab paper from the beginning to end. Let me say that "A Common Logic to Seeing Cats and Cosmos" is a sensationalist article about a trivial paper, which will have no impact whatsoever. The whole M-S paper is based on the fact that couplings of two systems appear in more than one context and that distributions can sometimes appear as marginal distributions on product spaces. There is no one-to-one mappings between renormalization group (RG) scheme of Kadanoff and Restricted Boltzmann Machines (RBM) in Deep Neural Networks (DNN) in their paper. What they show is that RBM can be represented as a RG scheme with a very specific choice of coupling function T in equation (18). Conveniently, this coupling function depends on the Hamiltonian of the spin system, which it normally should not. Equivalence in equations (8) and (9) is also not correct. Condition (9) of course implies that the scheme is exact, but not the other way around, unless the authors make some implicit assumptions about coupling function T not mentioned in the paper. The paper contains no non-trivial ideas, it does not "open up a door to something very exciting", and I will not hold my breath expecting new breakthroughs because of this connection.

https://www.quantamagazine.org/20141204-a-common-logic-to-seeing-cats-and-cosmos/

Dictionary: RG vs Deep Learning

Property	Variational RG	Deep Belief Networks
How input distribution is defined	Hamiltonian defining P(v)	Data samples drawn from P(v)
How interactions are defined	T(v,h)	E(v,h)
Exact transformation	$Tr_{\mathbf{h}} e^{T(\mathbf{v},\mathbf{h})} = 1$	KL divergence between P(v) and variational distribution is zero
Approximations	Minimize or bound free energy differences	Minimize the KL divergence
Method	Analytic (mostly)	Numerical
What happens under coarse-graining	Relevant operators grow/irrelevant shrink	New features emerge

From Schwab's talk at PI: http://pirsa.org/displayFlash.php?id=16080006

More on the DL-RG Connections

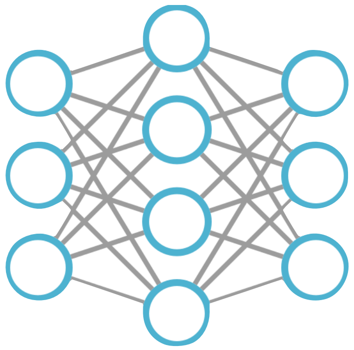
- “Why does deep and cheap learning work so well ”, Lin, Tegmark, Rolnick, 1608.08225

- Comment on the above paper, Schwab and Mehta, 1609.03541

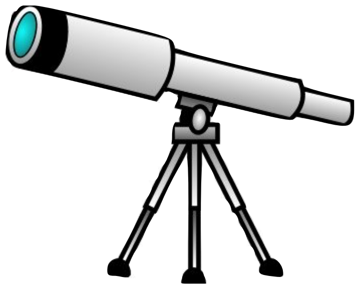
- PCA meets RG, Bradde and Bialek, 1610.09733

- Mutual information RG, Koch-Janusz and Ringel, 1704.06279

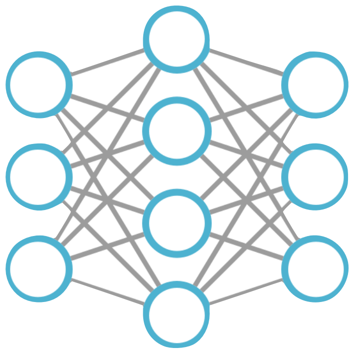
- Media coverage/blog posts/student term papers etc



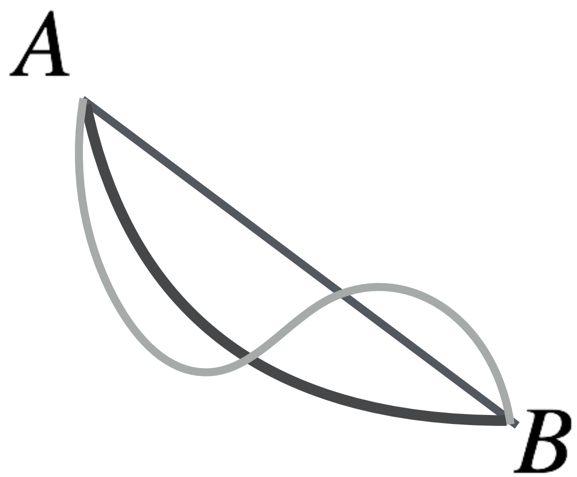
Neural Network Renormalization Group

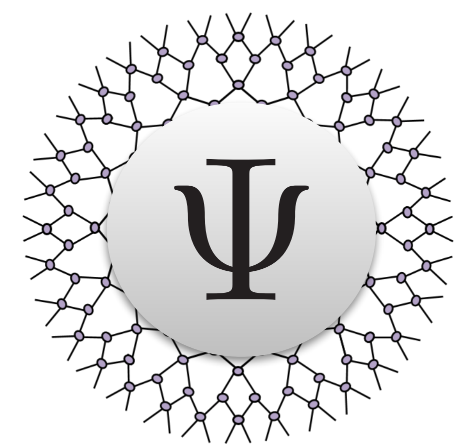


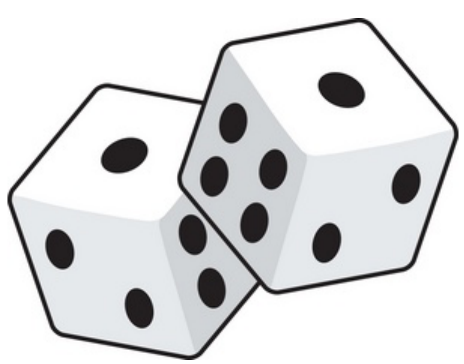
Shuo-Hui Li and LW, 1802.02840

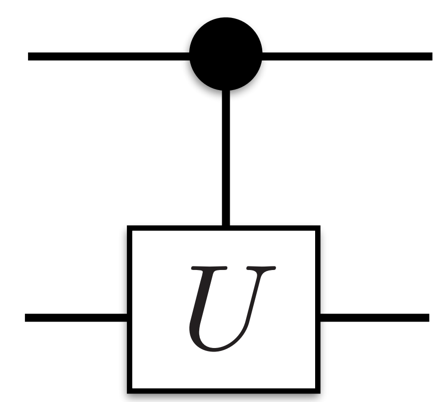


Neural Network

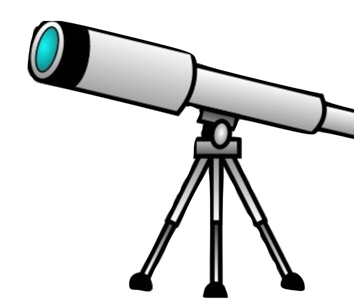






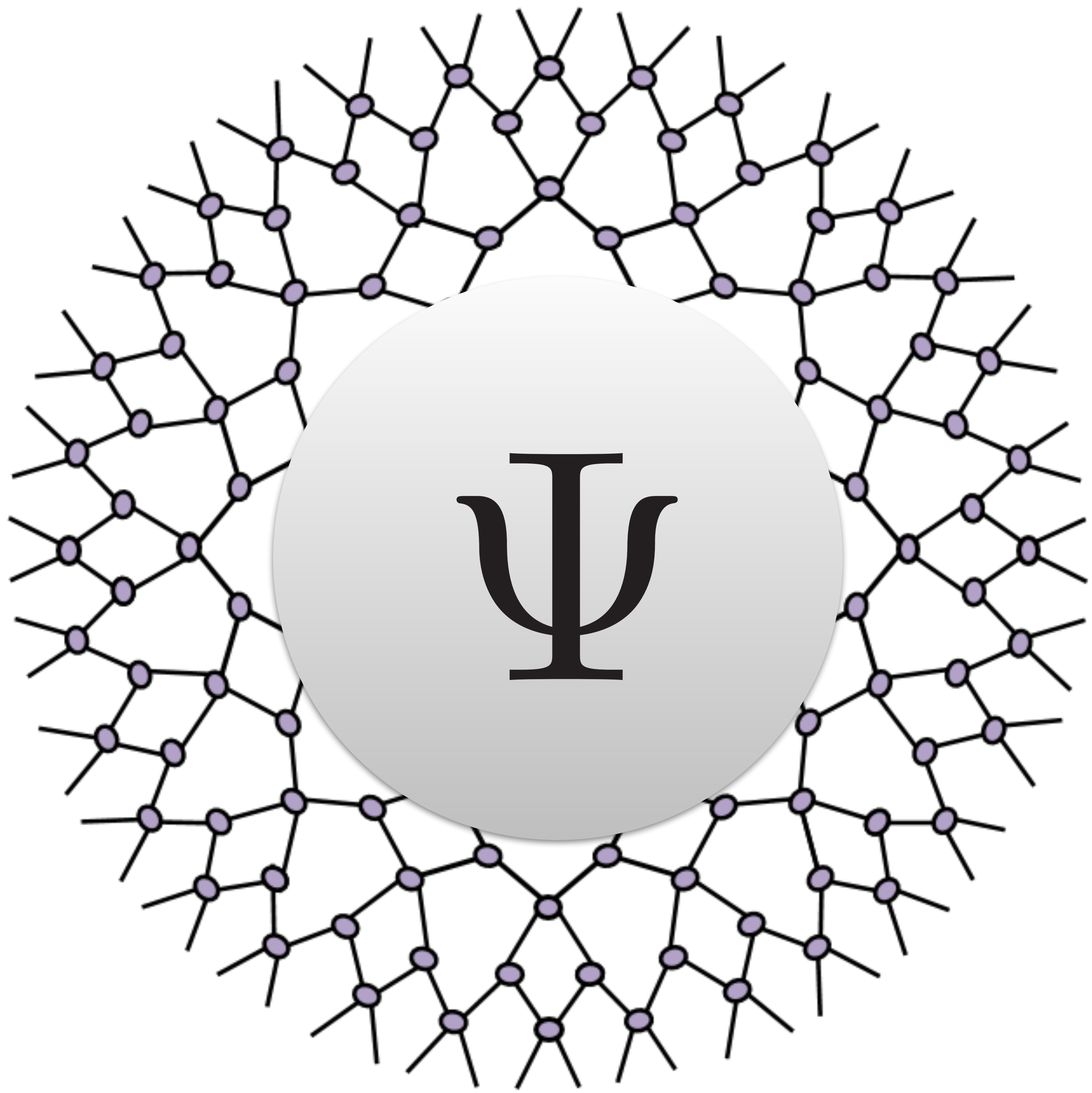


Renormalization Group



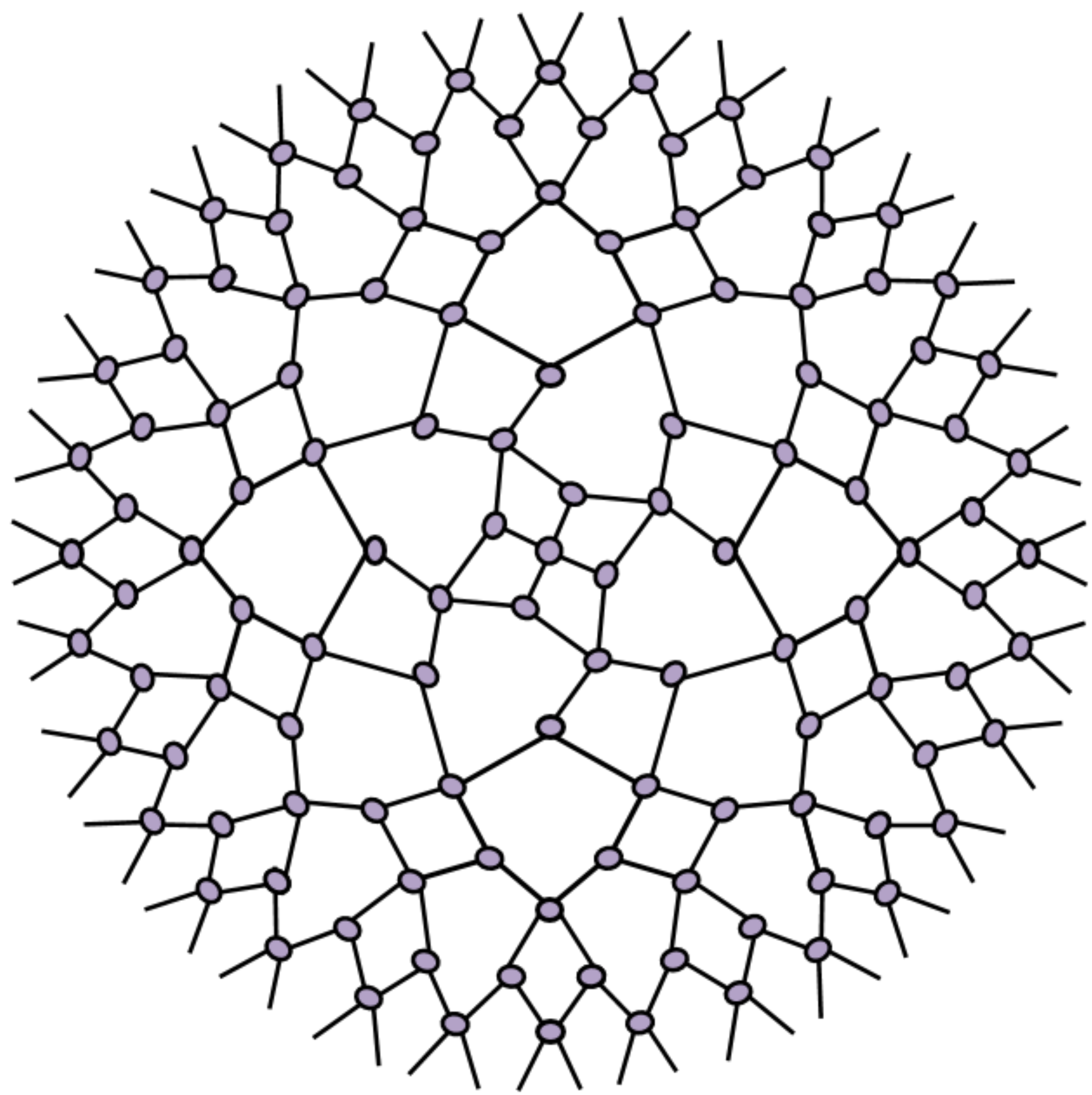
Shuo-Hui Li and LW, 1802.02840

Multi-Scale Entanglement Renormalization Ansatz



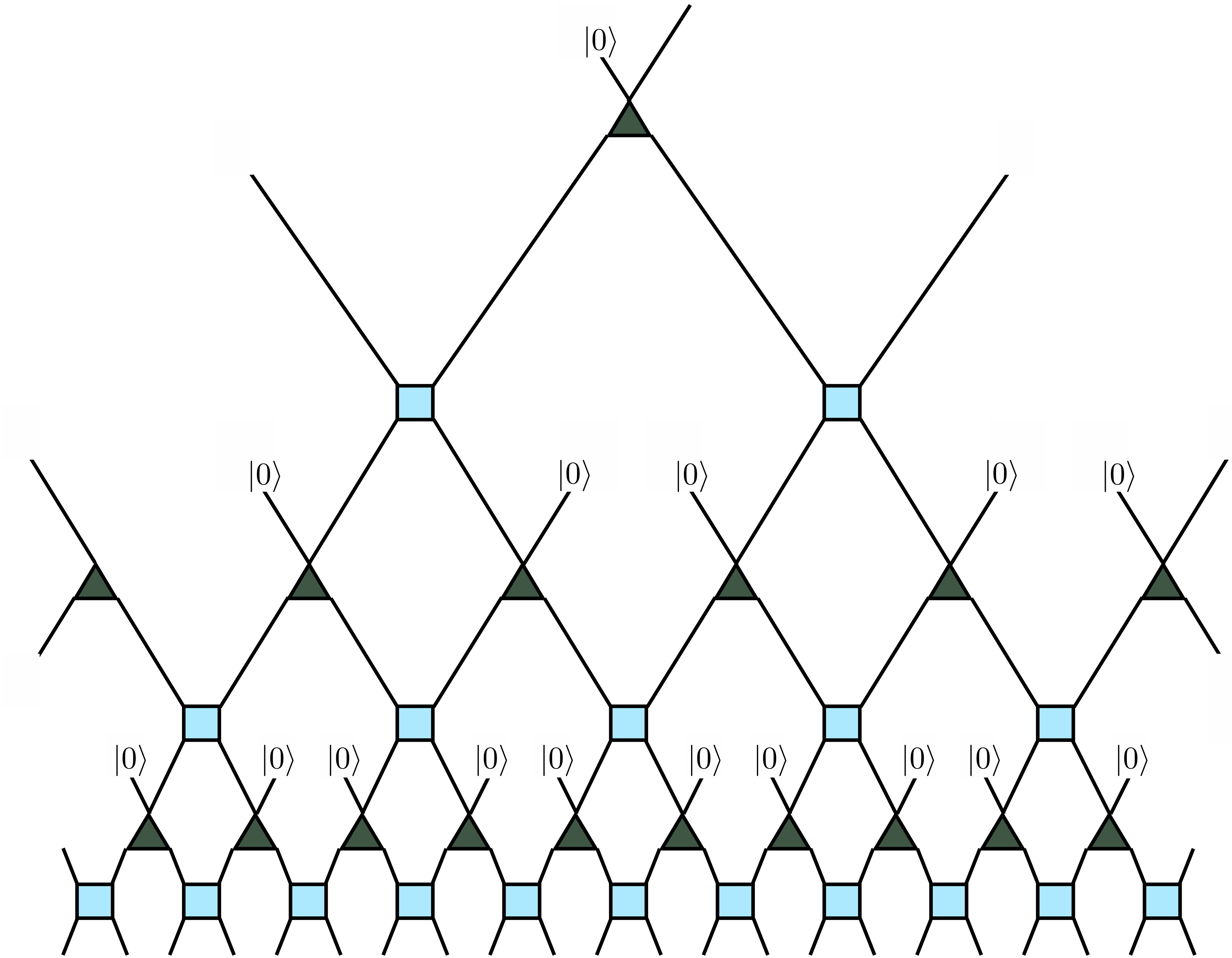
Vidal 2006

Multi-Scale Entanglement Renormalization Ansatz

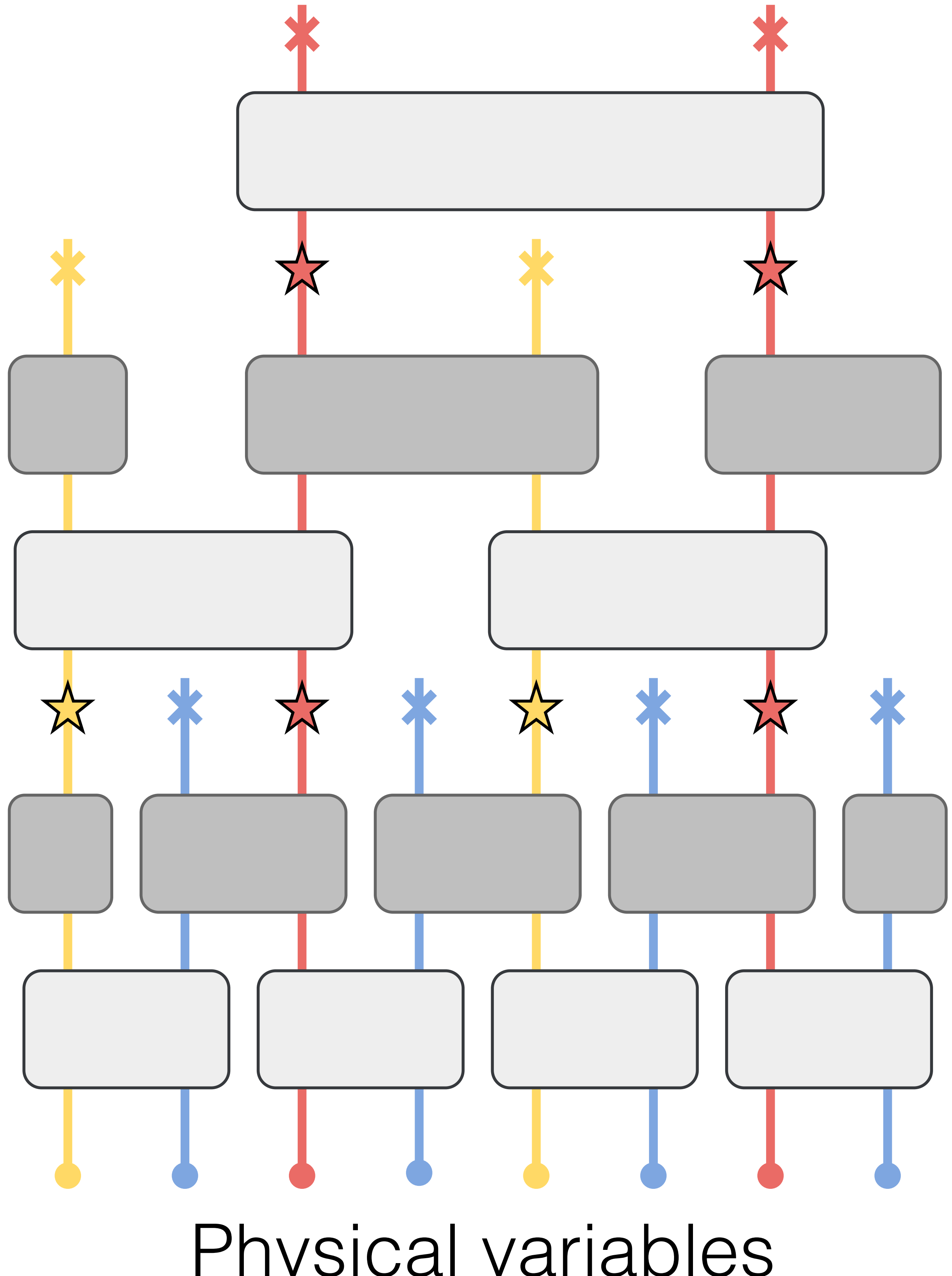


Vidal 2006

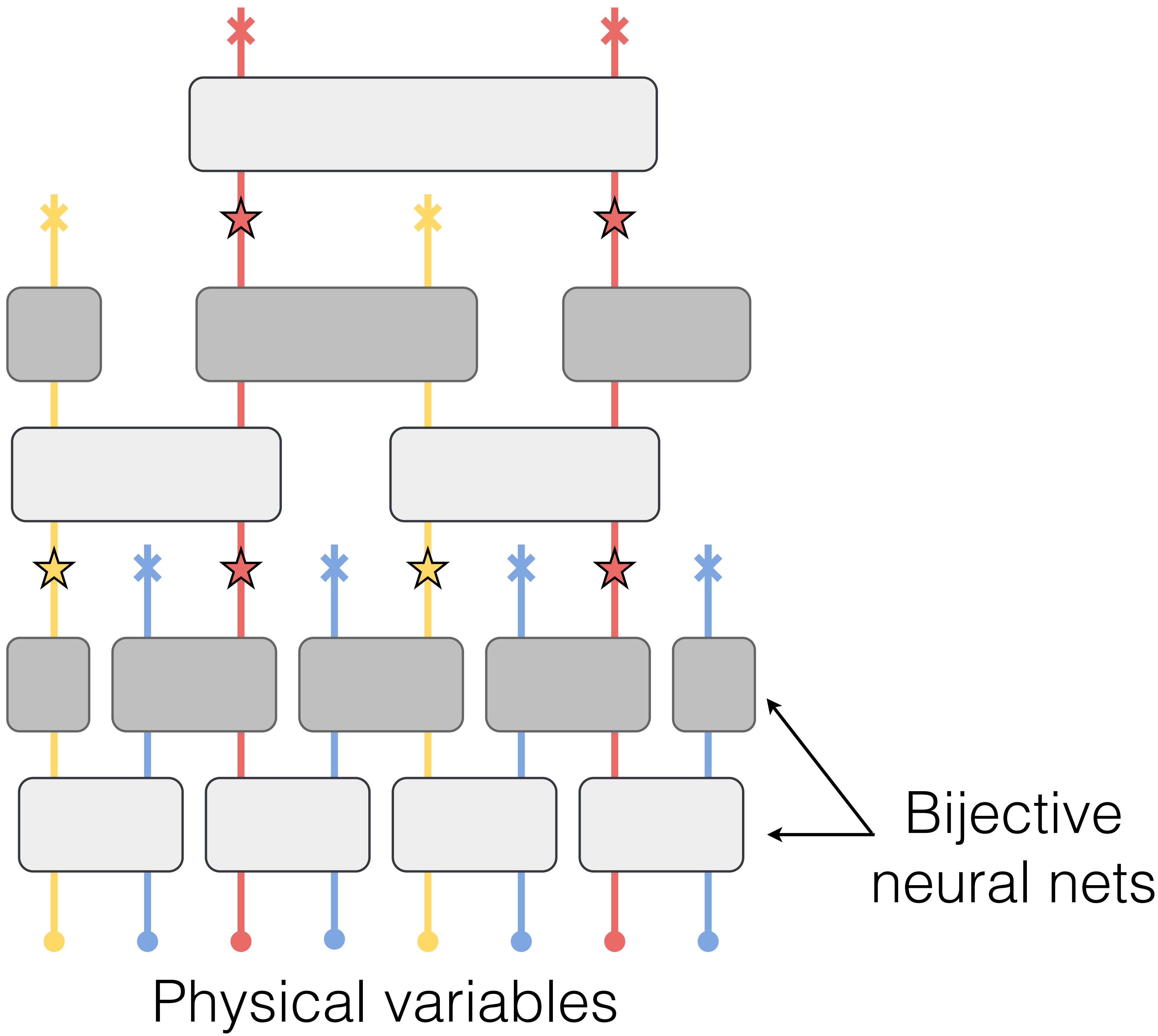
MERA as a quantum circuit



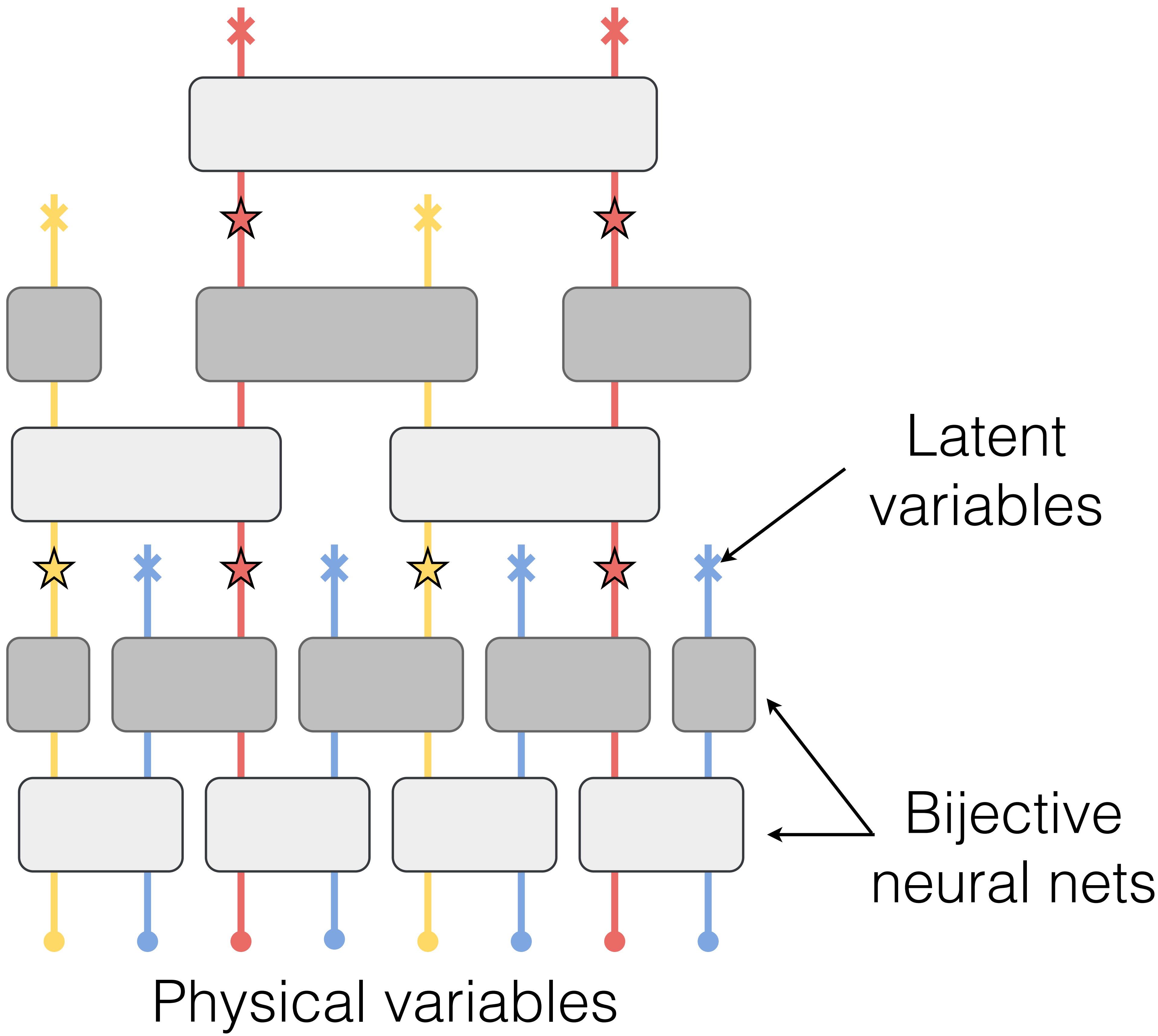
Neural Network Renormalization Group



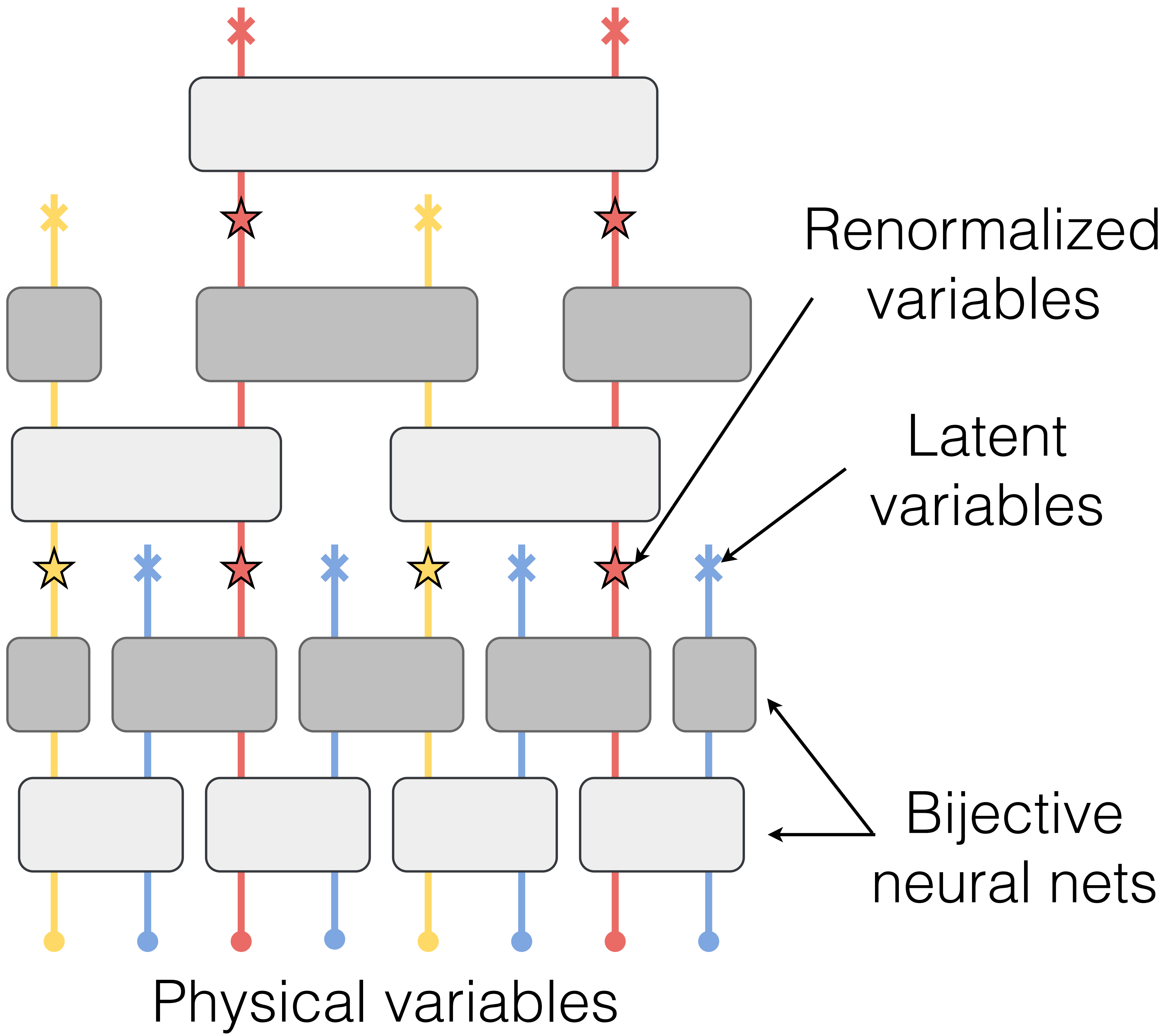
Neural Network Renormalization Group



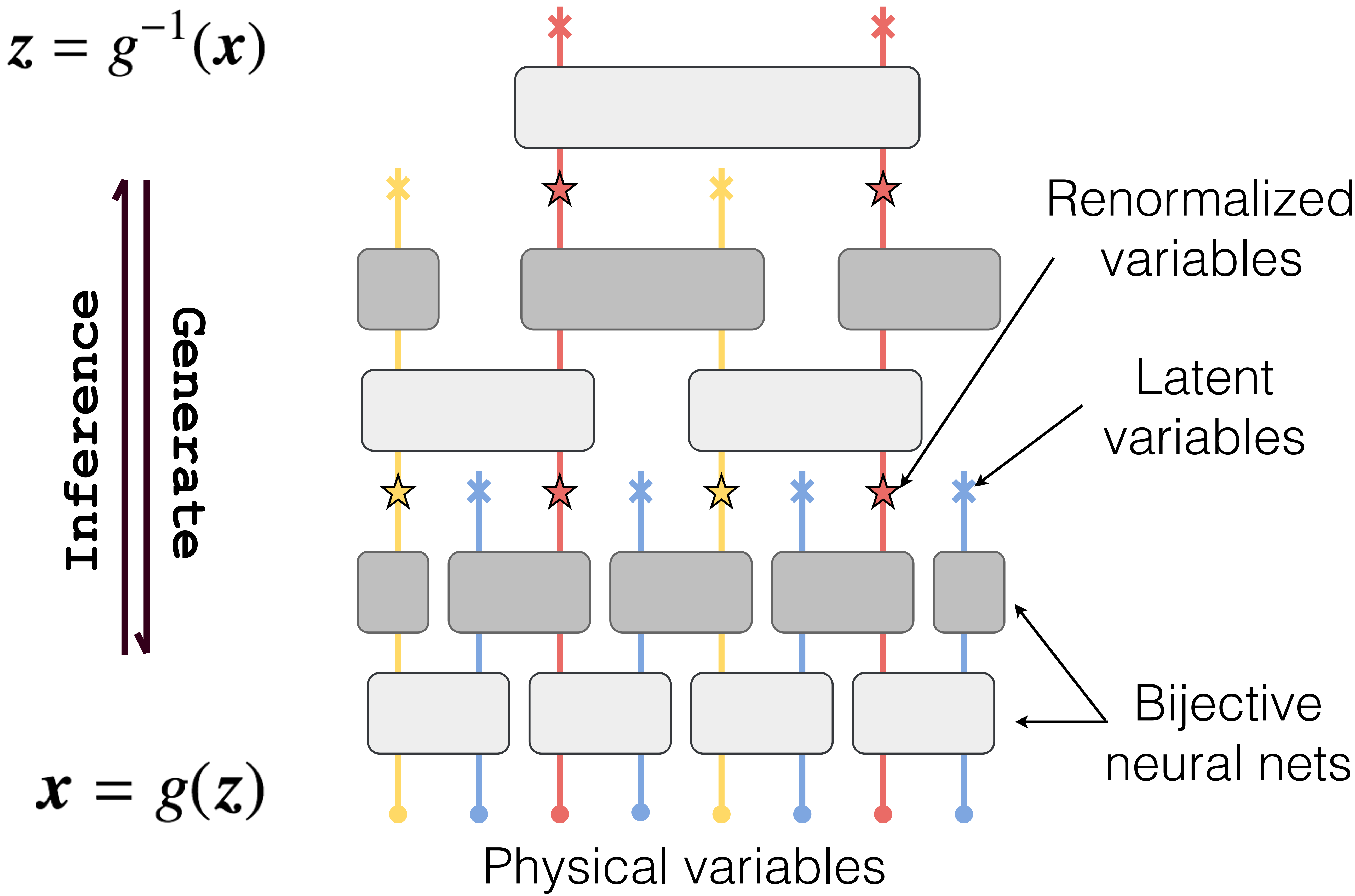
Neural Network Renormalization Group



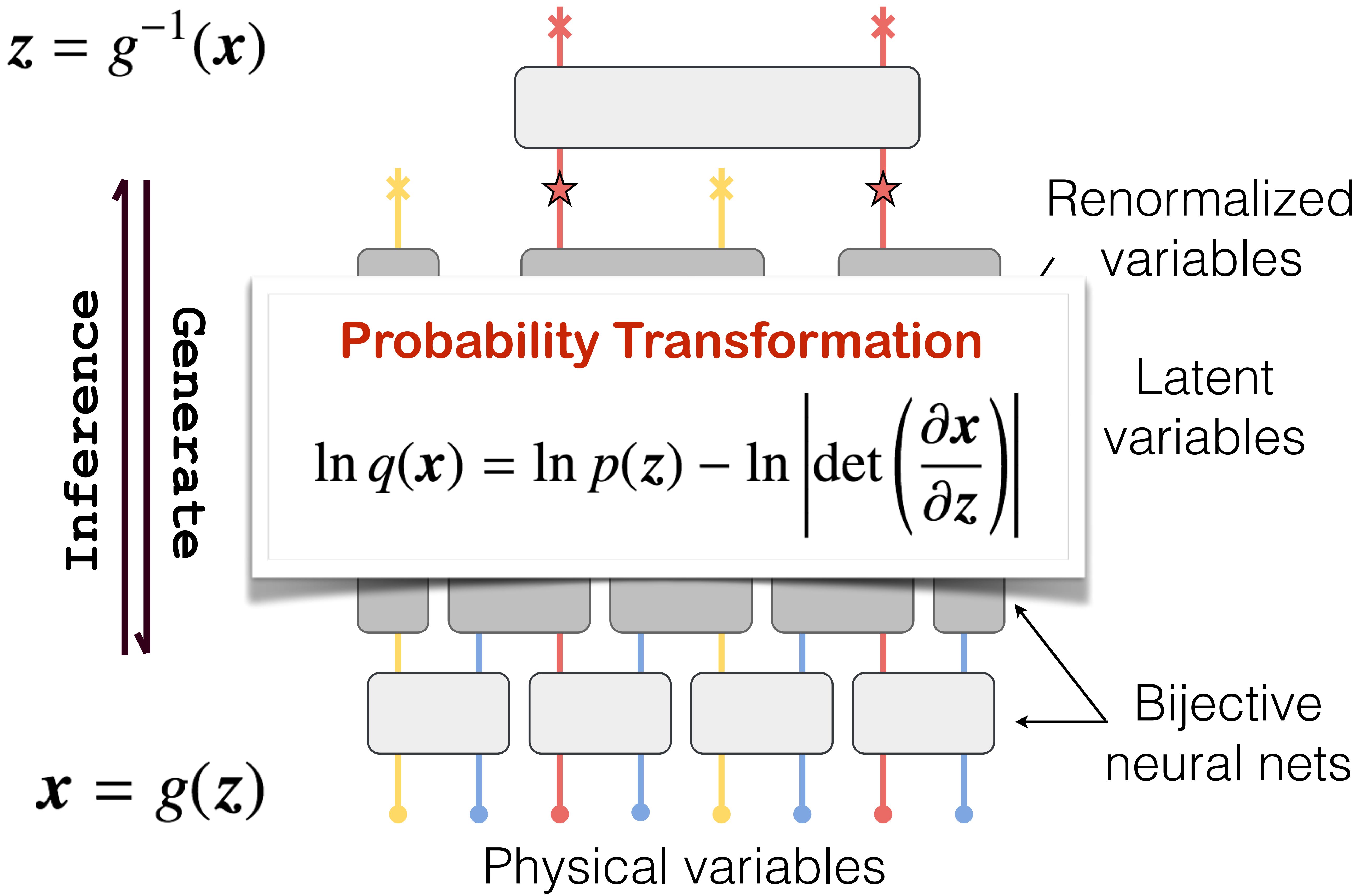
Neural Network Renormalization Group



Neural Network Renormalization Group



Neural Network Renormalization Group



Bijector Block

Bijective & Differentiable map, i.e., Diffeomorphism

Forward

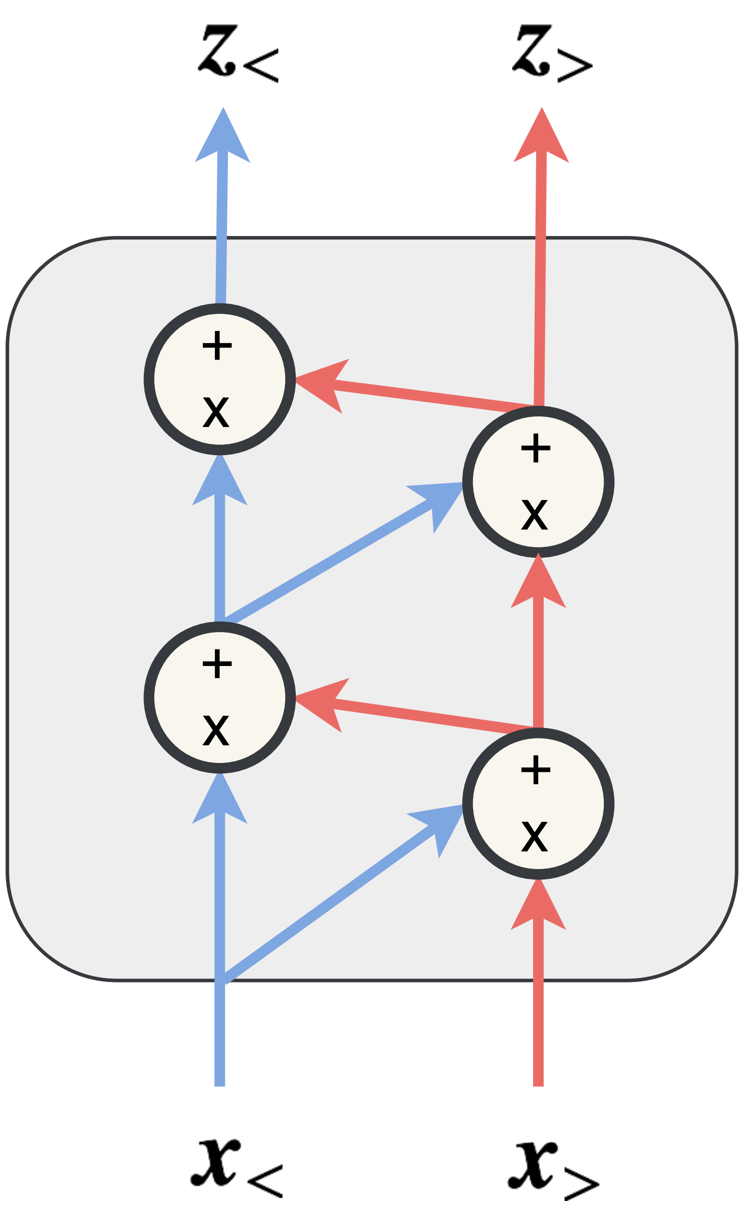
$$\begin{cases} \boldsymbol{x}_< = \boldsymbol{z}_< \\ \boldsymbol{x}_> = \boldsymbol{z}_> \odot e^{s(\boldsymbol{z}_<)} + t(\boldsymbol{z}_<) \end{cases}$$

Backward

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

Log-Det-Jacobian

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



Normalizing flow, Rezende et al,1505.05770

Special case: Real NVP, Dinh et al,1605.08803

Bijector Block

Bijective & Differentiable map, i.e., Diffeomorphism

Forward

Arbitrary neural nets

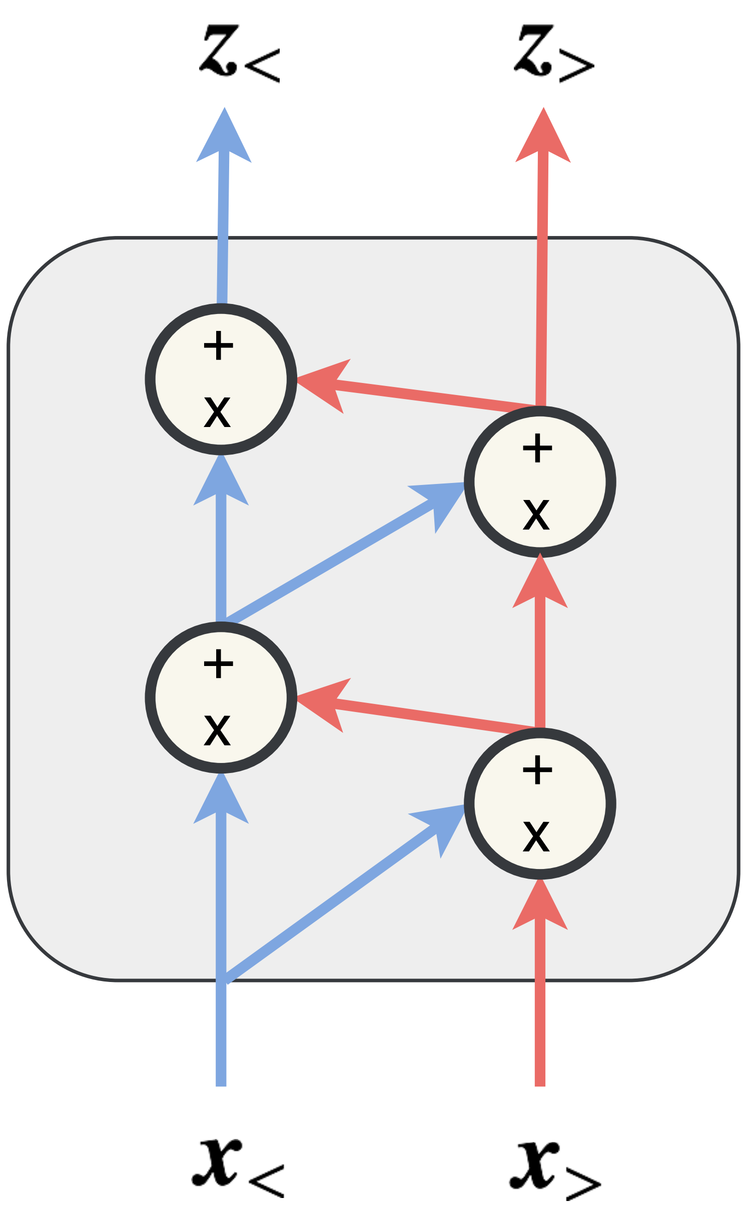
$$\begin{cases} \boldsymbol{x}_< = \boldsymbol{z}_< \\ \boldsymbol{x}_> = \boldsymbol{z}_> \odot e^{s(\boldsymbol{z}_<)} + t(\boldsymbol{z}_<) \end{cases}$$

Backward

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

Log-Det-Jacobian

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

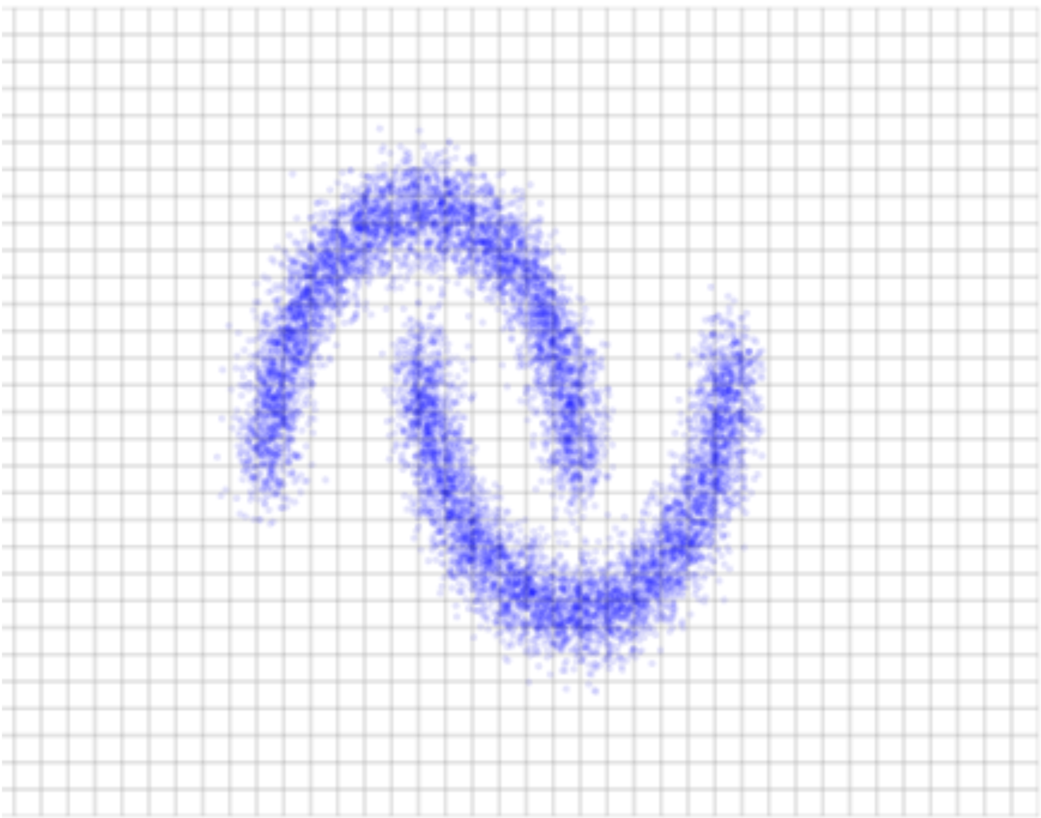


Normalizing flow, Rezende et al,1505.05770

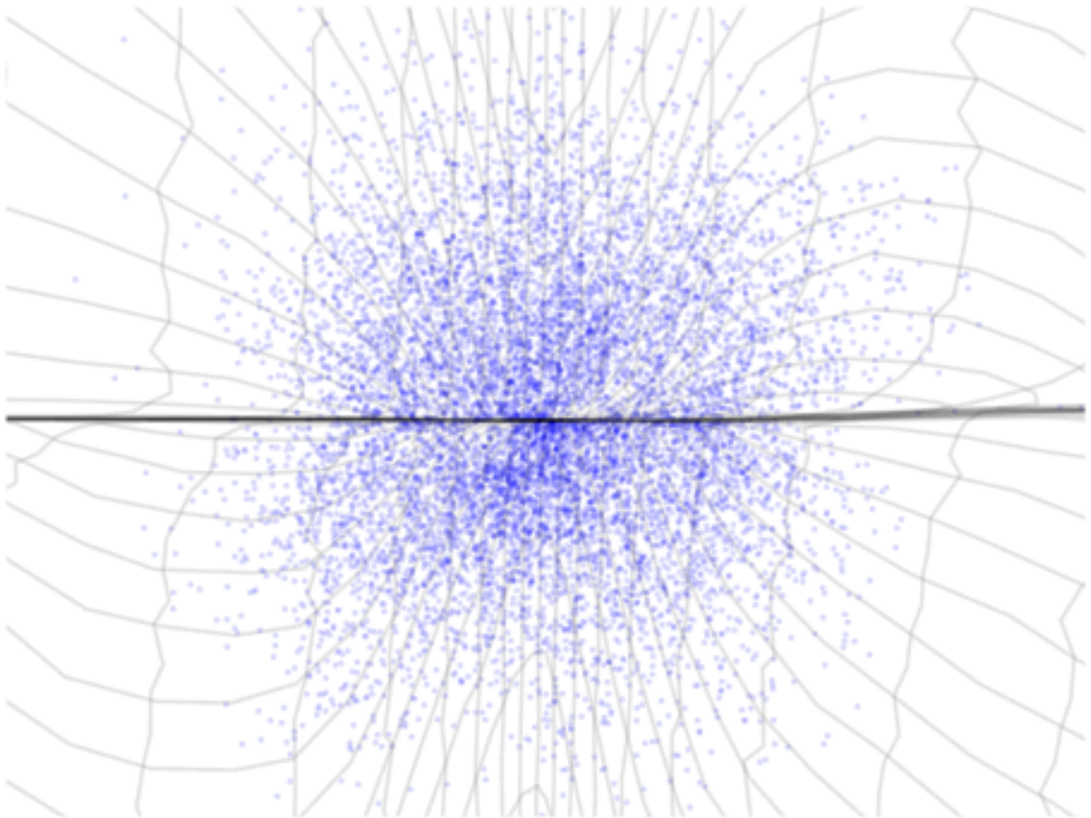
Special case: Real NVP, Dinh et al,1605.08803

Inference

Data space $g^{-1}(\boldsymbol{x}) = \boldsymbol{z}$ Latent space



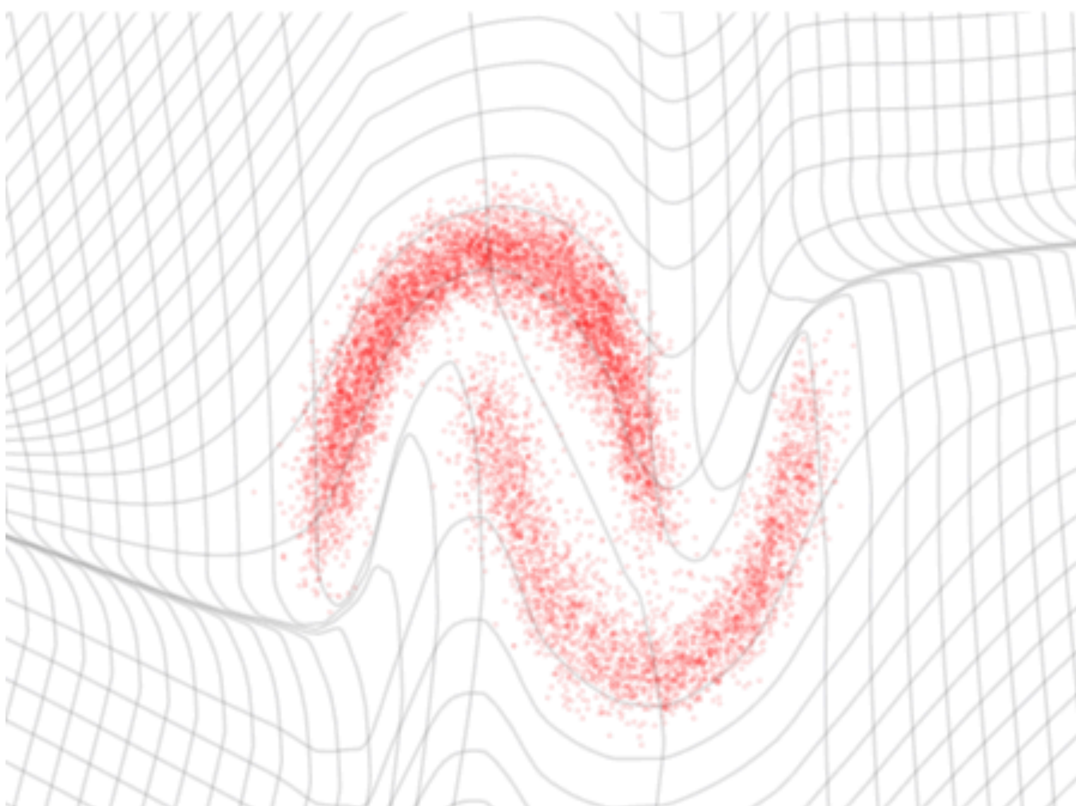
$\Rightarrow$



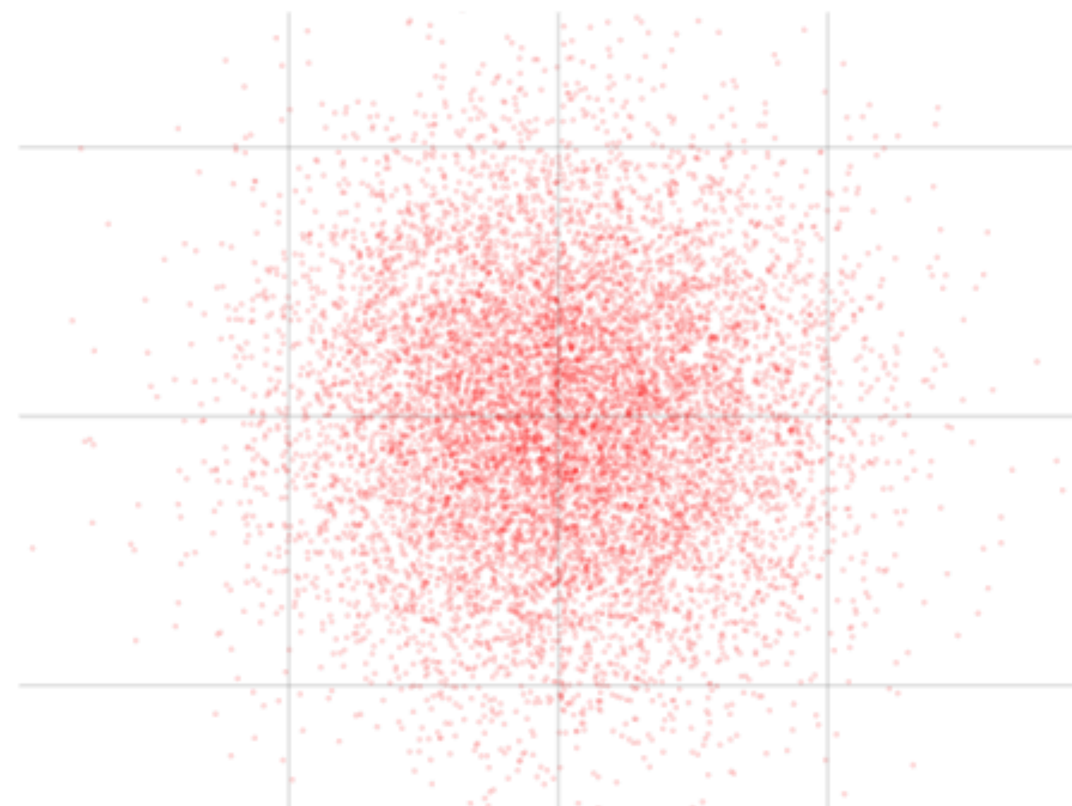
Real NVP, Dinh et al,1605.08803

Generate

Data space $\qquad x = g(z) \qquad$ Latent space



$\Leftarrow$

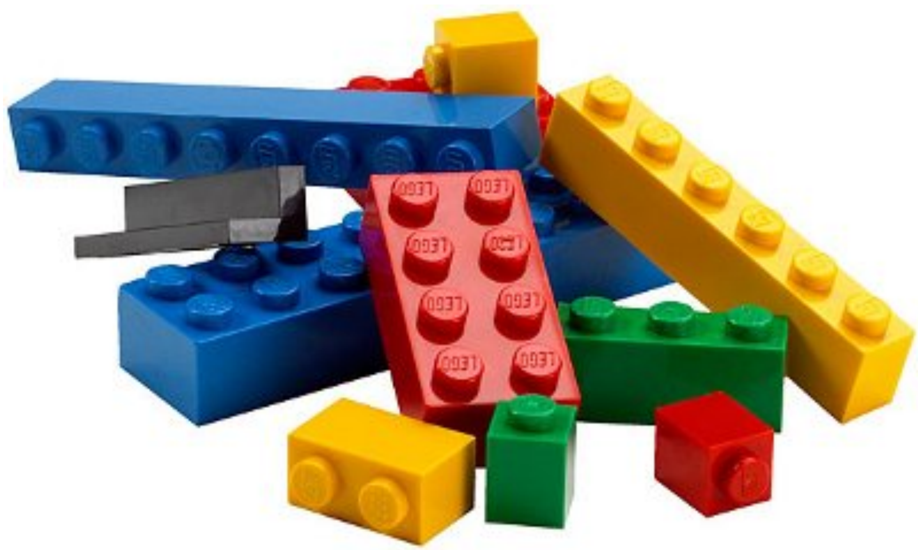


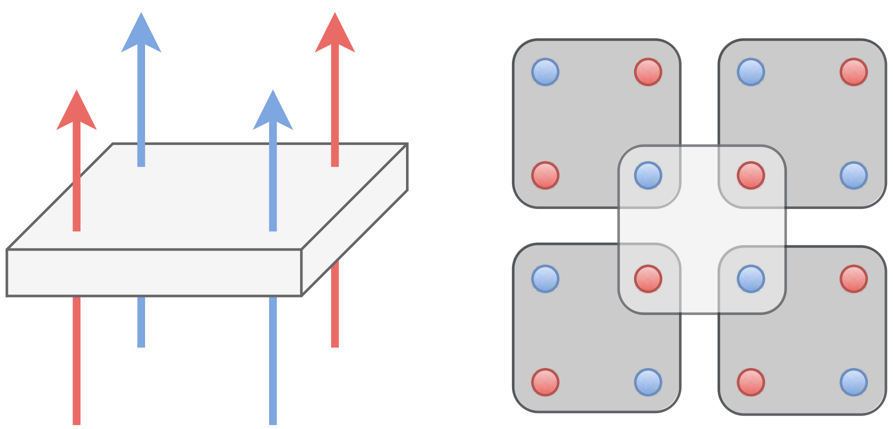
Real NVP, Dinh et al,1605.08803

Bijectors form a group

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

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

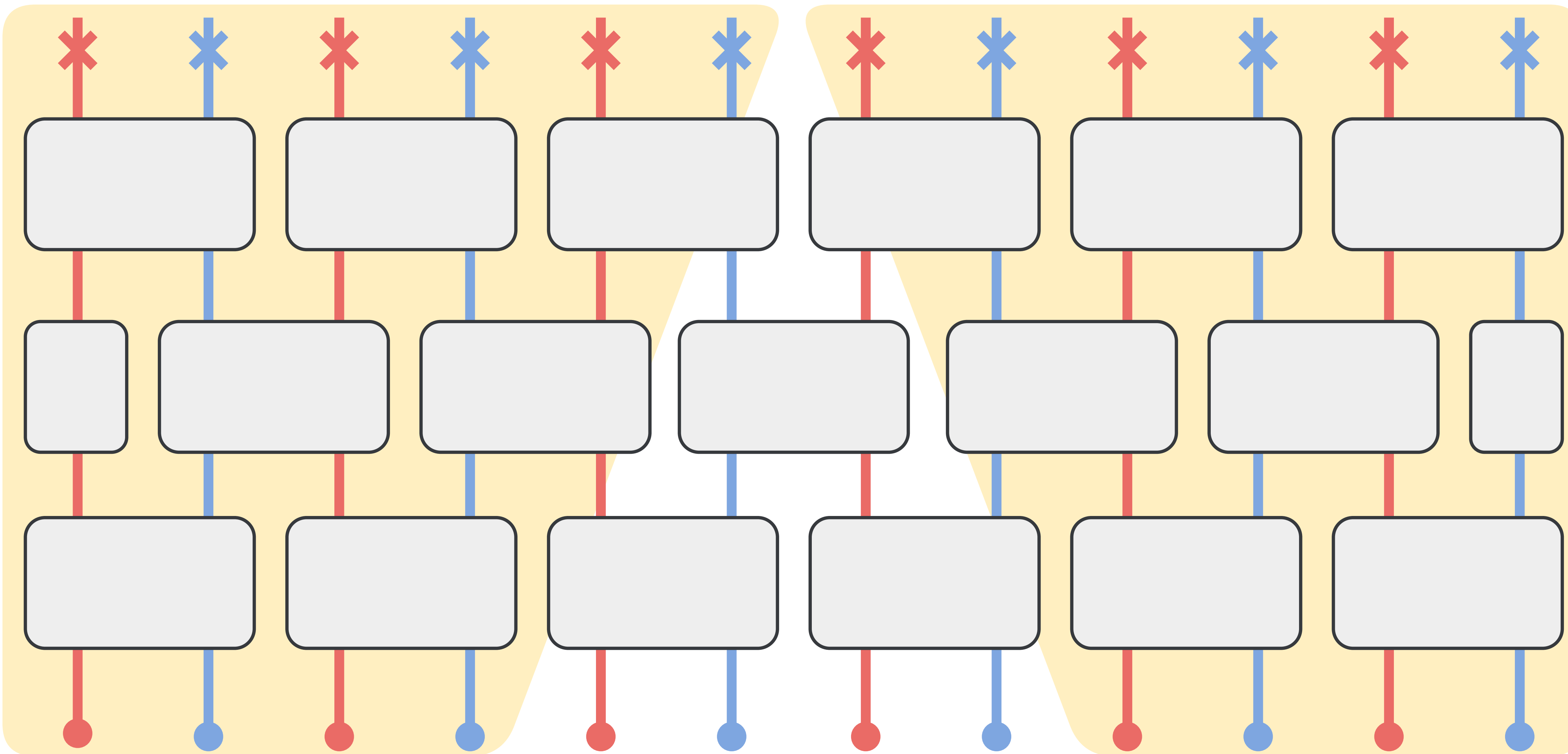




Modular design

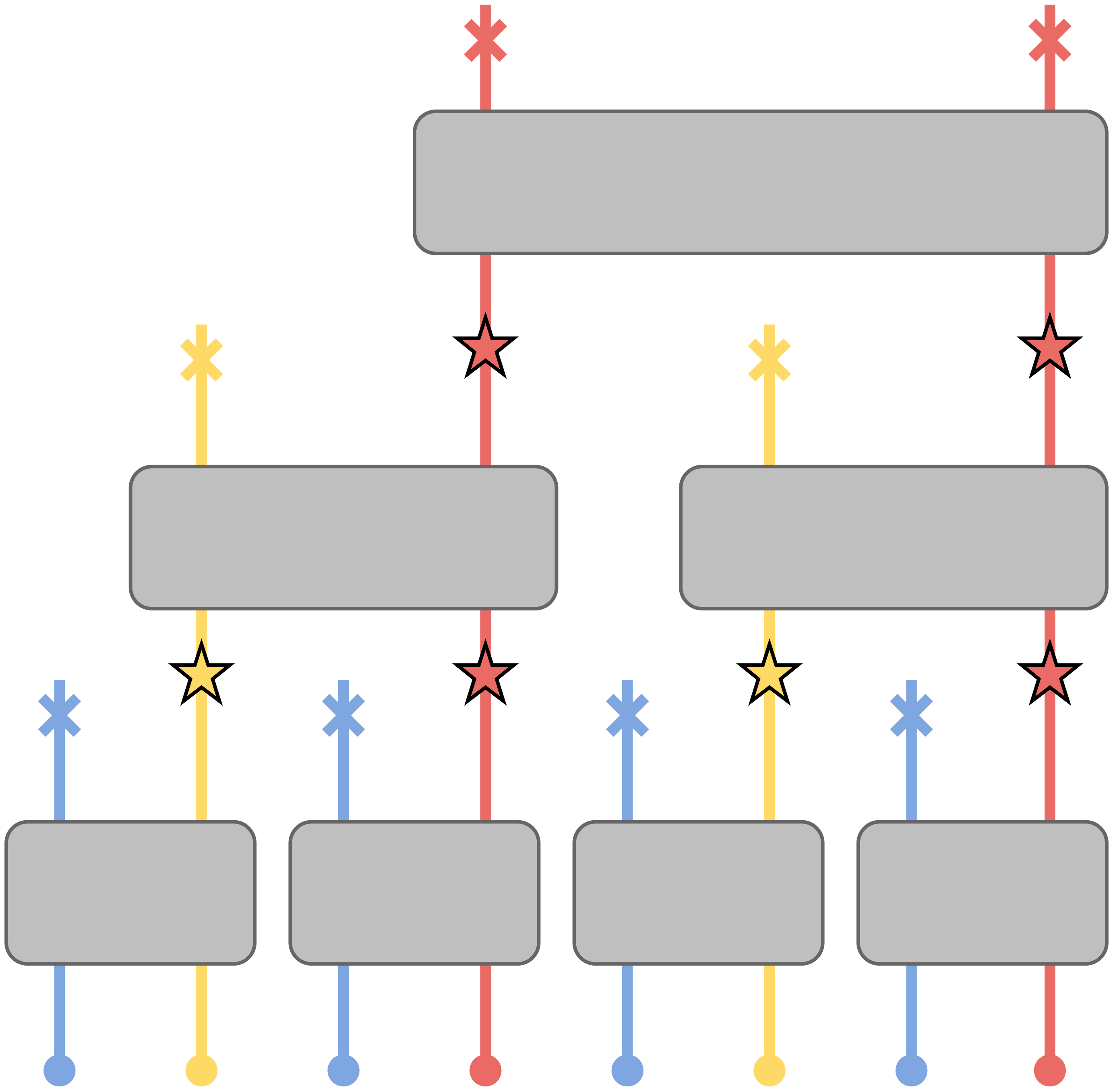
Flexible blocks and stacking

“Disentangler” only architecture

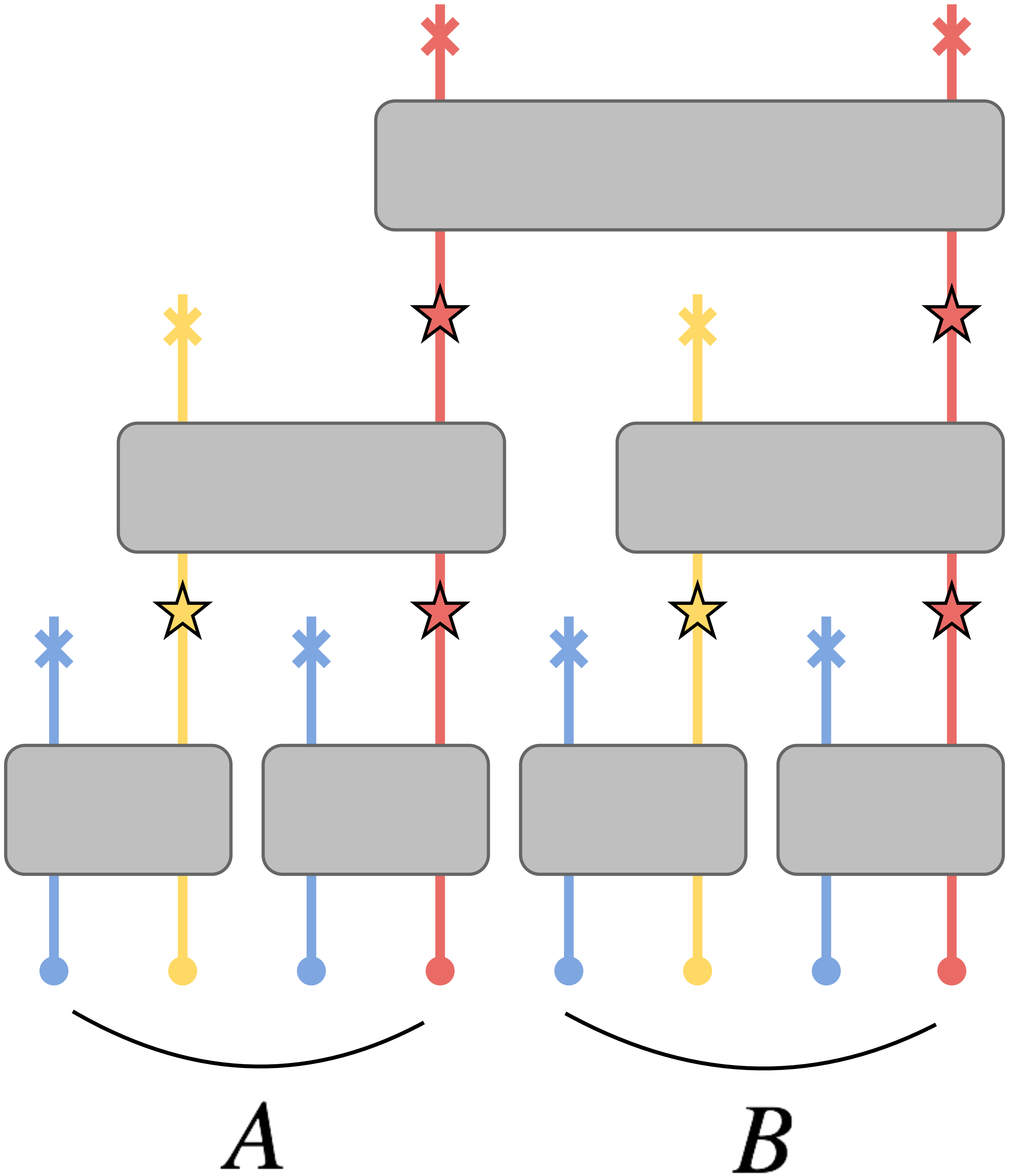


Correlation length ~ Network depth

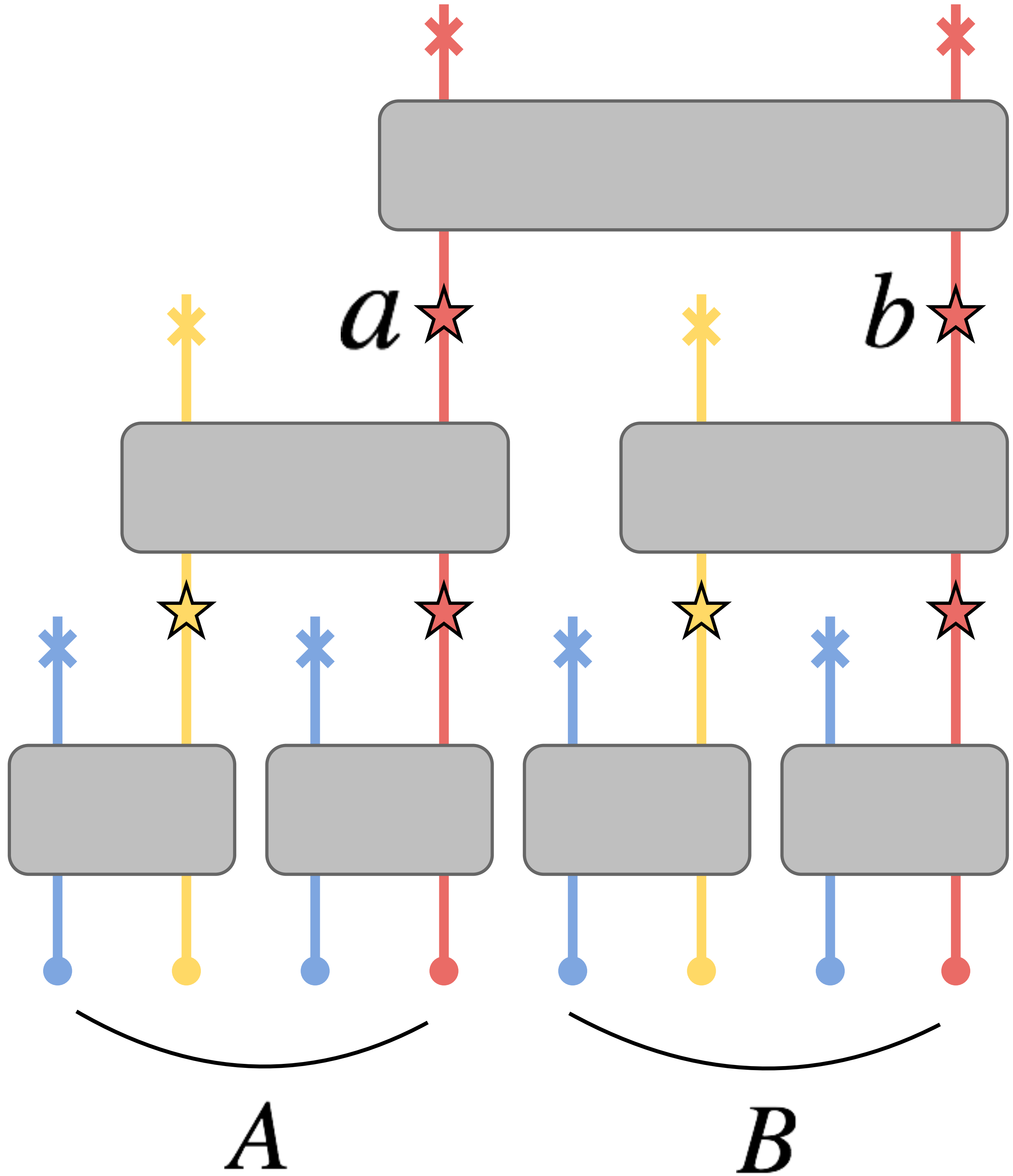
“Decimator” only architecture



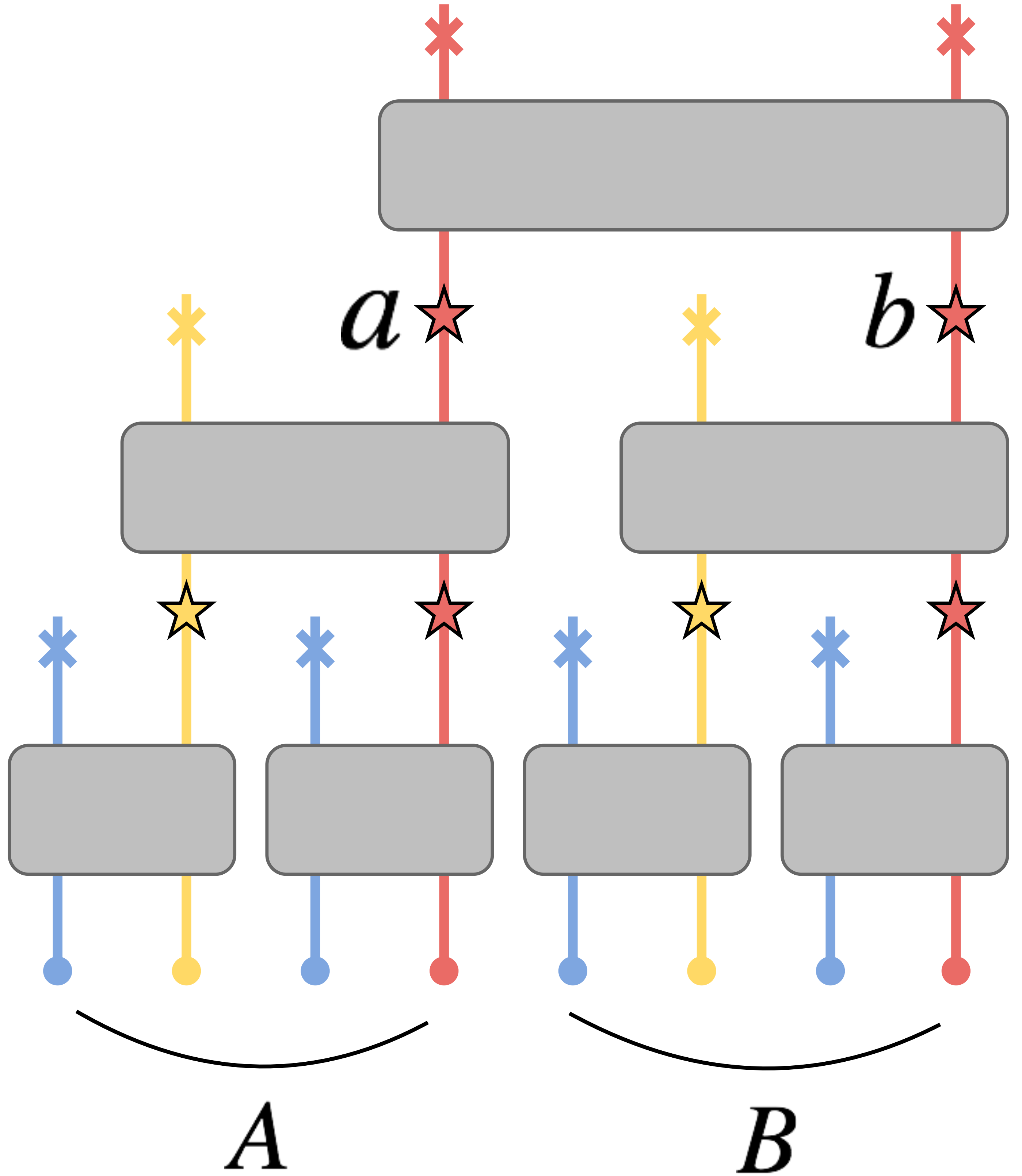
“Decimator” only architecture



“Decimator” only architecture



“Decimator” only architecture



$$I(A:B) = I(a:b)$$

Mutual Information Bottleneck

Probability Density *Estimation*

Given a dataset, learn its probability density by minimizing the negative likelihood

$$\mathrm{NLL}_{\boldsymbol{\theta}} = -\sum_{\boldsymbol{x} \in \mathrm{dataset}} \ln q_{\boldsymbol{\theta}}(\boldsymbol{x})$$

Network parameters

Equivalent to reduce the Kullback–Leibler divergence

$$\mathbb{KL}\left(\frac{\pi(\boldsymbol{x})}{Z} \,\Big\|\, q_{\boldsymbol{\theta}}(\boldsymbol{x})\right)$$

However, typical Stat-Mech problem has access only to the energy function

Probability Density *Distillation*

Learn from the data generated by the network itself

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

Probability Density *Distillation*

Learn from the data generated by the network itself

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

Entropy

Probability Density *Distillation*

Learn from the data generated by the network itself

$$\mathcal{L}_{\boldsymbol{\theta}} = \int \mathrm{d}\boldsymbol{x}\, q_{\boldsymbol{\theta}}(\boldsymbol{x}) \big[\underbrace{\ln q_{\boldsymbol{\theta}}(\boldsymbol{x})}_{\text{Entropy}} - \underbrace{\ln \pi(\boldsymbol{x})}_{\text{Energy}} \big]$$

Probability Density *Distillation*

Learn from the data generated by the network itself

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

"Variational Free Energy"

Entropy

Energy

Probability Density *Distillation*

Learn from the data generated by the network itself

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

"Variational Free Energy"

Entropy

Energy

$$Z = \int \mathrm{d}\boldsymbol{x}\, \pi(\boldsymbol{x})$$

Partition function

Probability Density *Distillation*

Learn from the data generated by the network itself

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

"Variational Free Energy"

Entropy

Energy

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

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

"Reparametrization trick"

How to compute the gradient w.r.t random sampling ?

$$\mathcal{L}_{\theta} = \mathop{\mathbb{E}}_{z \sim p(z)} \left[\ln q(g_{\theta}(z)) - \ln \pi(g_{\theta}(z)) \right]$$

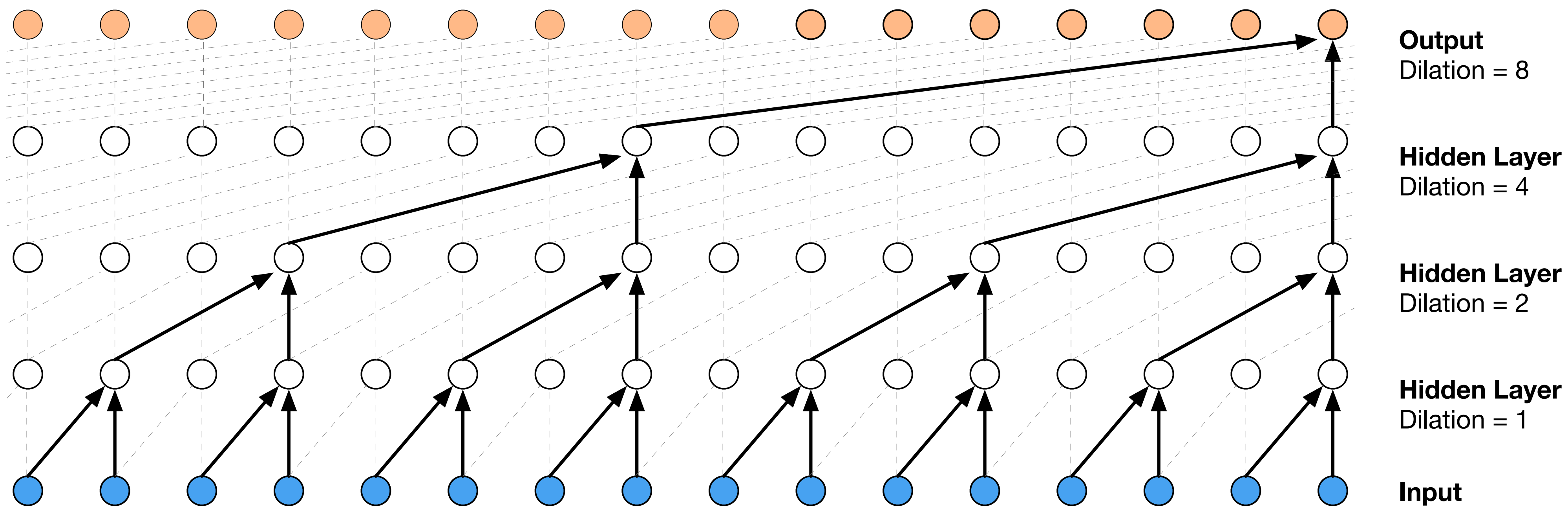
Sample from the prior distribution

Network parameters

End-to-end training using **back-propagation**

Interlude: WaveNet Story

generated signal



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:

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

DeepMind

https://deepmind.com/blog/wavenet-generative-model-raw-audio/ 1609.03499

https://deepmind.com/blog/high-fidelity-speech-synthesis-wavenet/ 1711.10433

Demo: Ising model

$$\pi(\boldsymbol{s}) = \exp\left(\frac{1}{2}\boldsymbol{s}^T K \boldsymbol{s}\right)$$

Demo: Ising model

$$\pi(\boldsymbol{s}) = \exp\left(\frac{1}{2}\boldsymbol{s}^T K \boldsymbol{s}\right)$$

decouple

$$\propto \int \mathrm{d}\boldsymbol{x} \exp\left(-\frac{1}{2}\boldsymbol{x}^T \left(K + \alpha I\right)^{-1} \boldsymbol{x} + \boldsymbol{s}^T \boldsymbol{x}\right)$$

Demo: Ising model

$$\pi(\boldsymbol{s}) = \exp\left(\frac{1}{2}\boldsymbol{s}^T K \boldsymbol{s}\right)$$

decouple

$$\propto \int \mathrm{d}\boldsymbol{x} \exp\left(-\frac{1}{2}\boldsymbol{x}^T \left(K + \alpha I\right)^{-1} \boldsymbol{x} + \boldsymbol{s}^T \boldsymbol{x}\right)$$

trace out s

$$\boxed{\pi(\boldsymbol{x}) = \exp\left(-\frac{1}{2}\boldsymbol{x}^T \left(K + \alpha I\right)^{-1} \boldsymbol{x}\right) \prod_i \cosh(x_i)}$$

Demo: Ising model

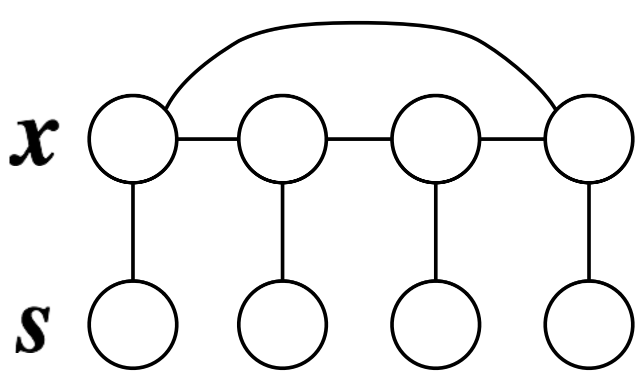
$$\pi(\boldsymbol{s}) = \exp\left(\frac{1}{2}\boldsymbol{s}^T K \boldsymbol{s}\right)$$

decouple

$$\propto \int \mathrm{d}\boldsymbol{x} \exp\left(-\frac{1}{2}\boldsymbol{x}^T (K + \alpha I)^{-1} \boldsymbol{x} + \boldsymbol{s}^T \boldsymbol{x}\right)$$

trace out s

$$\boxed{\pi(\boldsymbol{x}) = \exp\left(-\frac{1}{2}\boldsymbol{x}^T (K + \alpha I)^{-1} \boldsymbol{x}\right) \prod_i \cosh(x_i)}$$



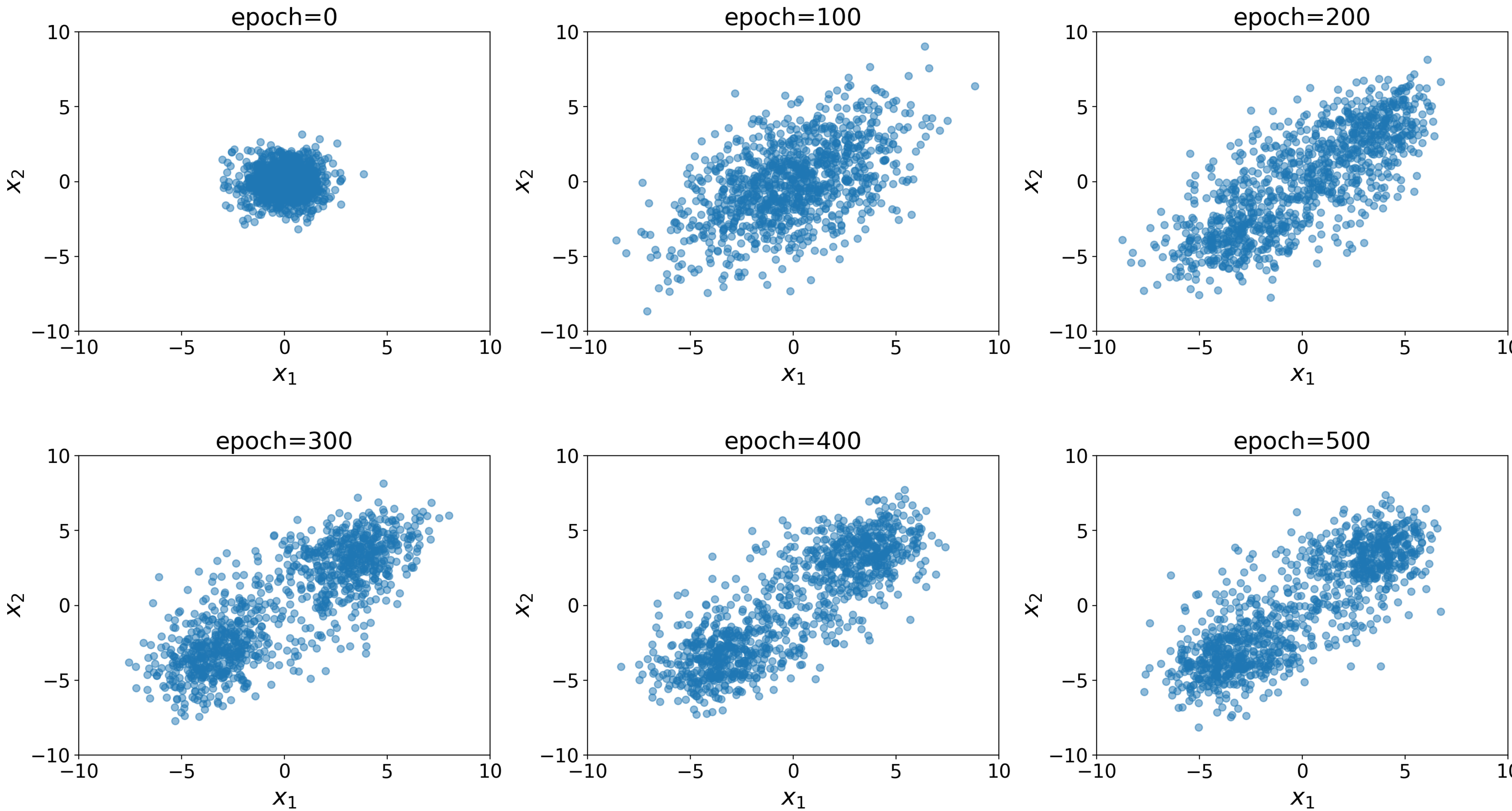
$$\pi(\boldsymbol{s}|\boldsymbol{x}) = \prod_i \left(1 + e^{-2 s_i x_i}\right)^{-1}$$

continuous dual of the Ising model

"Gaussian-Bernoulli Boltzmann Machine"

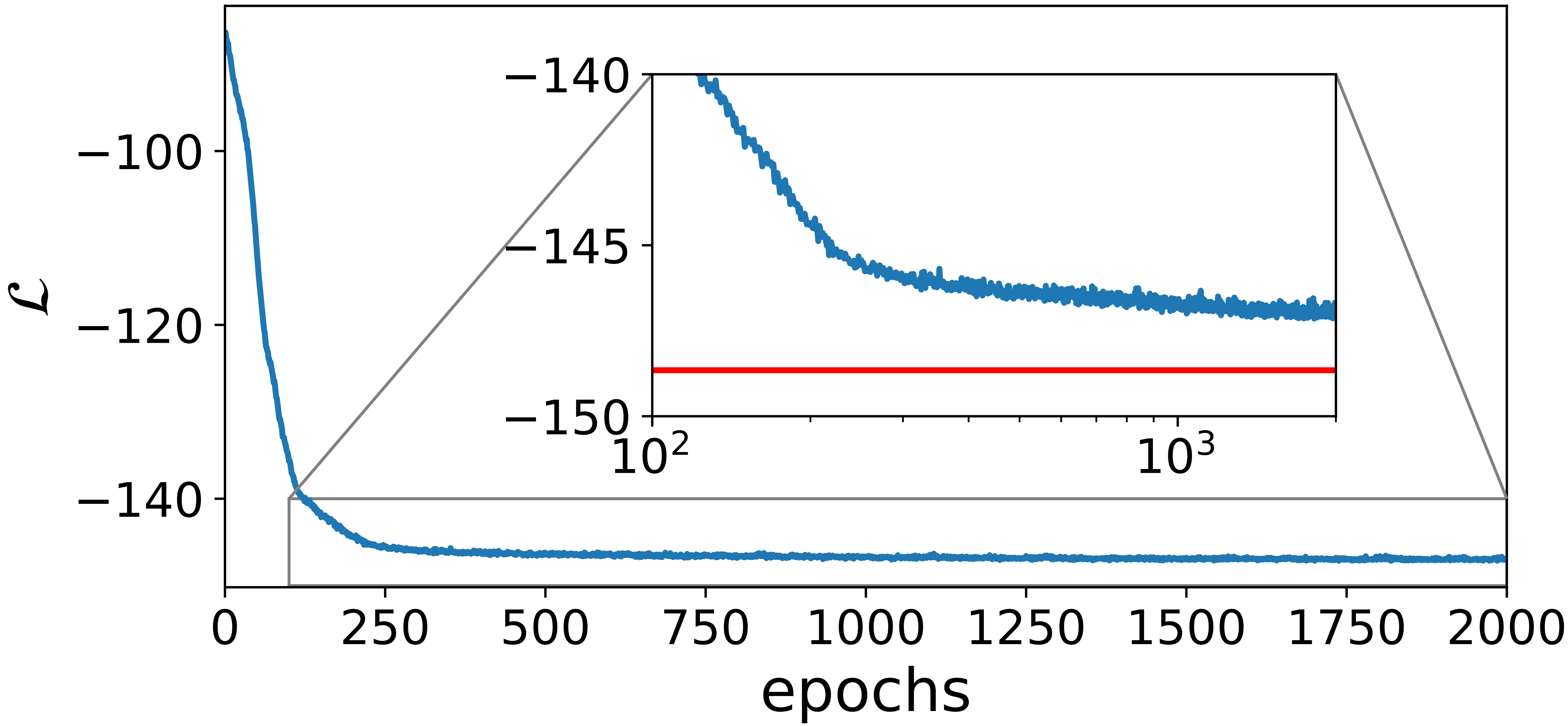
Zhang, Sutton, Storkey, Ghahramani, NIPS 2012

Generated Samples



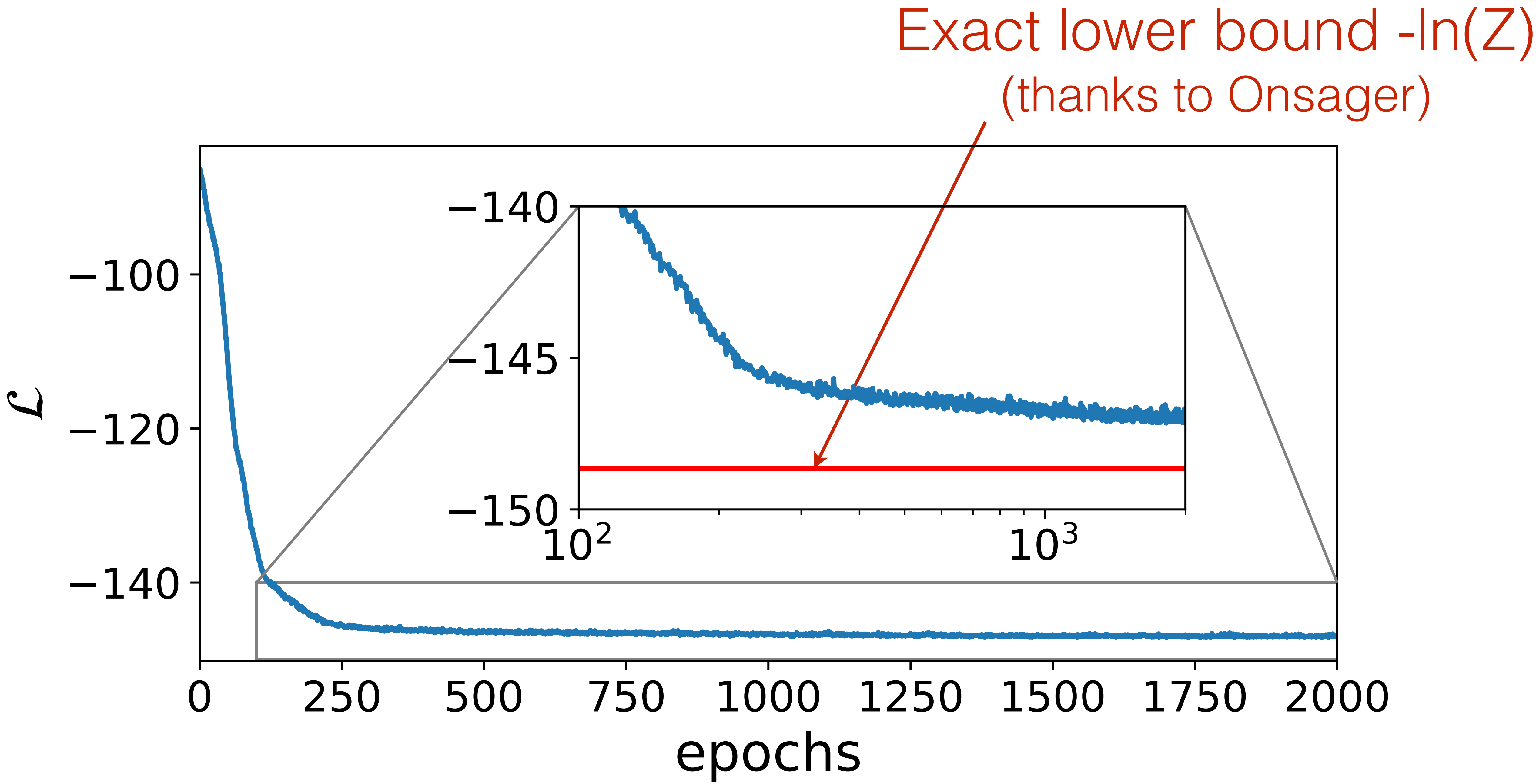
x's are continuous fields dual to Ising spins

Variational Loss



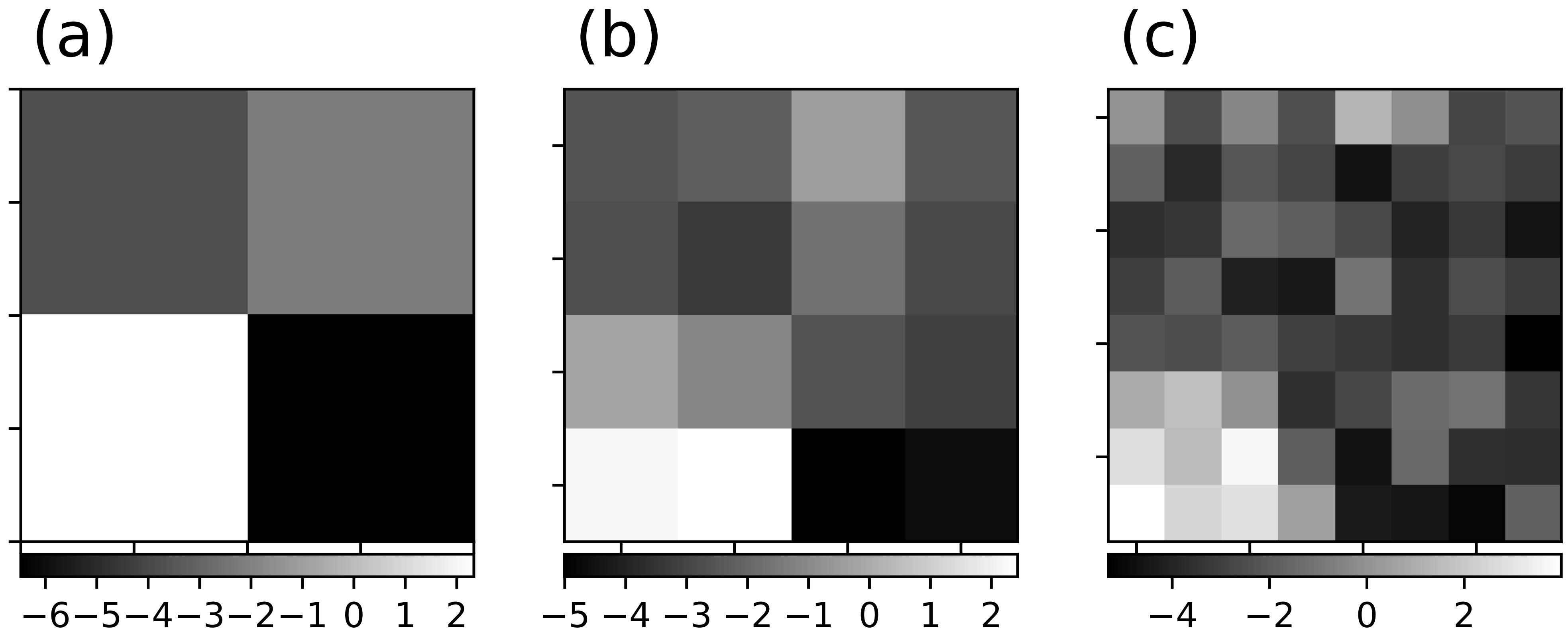
Loss can be further improved by
using more expressive networks

Variational Loss



Loss can be further improved by using more expressive networks

Renormalized Collective Variables



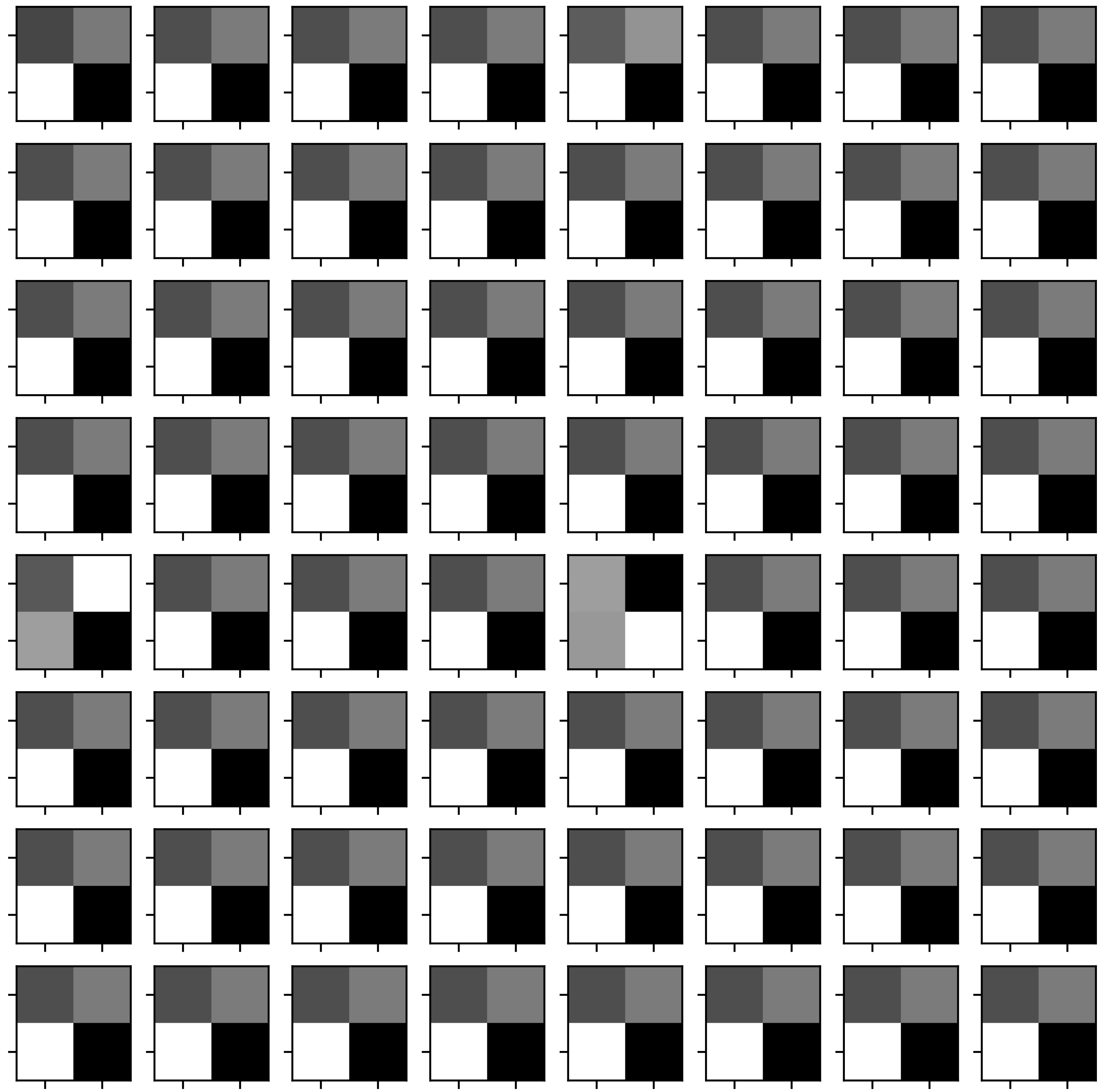
2x2

4x4

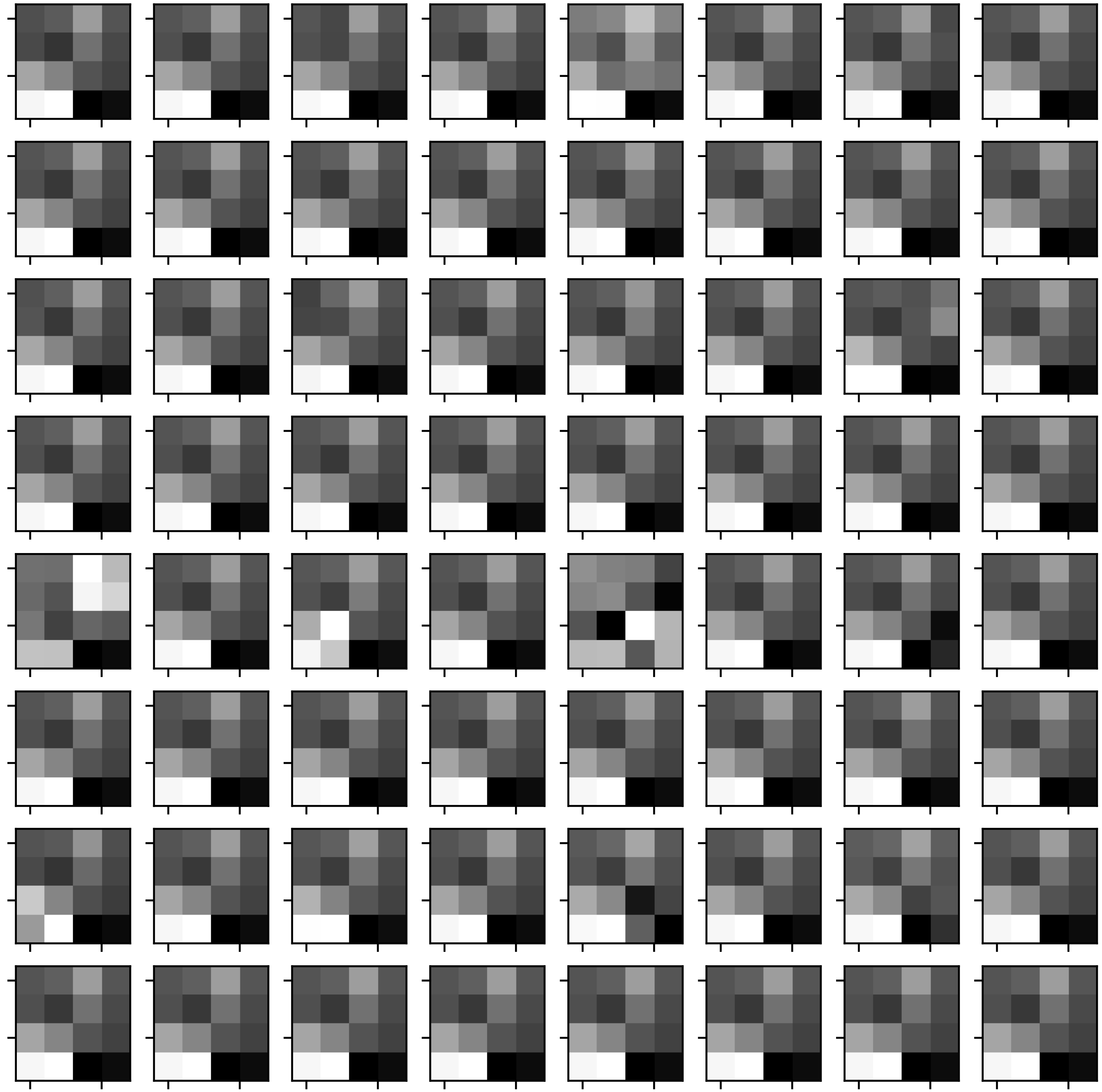
Physical Variables

Also know their effective couplings
=> renormalized energy function

Wander in the Latent Space



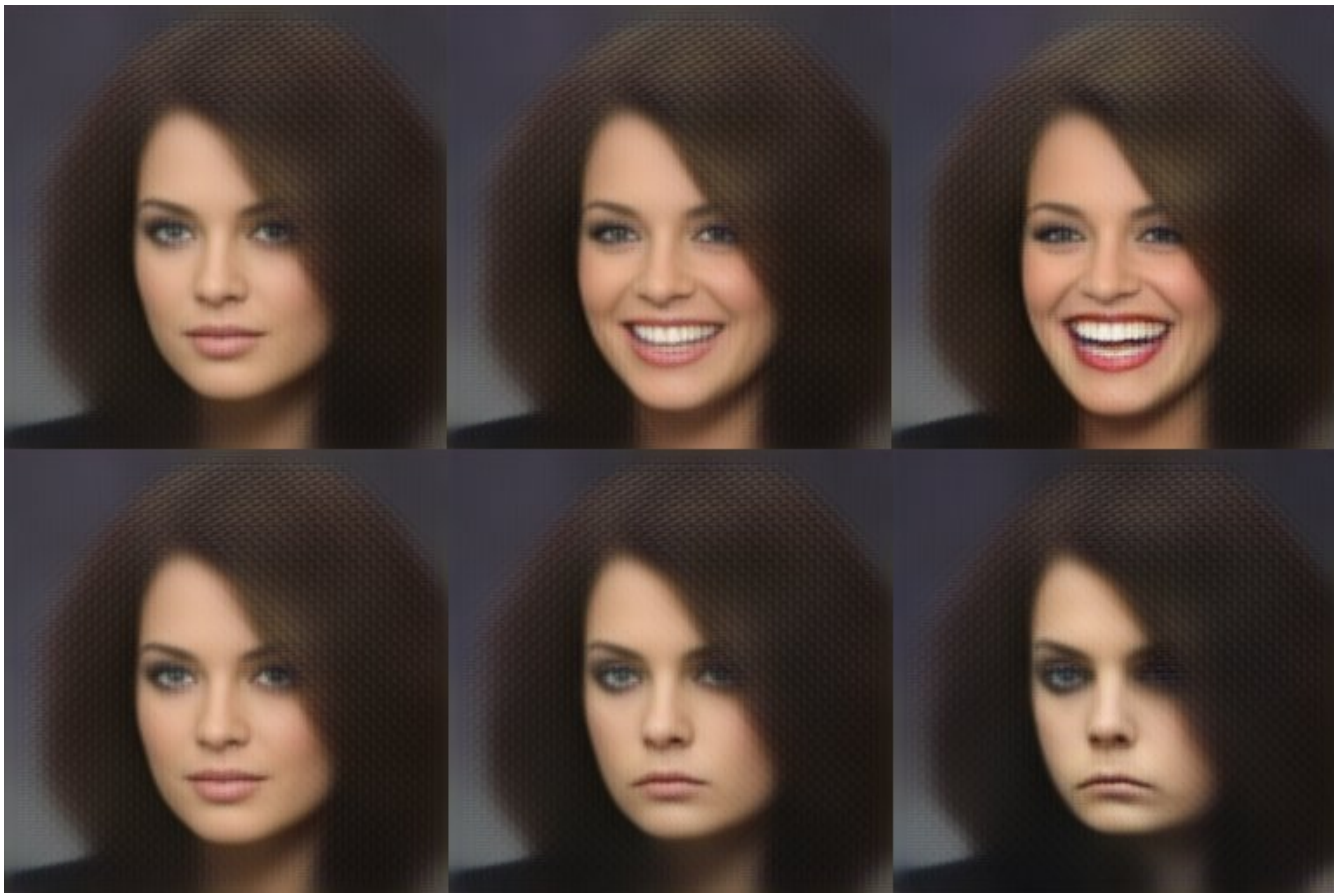
Wander in the Latent Space



Wander in the Latent Space

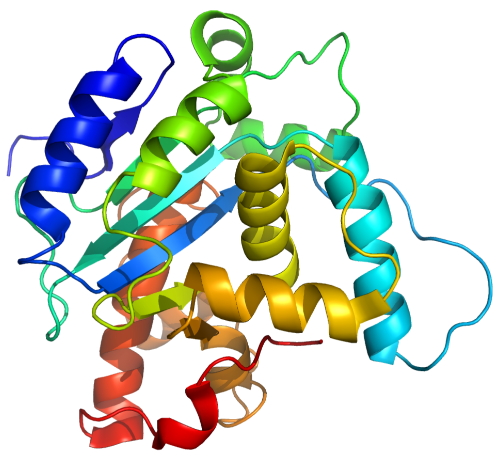
Wander in the Latent Space

Arithmetics of the "smile vector"

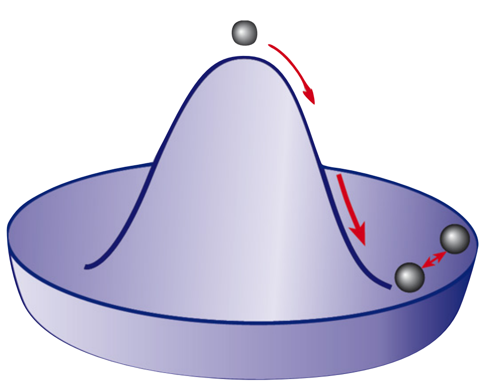


White, 1609.04468 implemented using the variational autoencoder by Kingma and Welling, 1312.6114

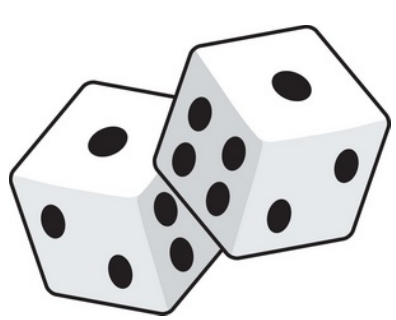
How is it useful ?



Automatically identify collective variables (metadynamics molecular simulation)



Automatically derive effective field theory (free energy surface)



Monte Carlo update proposals

Sampling in the latent space

Change-of-variables in a learnable way

$$Z = \int \mathrm{d}\boldsymbol{x}\, \boxed{\pi(\boldsymbol{x})} = \int \mathrm{d}\boldsymbol{z}\, \boxed{\pi(g(\boldsymbol{z})) \left| \det \left(\frac{\partial \boldsymbol{x}}{\partial \boldsymbol{z}} \right) \right|}$$

Physical
Prob. Dist.

Latent variable
Prob. Dist.

Latent space is less correlated,
therefore, easier to sample

Metropolized Independent Sampler

Acceptance rate with detailed balance condition

$$A(\boldsymbol{x} \to \boldsymbol{x}') = \min\left[1, \underbrace{\boxed{\frac{q(\boldsymbol{x})}{q(\boldsymbol{x}')}}}_{\text{Propose Ratio}} \cdot \underbrace{\boxed{\frac{\pi(\boldsymbol{x}')}{\pi(\boldsymbol{x})}}}_{\text{Physical Probability}}\right]$$

- Unbiased physics even for imperfect proposals
- Proposals are independent

Surrogate energy function:

Li Huang and LW, 1610.02746

Liu, Qi, Meng, Fu, 1610.03137

Trainable transition kernel:

Song, Zhao, Ermon, 1706.07561

Levy, Hoffman, Sohl-Dickstein, 1711.09268

Remarks on TNS Connection

- What we had is a classical downgrade of MERA

(Bény 2013)

Probability Density~ Quantum Wavefuntion

Classical Mutual Information ~ Entanglement Entropy

“Decorrelator" ~ Disentangler

Decimator~Isometry

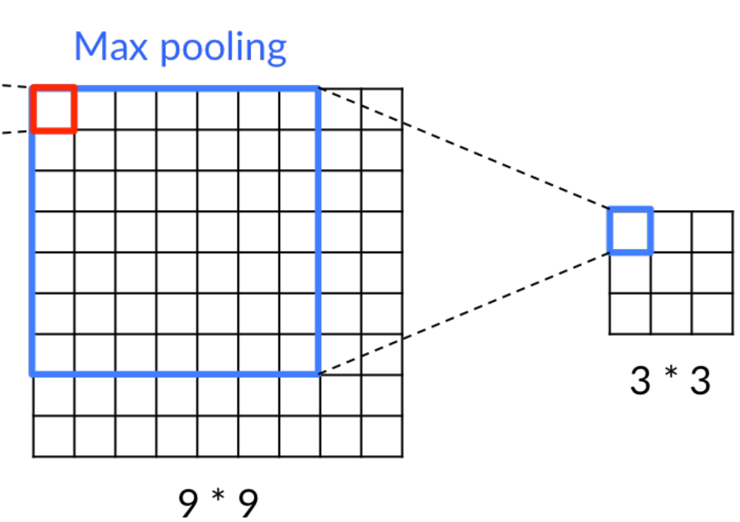
Bijector~Unitary

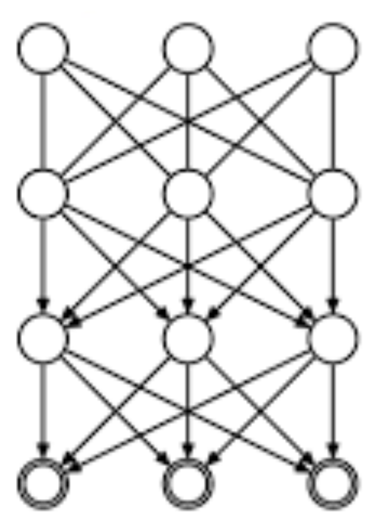
- Deep Learning machinery provides structural flexibility, modular abstraction, end-to-end training

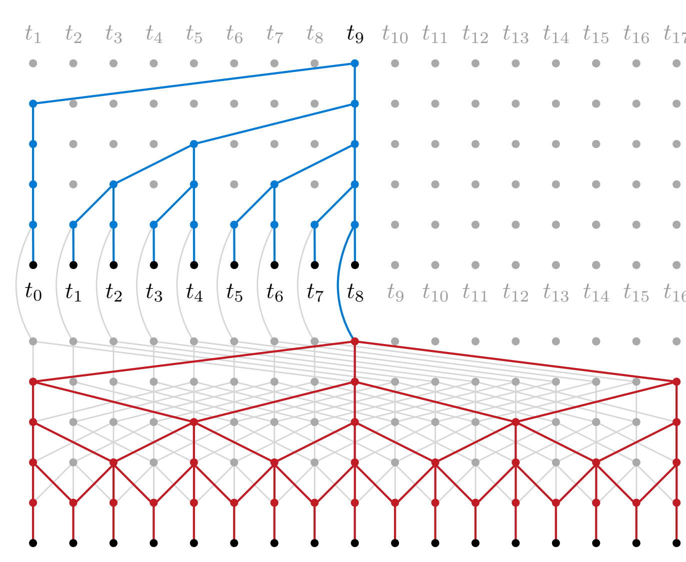
- We give back to DL understandings of what are they doing (and hopefully, how to do better)

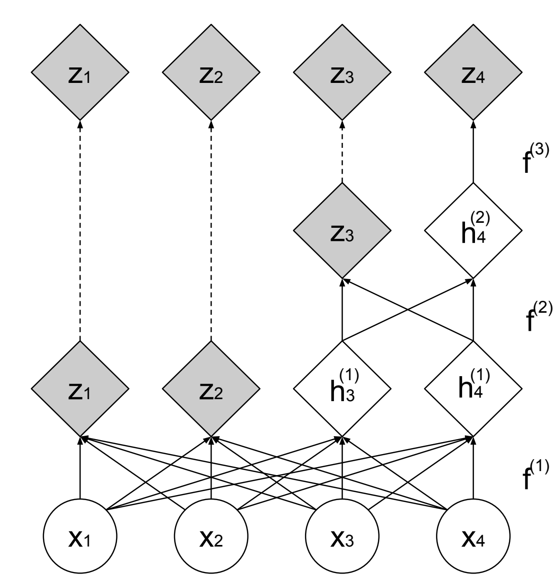
Remarks on DL

Old Wisdoms	**New Insights**
Pooling layer in ConvNets ~ Decimation	Dialed convolution + Factor out layers = Decimation
Hidden nodes in RBM/DBN ~ Renormalized Variables	Kept latent variables = Renormalized Variables



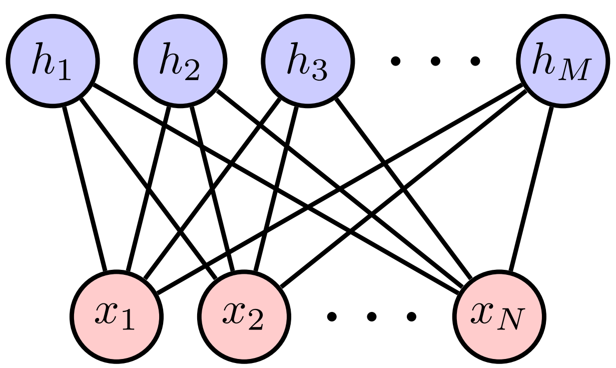


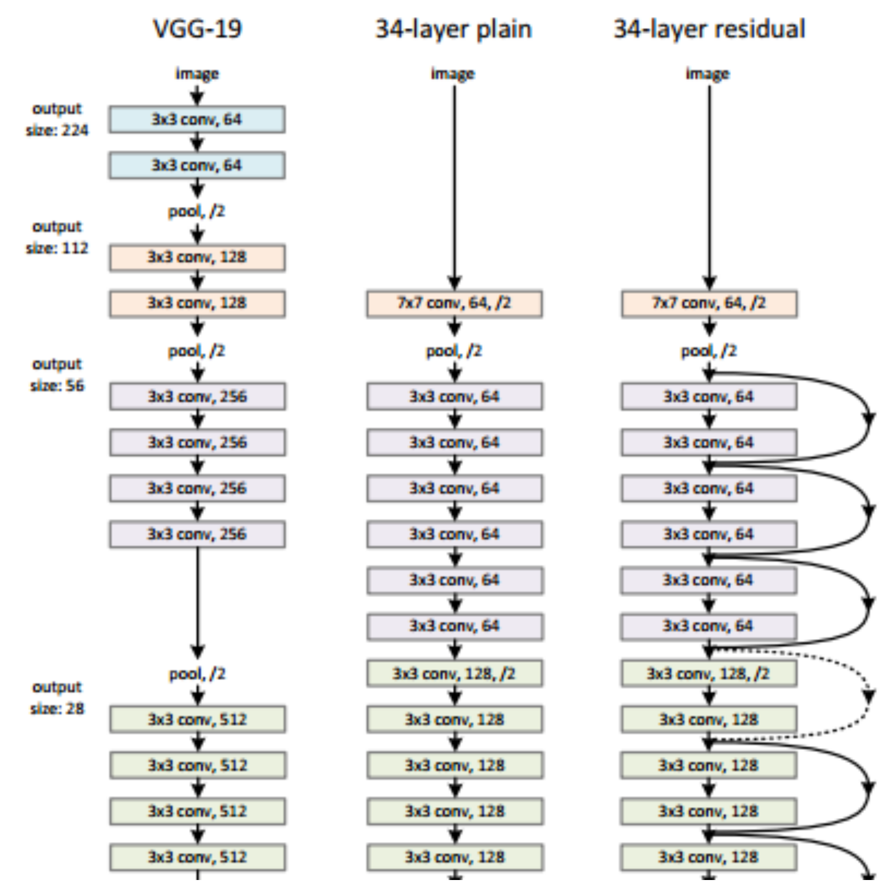




Spherical chicken
in vacuum

Animals
in the wild

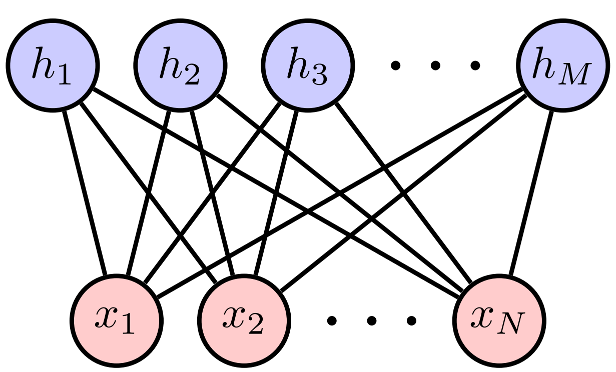


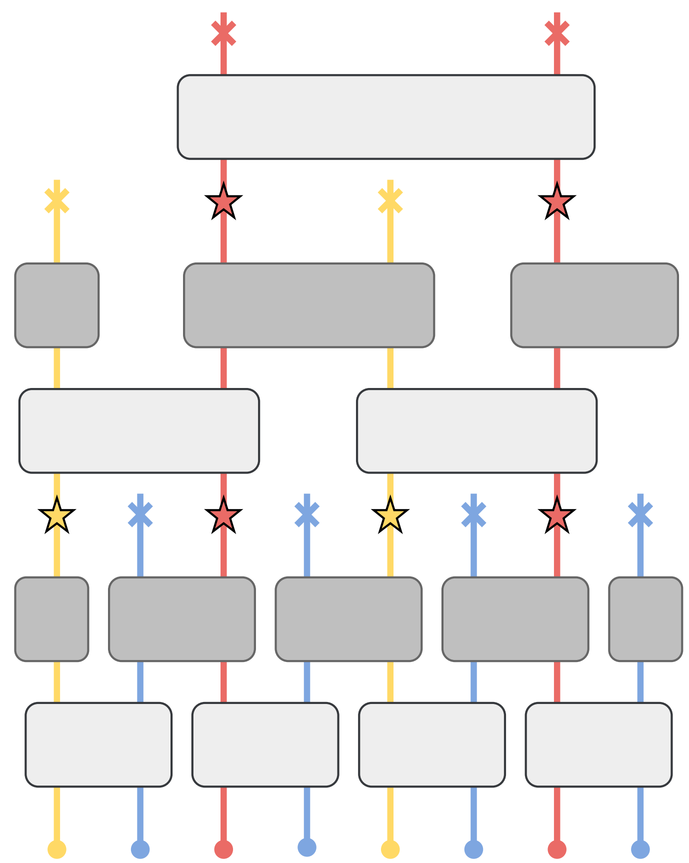


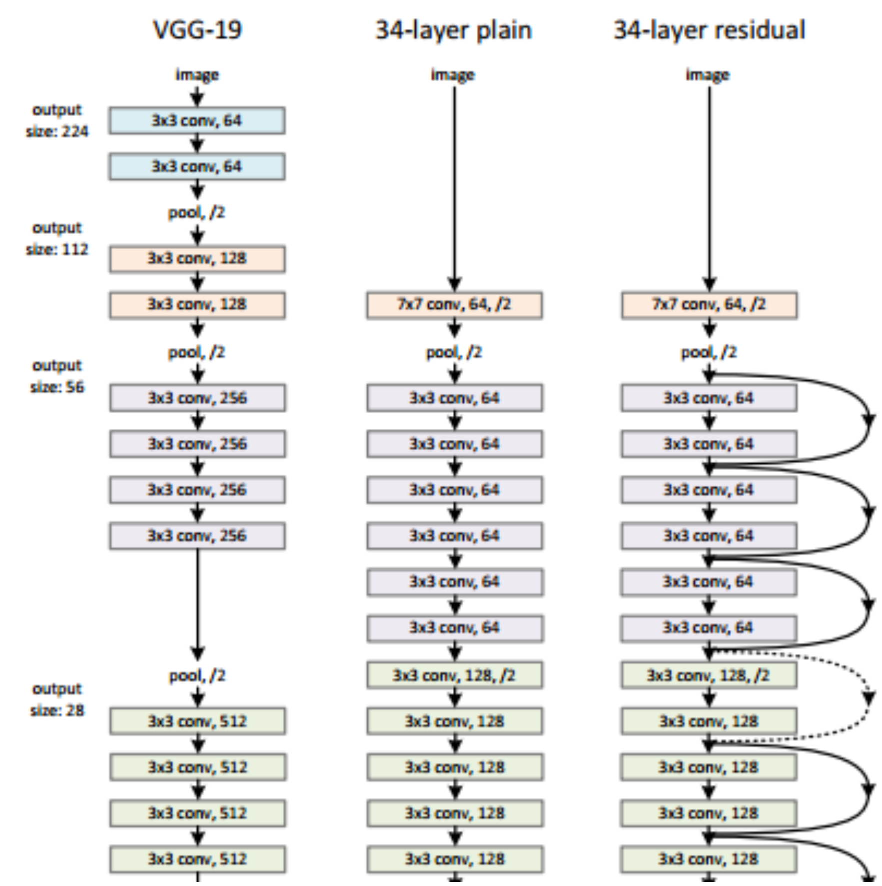
Here we are

Spherical chicken in vacuum

Animals in the wild







Simplified, but not oversimplified model with balanced interpretability and expressibility

Remarks on RG

- Conventionally, RG is a semi-group, not a group

- NeuralRG builds on bijectors, hence a group (coarse-graining due to the multiscale structure)

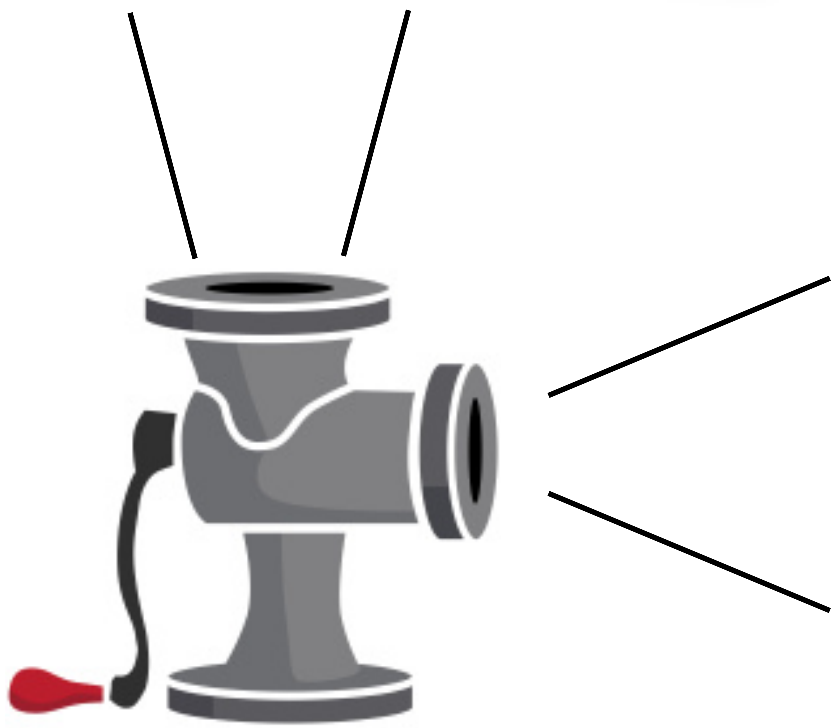
- Probabilistic (Jona-Lasinio 75') and Information Theory (Apenko 09') views on RG (same is true for neural & tensor networks)

- Diffeomorphism does not change topology of the manifolds, therefore, may be limited.

The Universe as a Generative Model

$$\mathcal{L} = \int d^4x \sqrt{-g} \Big[\frac{m_p^2}{2} R - \frac{1}{4} F^a_{\mu\nu} F_a^{\mu\nu} + i\bar{\psi}^i \gamma^\mu D_\mu \psi^i + \left(\bar{\psi}^i_L V_{ij} \Phi \psi^j_R + h.c. \right) - |D_\mu \Phi|^2 - V(\Phi) \Big]$$

RG = Infer the Effective Field Theory



Thank you!