

从伊辛模型到神经网络：2024物理诺奖解读

1933年生
1958 物理博士
1959至今：贝尔实验室、伯克利、加州理工、普林斯顿（系统生物学教授）
1969 巴克利奖
2001 狄拉克奖
2022 玻尔兹曼奖
2024 物理诺贝尔奖



Ill. Niklas Elmehed © Nobel Prize Outreach
John J. Hopfield



Ill. Niklas Elmehed © Nobel Prize Outreach
Geoffrey E. Hinton

1947年生，Bool曾孙
1970 试验心理学学士
当了一段时间木匠
1978 艾丁顿大学
人工智能博士
1987至今 多伦多大学
2013-2023 Google
2018 图灵奖
2024 物理诺贝尔奖
学生占据AI半壁江山

Pan Zhang
ITP,CAS



The Nobel Prize in Physics 2024

The Royal Swedish Academy of Sciences has decided to award the Nobel Prize in Physics 2024 to

John J. Hopfield

Princeton University, NJ, USA

Geoffrey E. Hinton

University of Toronto, Canada

“for foundational discoveries and inventions that enable machine learning with artificial neural networks”

They trained artificial neural networks using physics

This year’s two Nobel Laureates in Physics have used tools from physics to develop methods that are the foundation of today’s powerful machine learning. John Hopfield created an associative memory that can store and reconstruct images and other types of patterns in data. Geoffrey Hinton invented a method that can autonomously find properties in data, and so perform tasks such as identifying specific elements in pictures.

When we talk about artificial intelligence, we often mean machine learning using artificial neural networks. This technology was originally inspired by the structure of the brain. In an artificial neural network, the brain’s neurons are represented by nodes that have different values. These nodes influence each other through connections that can be likened to synapses and which can be made stronger or weaker. The network is *trained*, for example by developing stronger connections between nodes with simultaneously high values. This year’s laureates have conducted important work with artificial neural networks from the 1980s onward.

John Hopfield invented a network that uses a method for saving and recreating patterns. We can imagine the nodes as pixels. The *Hopfield network* utilises physics that describes a material’s characteristics due to its atomic spin – a property that makes each atom a tiny magnet. The network as a whole is described in a manner equivalent to the energy in the spin system found in physics, and is trained by finding values for the connections between the nodes so that the saved

images have low energy. When the Hopfield network is fed a distorted or incomplete image, it methodically works through the nodes and updates their values so the network’s energy falls. The network thus works stepwise to find the saved image that is most like the imperfect one it was fed with.

Geoffrey Hinton used the Hopfield network as the foundation for a new network that uses a different method: the *Boltzmann machine*. This can learn to recognise characteristic elements in a given type of data. Hinton used tools from statistical physics, the science of systems built from many similar components. The machine is trained by feeding it examples that are very likely to arise when the machine is run. The Boltzmann machine can be used to classify images or create new examples of the type of pattern on which it was trained. Hinton has built upon this work, helping initiate the current explosive development of machine learning.

“The laureates’ work has already been of the greatest benefit. In physics we use artificial neural networks in a vast range of areas, such as developing new materials with specific properties,” says Ellen Moons, Chair of the Nobel Committee for Physics.

John J. Hopfield, born 1933 in Chicago, IL, USA. PhD 1958 from Cornell University, Ithaca, NY, USA. Professor at Princeton University, NJ, USA.

Geoffrey E. Hinton, born 1947 in London, UK. PhD 1978 from The University of Edinburgh, UK. Professor at University of Toronto, Canada.

Hopfield network

Boltzmann machine

无监督学习 (生成学习)

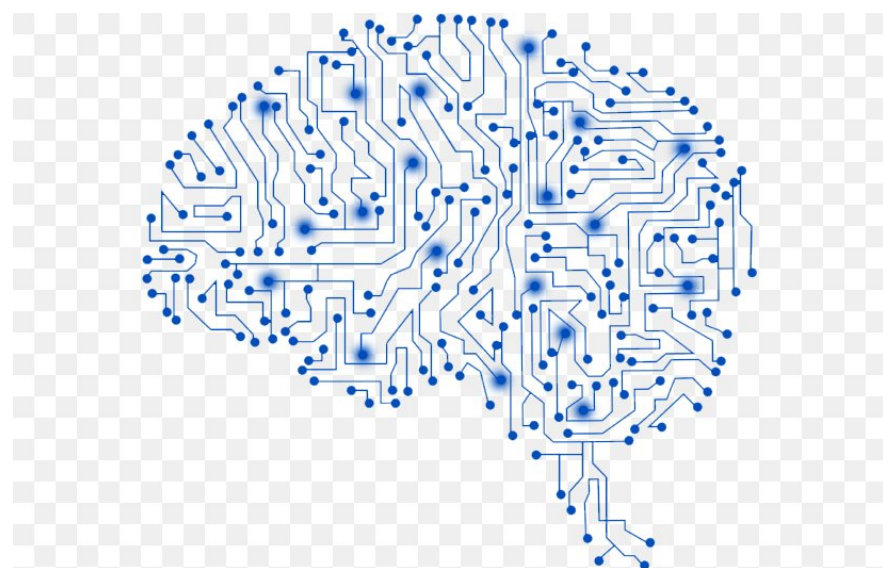


4	1	9	2	1	3
3	5	3	6	1	7
6	9	4	0	9	1
4	3	2	7	3	8
0	5	6	0	7	6
1	9	3	9	8	5

床前明月光，
疑是地上霜。
举头望明月，
低头思故乡。



训练



$P(X)$

9	3	6	6	5	7
5	3	9	4	4	7
5	4	1	2	6	0
7	4	6	2	2	2
2	9	8	9	3	9
2	0	6	7	1	9

千山鸟飞绝，
万径人踪灭。
孤舟蓑笠翁，
独钓寒江雪。



↑ ↑ ↑ ↓ ↑ ↓ ↓ ↓ ↓ ↓ ↑ ↑

↑ ↑ ↑ ↓ ↑ ↓ ↓ ↓ ↓ ↓ ↑ ↑

The Forward-Forward Algorithm: Some Preliminary Investigations

Geoffrey Hinton
Google Brain
geoffhinton@google.com

2022年NeurIPS
Hinton 75岁

Abstract

The aim of this paper is to introduce a new learning procedure for neural networks and to demonstrate that it works well enough on a few small problems to be worth

In the early 1980s there were two promising learning procedures for deep neural networks. One was backpropagation and the other was Boltzmann Machines (Hinton and Sejnowski, 1986) which performed unsupervised contrastive learning. A Boltzmann Machine is a network of stochastic binary

The Boltzmann machine can be seen as a combination of two ideas:

1. Learn by minimizing the free energy on real data and maximizing the free energy on negative data generated by the network itself.
2. Use the Hopfield energy as the energy function and use repeated stochastic updates to sample global configurations from the Boltzmann distribution defined by the energy function.

大自然的分布与采样



Low temperature

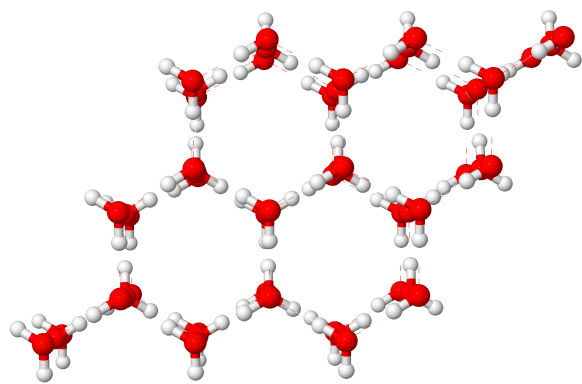


High temperature

大自然的分布与采样

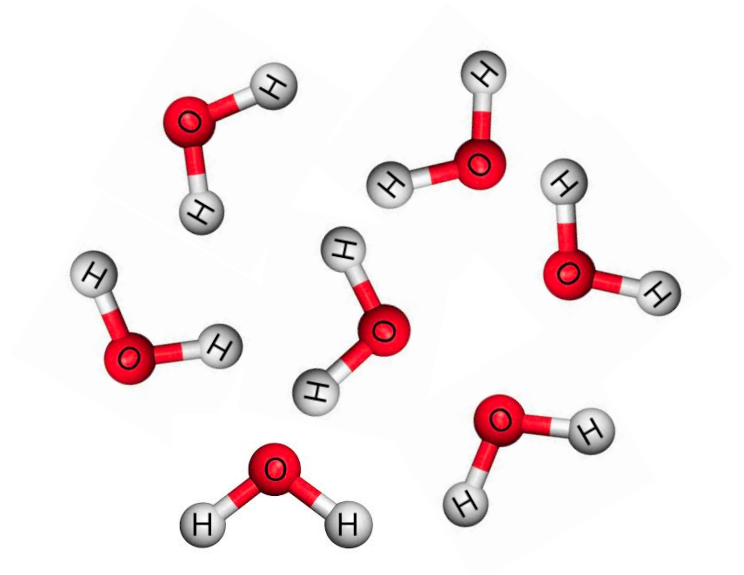


Low temperature



molecules are the same

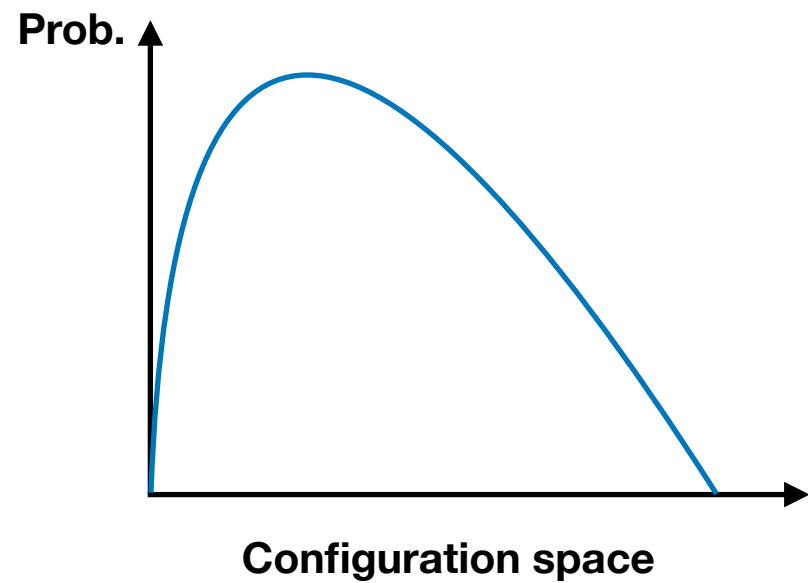
High temperature



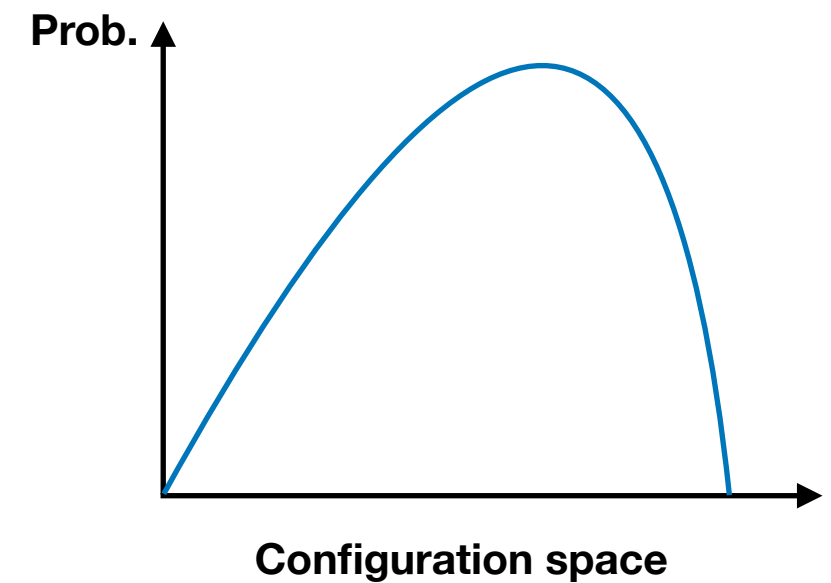
大自然的分布与采样



Low temperature



High temperature

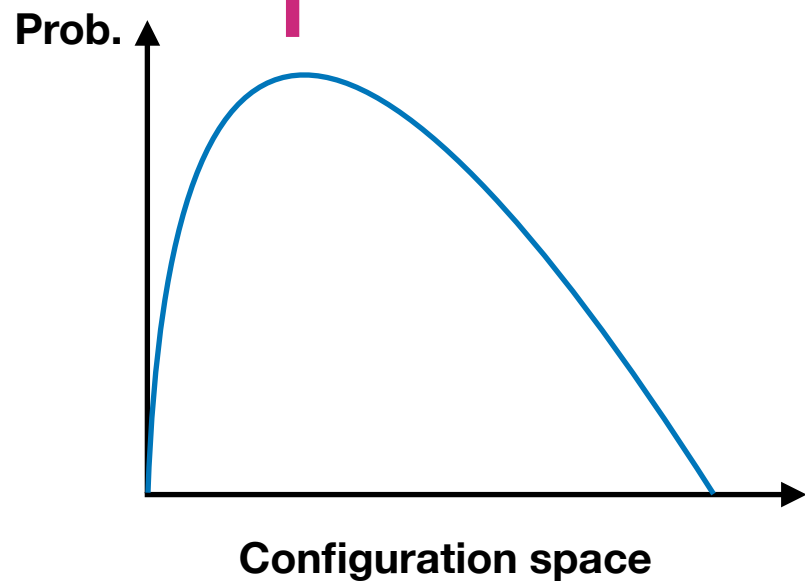


molecules are the same, but *distributions* are different

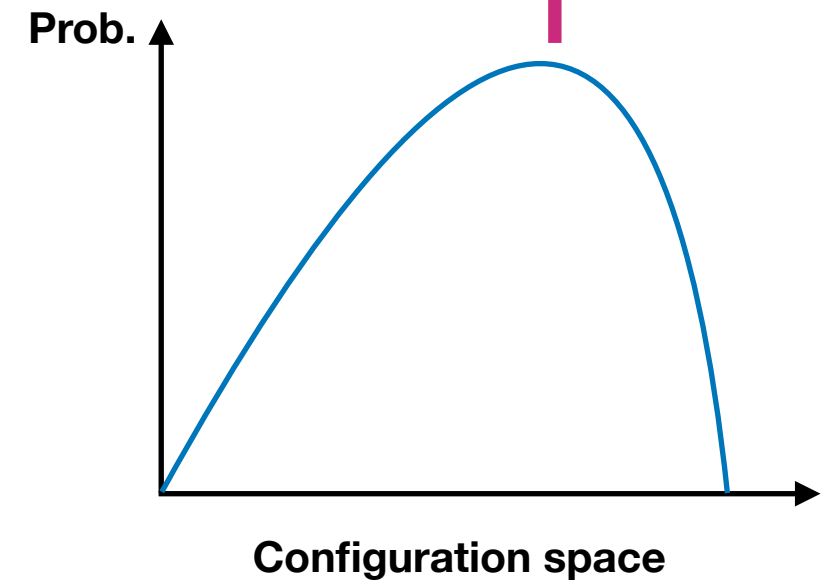
大自然的分布与采样



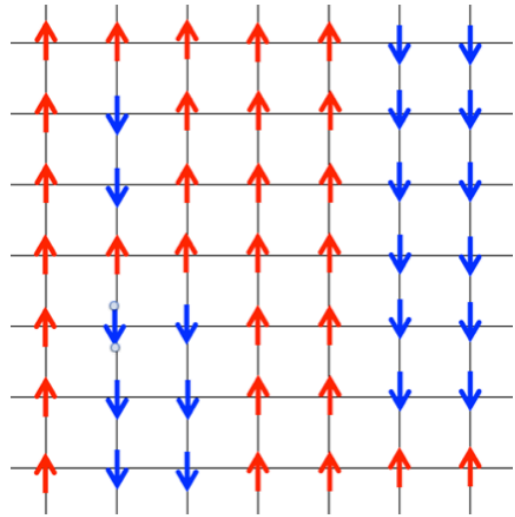
采样



采样



统计物理与机器学习



Joint distribution of micro-configurations

$$P(\sigma) = \frac{1}{Z} \exp(-\beta E(\sigma))$$



Joint distribution of data variables

$$P(\text{Data})$$

Exponential-large space
Efficient methods
Computational power

AN INTRODUCTION TO LEARNING AND GENERALISATION



Giorgio Parisi

Dipartimento di Fisica

Piazzale delle Scienze

Roma Italy 00185

ABSTRACT. In this lecture I will present some basic ideas on how computers may learn rules from examples and how generalisation may be achieved. The general prospective is presented. Some comments are also done on the definition of intelligence.

Learning – Generalisation – Intelligence

Giorgio Parisi 1992'

Boltzmann Medal, Lars Onsager Medal, Dirac Medal, Nobel Prize

Statistical Physics (machine learning related)

Machine Learning (neural network related)

19th century

Boltzmann, Clausius, Maxwell,
Statistical Mechanics

Building of statistical methods
Least squares, Bayes rule

19th century

1895'
1920'
1935'
1944'
1953'
1957'

Curie-Weiss mean-field
Ising model
Bethe Approximation
Onsager's solution to 2D Ising model
Metropolis, MCMC
Jaynes, Maximum Entropy principle

Markov Chain

1913'
1948'
1950'
1957'
1969'

Shannon, Information Theory
Turin's learning machine
Perceptron
Minsky and Papert,
Limitations of perceptron

1975'
1978'
1979'
1980'
1982'
1982'
1985'

Edward and Anderson spin-glass model
Sherrington-Kirkpatrick spin-glass model
Parisi' replica symmetry breaking
Nishimori line
Hopfield neural networks
Simulated annealing
Amit-Gutfreund-Sompolinsky
Phase diagram of the Hopfield NN
Gardner and Derrida, capacity of perceptron

Necognitron,
early convolution neural networks
Hopfield neural networks
Boltzmann Machine
Backpropagation
Reinforcement learning

1980'
1982'
1985'
1986'
1989'

1989'

Krauth and Mezard
Capacity of binary perceptrons

Support Vector Machine

1995'

1990 - 2009'

Applications of spin glass theory in
optimizations, Mean-field approximations

IBM Deep Blue
Variational inference

1997'
2000'

Image Net

2009'

AlexNet convolution networks

2012'

2009 - now

message passings algorithms for
inference and learning

Applications of Deep Neural Networks
Alpha Go, alphaFold, chatGPT

2012 - now

4.4. NEURAL NETWORKS

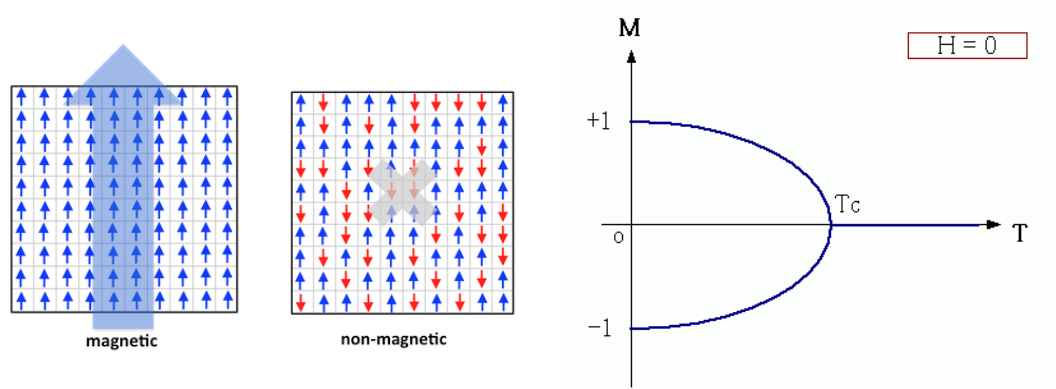
One wide ranging development, in the statistical physics of neural networks, has been the so-called Gardner approach, namely a statistical analysis in parameter space, i.e. the space of interactions (e.g. synaptic weights). It has been called the inverse problem of statistical mechanics, because in ordinary statistical mechanics the interactions are given and the statistical analysis is done in variable space (e.g. the space of neural activities). At this point,

Gerard Toulouse 1992'

Langevin Prize, Holweck Prize

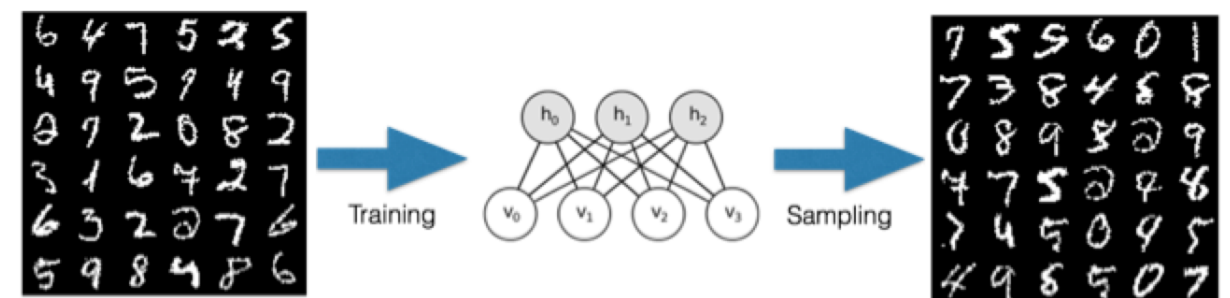
The Ising model
(*Ising, 1924*)

Ordinary Statistical Mechanics

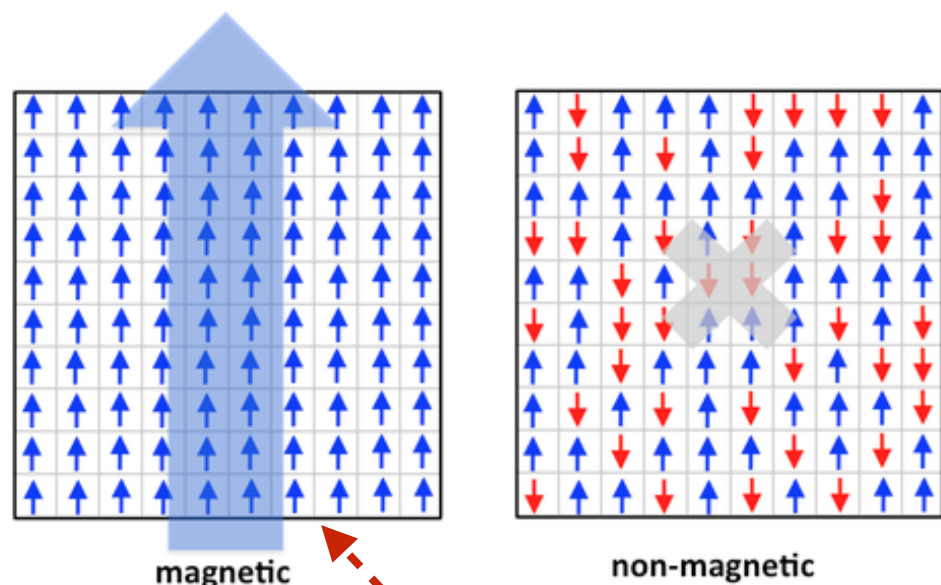


Restricted Boltzmann Machine
(*Ackley, **Hinton**, Sejnowski, 1985*)

Inverse problem of statistical mechanics

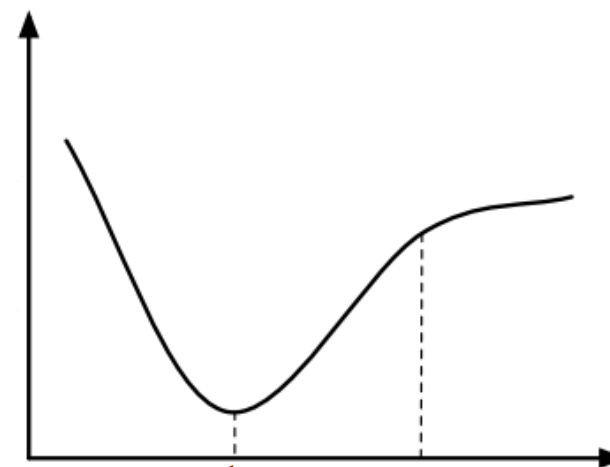
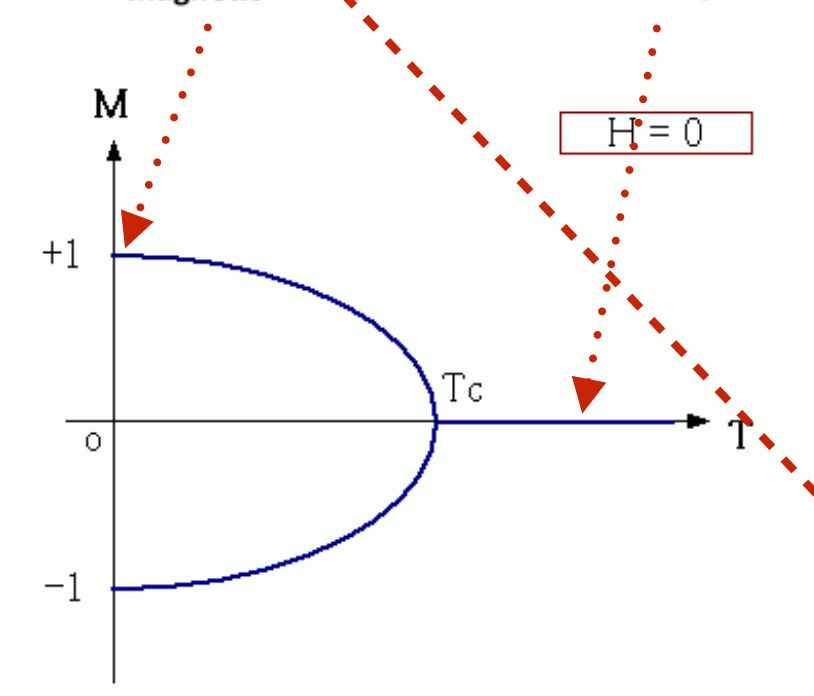


伊辛模型 (Ising, 1924)



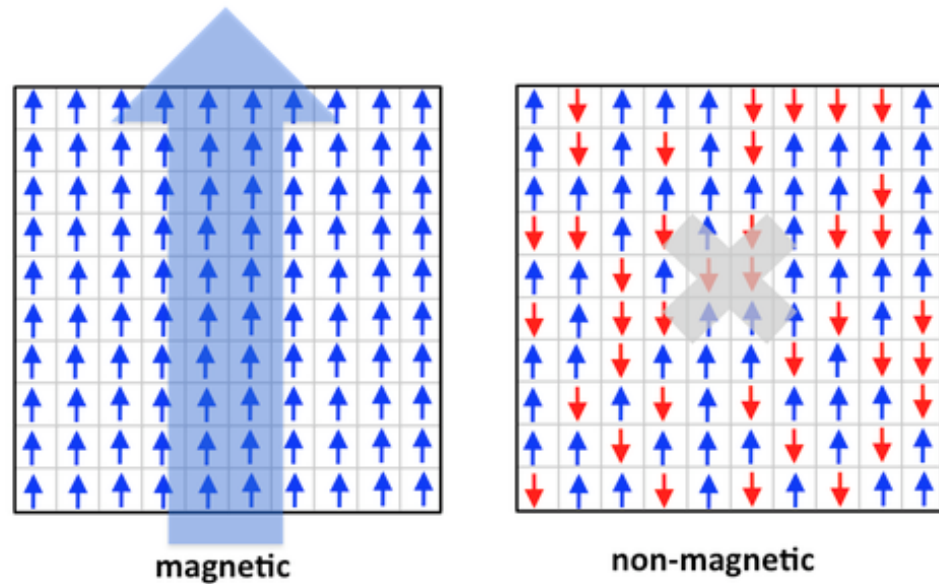
$$S = \{+1, -1\}^n \quad \uparrow \uparrow \uparrow \downarrow \uparrow \downarrow \downarrow \downarrow \downarrow \downarrow \uparrow \uparrow$$

$$P(S) = \frac{1}{Z} e^{\beta \sum_{\langle ij \rangle} J_{ij} S_i S_j} \quad J_{ij} = 1$$



- 高温：记不住任何数据
- 低温：记住一个数据（全黑或者全白）！

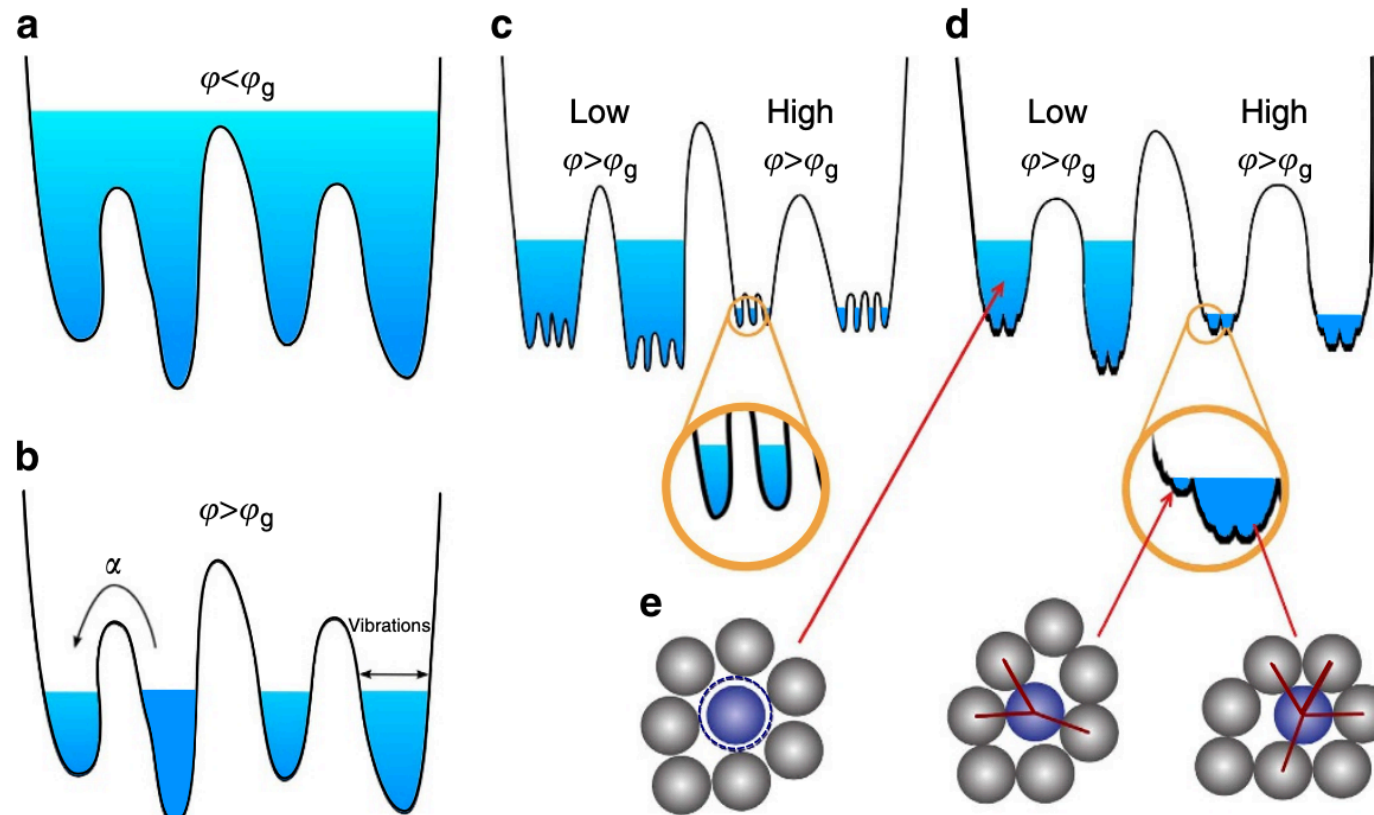
自旋玻璃模型 (忘掉所有的数据)



$$S = \{+1, -1\}^n \quad \uparrow \uparrow \uparrow \downarrow \uparrow \downarrow \downarrow \downarrow \downarrow \downarrow \uparrow \uparrow$$

$$P(S) = \frac{1}{Z} e^{\beta \sum_{(ij)} J_{ij} S_i S_j} \quad J_{ij} \sim \mathcal{N}(0, 1/n)$$

Sherrington-Kirkpatrick 1975'
Parisi (full RSB solution) 1979'



Hopfield model: 记住多个数据 (Hopfield, 1982)



Associative memory (Hopfield, 1982, Amari 1977, Little 1974)

$$P(S) = \frac{1}{Z} e^{\beta \sum_{(ij)} J_{ij} S_i S_j} \quad \left\{ \xi_i^\mu \right\} \in \{+1, -1\}^{\alpha n \times n}$$

$$J_{ij} = \frac{1}{\alpha n} \sum_{\mu=1}^{\alpha n} \xi_i^\mu \xi_j^\mu$$

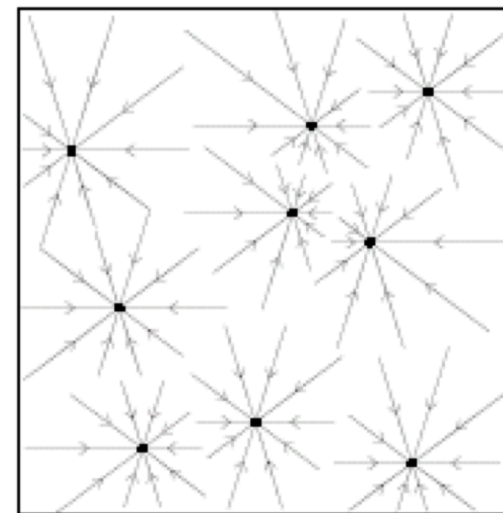
Hebb's learning rule
Hebb 1949

网络更新规则: Glauber dynamics (Glauber 1963)

$$P(S_i) = \frac{e^{\sum_{j \neq i} J_{ij} S_i S_j}}{2 \cosh(\sum_{j \neq i} J_{ij} S_i S_j)} \propto e^{\beta \frac{1}{\alpha n} \sum_{j \neq i} \sum_{\mu} \xi_i^\mu \xi_j^\mu S_i S_j}$$

如果网络状态是第一个数据, $S_i = \xi_i^1$

$$\begin{aligned} P(S_i) &\propto e^{\frac{1}{\alpha n} \sum_{j \neq i} \xi_i^1 \xi_j^1 S_i S_j + \sum_{\mu \neq 1} \sum_{j \neq i} \xi_i^\mu \xi_j^\mu S_i S_j} \\ &= e^{\frac{1}{\alpha n} \sum_{j \neq i} 1 + \sum_{\mu \neq 1} \sum_{j \neq i} \xi_i^\mu \xi_j^\mu S_i S_j} \end{aligned}$$



数据被存储为不动点

Hopfield model: 记住多个数据 (Hopfield, 1982)

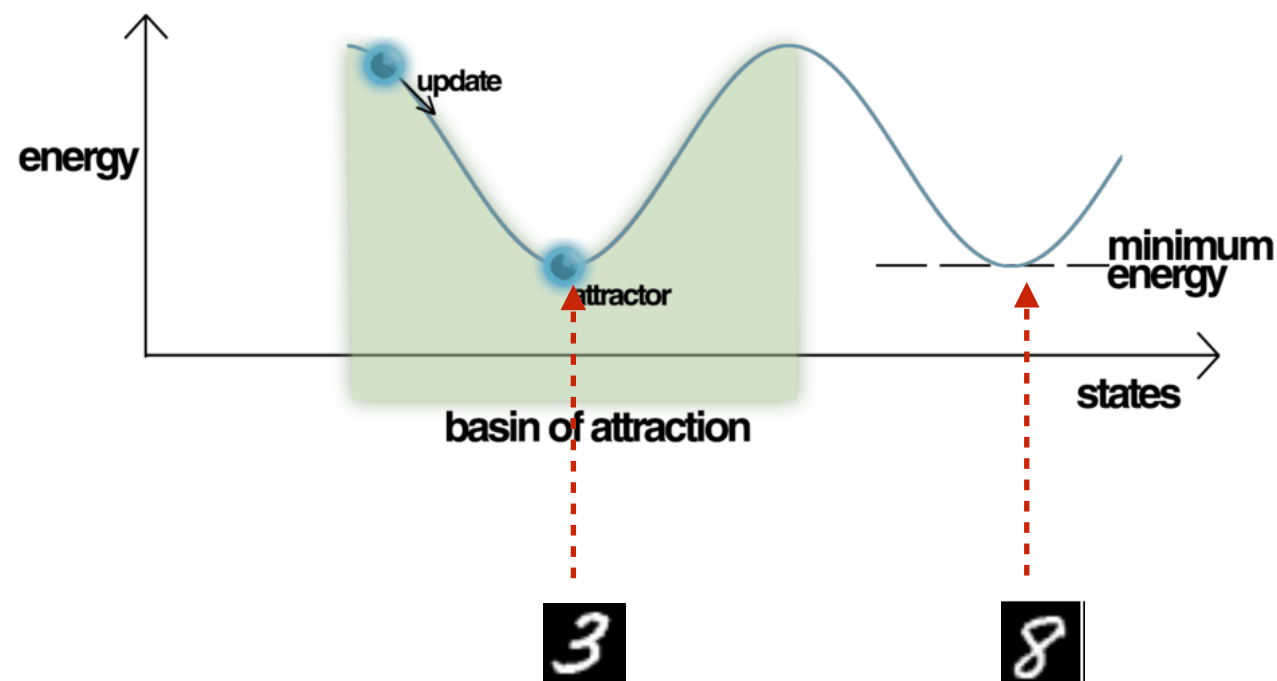


Associative memory (Hopfield, 1982, Amari 1977, Little 1974)

$$P(S) = \frac{1}{Z} e^{\beta \sum_{(ij)} J_{ij} S_i S_j} \quad \left\{ \begin{matrix} \xi_i^\mu \\ \xi_j^\mu \end{matrix} \right\} \in \{+1, -1\}^{\alpha n \times n}$$

$$J_{ij} = \frac{1}{\alpha n} \sum_{\mu=1}^{\alpha n} \xi_i^\mu \xi_j^\mu$$

Hebb's learning rule
Hebb 1949



Hopfield model的统计物理理论



Associative memory (Hopfield, 1982, Amari 1977, Little 1974)

$$P(S) = \frac{1}{Z} e^{\beta \sum_{(ij)} J_{ij} S_i S_j} \quad \left\{ \begin{matrix} \xi^\mu \\ \zeta_i \end{matrix} \right\} \in \{+1, -1\}^{\alpha n \times n}$$

$$J_{ij} = \frac{1}{\alpha n} \sum_{\mu=1}^{\alpha n} \xi_i^\mu \xi_j^\mu$$

Hebb's learning rule
Hebb 1949

VOLUME 55, NUMBER 14 PHYSICAL REVIEW LETTERS 30 SEPTEMBER 1985

Storing Infinite Numbers of Patterns in a Spin-Glass Model of Neural Networks

Daniel J. Amit and Hanoach Gutfreund
Racah Institute of Physics, Hebrew University, Jerusalem 91904, Israel

and

H. Sompolinsky
Department of Physics, Bar Ilan University, Ramat Gan, Israel
(Received 11 July 1985)

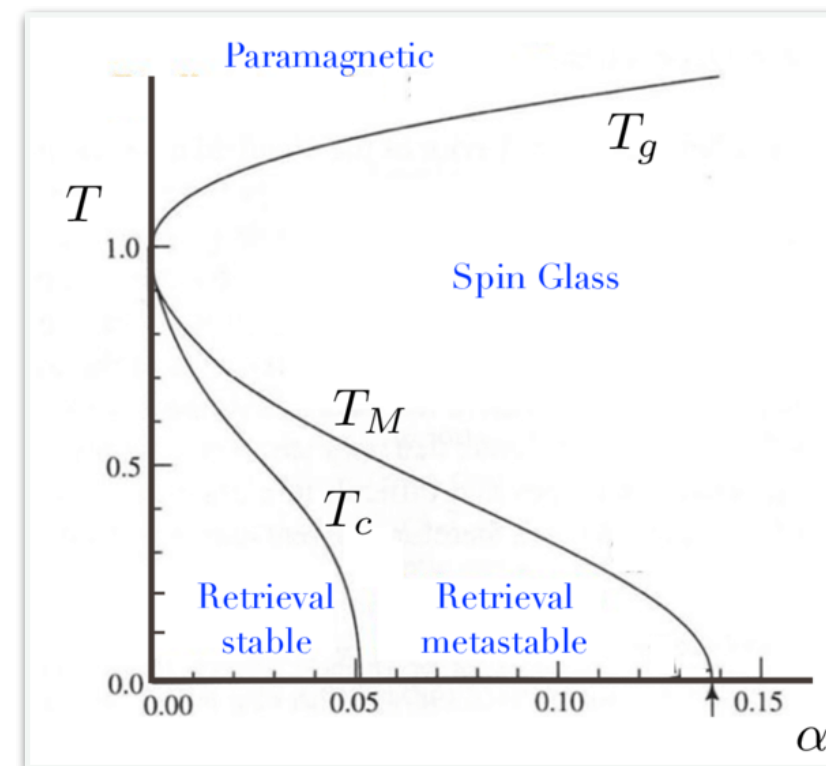
VOLUME 71, NUMBER 23 PHYSICAL REVIEW LETTERS 6 DECEMBER 1993

Dynamics of Fully Connected Attractor Neural Networks near Saturation

A. C. C. Coolen and D. Sherrington

Department of Physics-Theoretical Physics, University of Oxford, 1 Keble Road, Oxford OX1 3NP, United Kingdom
(Received 16 August 1993)

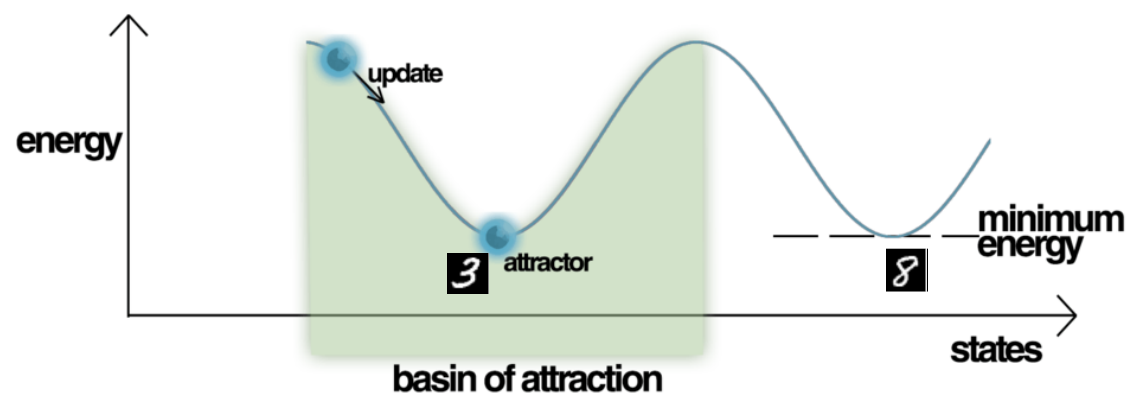
We present an exact dynamical theory, valid on finite time scales, to describe the fully connected Hopfield model near saturation in terms of deterministic flow equations for order parameters. Two transparent assumptions allow us to perform a replica calculation of the distribution of intrinsic noise components of the alignment fields. Numerical simulations indicate that our equations describe the dynamics correctly in the region where replica symmetry is stable. In equilibrium our theory reproduces



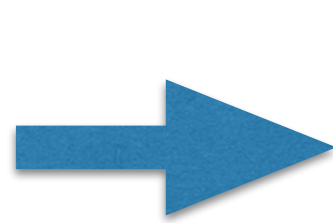
局限性：数据需要是正交的，线性capacity

Phase diagram
Amit, Gutfreund, Sompolinsky 1985

Inverse-Ising model: 根据数据学习权重

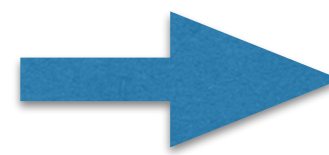


6	1	9	4	2	5
7	8	7	1	3	0
0	7	2	4	8	0
8	4	5	3	0	7
6	9	8	4	5	8
7	7	3	6	8	2



改变 J_{ij}

Ising Model



采样

6	1	9	4	2	5
7	8	7	1	3	0
0	7	2	4	8	0
8	4	5	3	0	7
6	9	8	4	5	8
7	7	3	6	8	2

$$P(S) = \frac{1}{Z} e^{\beta \sum_{(ij)} J_{ij} S_i S_j}$$

Inverse-Ising model: 根据数据学习权重

$$P(S) = \frac{1}{Z} e^{\beta \sum_{(ij)} J_{ij} S_i S_j}$$

- 全连接自旋玻璃模型

➔ 需要学习所有的 J_{ij}

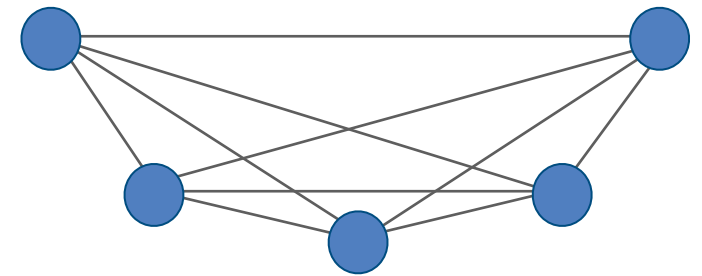
- 给定数据一阶矩和二阶矩的最大熵模型

$$P(s) = \arg \max H(P), \text{ s.t. } \langle s_i \rangle_P = m_i, \langle s_i s_j \rangle_P = \langle S_i S_j \rangle_{\text{data}}$$

- 指数函数族

➔ Sufficient Statistics 是一阶矩和二阶矩

$$\text{➔ } \frac{\partial \log P}{\partial J_{ij}} = \langle S_i S_j \rangle_{\text{data}} - \langle S_i S_j \rangle_P$$



缺点：参数少，表达能力不够

玻尔兹曼机：通过隐变量增加表述能力

$$P(v, h) = \frac{1}{Z} e^{\sum_{(ij)} J_{ij} v_i v_j + \sum_{ab} J_{ab} h_a h_b + \sum_{ia} W_{ia} v_i h_a}$$

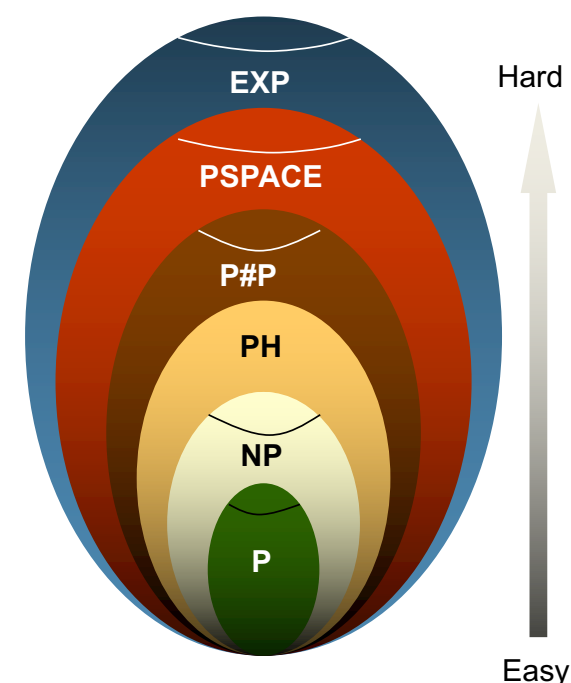
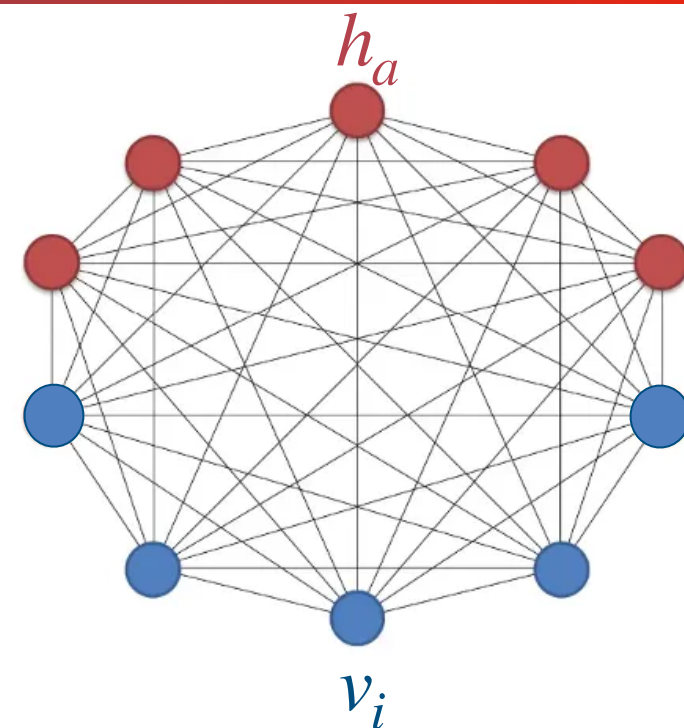
$$P(v) = \frac{1}{Z} \sum_h e^{\sum_{(ij)} J_{ij} v_i v_j + \sum_{ab} J_{ab} h_a h_b + \sum_{ia} W_{ia} v_i h_a}$$

- 玻尔兹曼机是带有隐变量的Inverse Ising model (Ackley, **Hinton**, Sejnowski, 1985)

- 增加参数数目，模型表述能力
- 表达数据中的高阶关联

- 缺点：难以训练

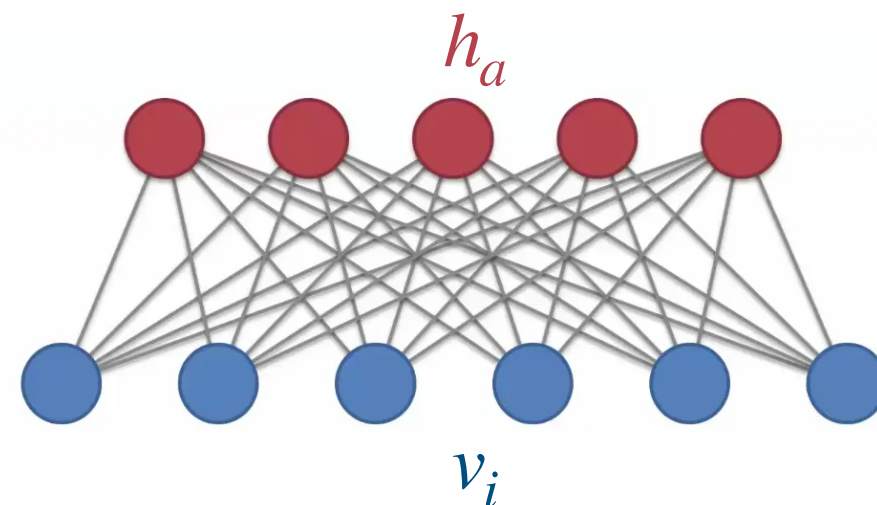
$$-\frac{\partial \log P}{\partial W_{ia}} = \langle v_i h_a \rangle_{\text{data+model}} - \langle v_i h_a \rangle_{\text{model}}$$



受限玻尔兹曼机 (RBM): 有效的训练算法

$$P(v, h) = \frac{1}{Z} e^{\sum_{ia} W_{ia} v_i h_a}$$

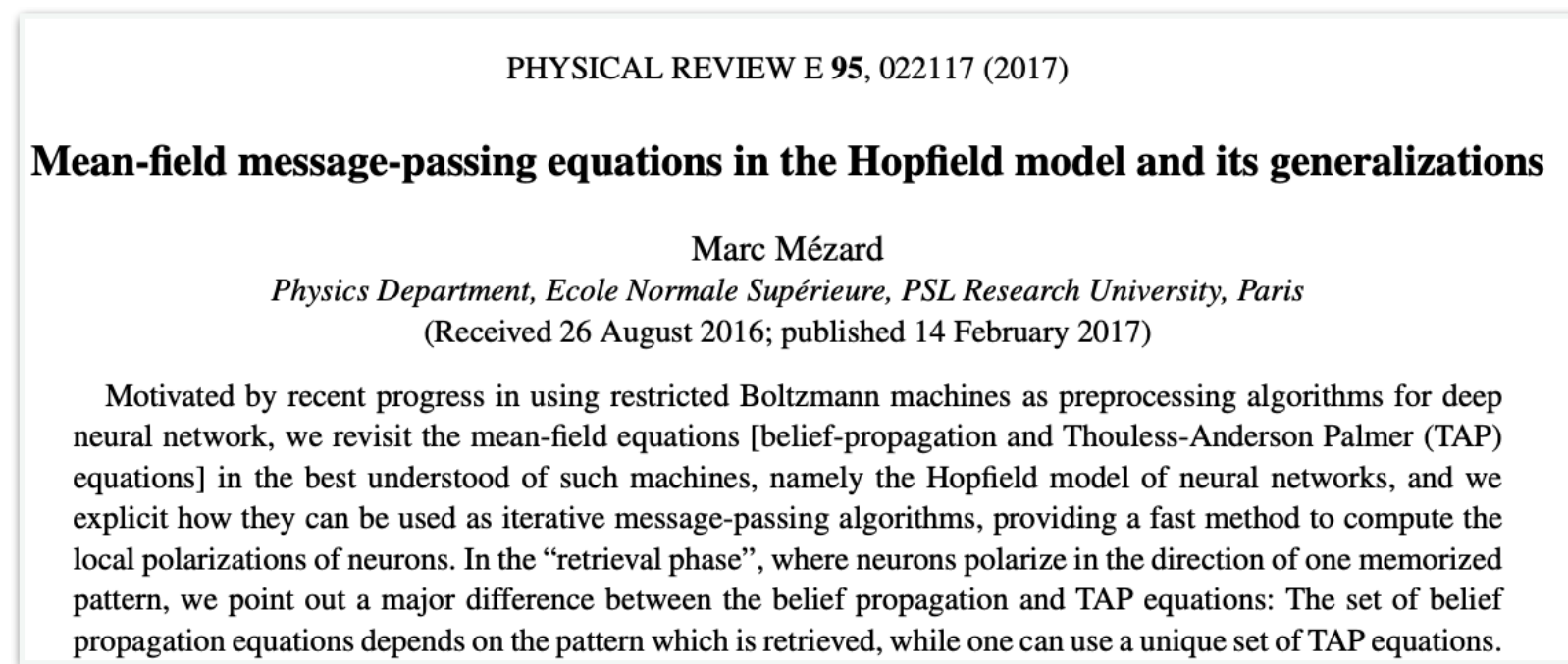
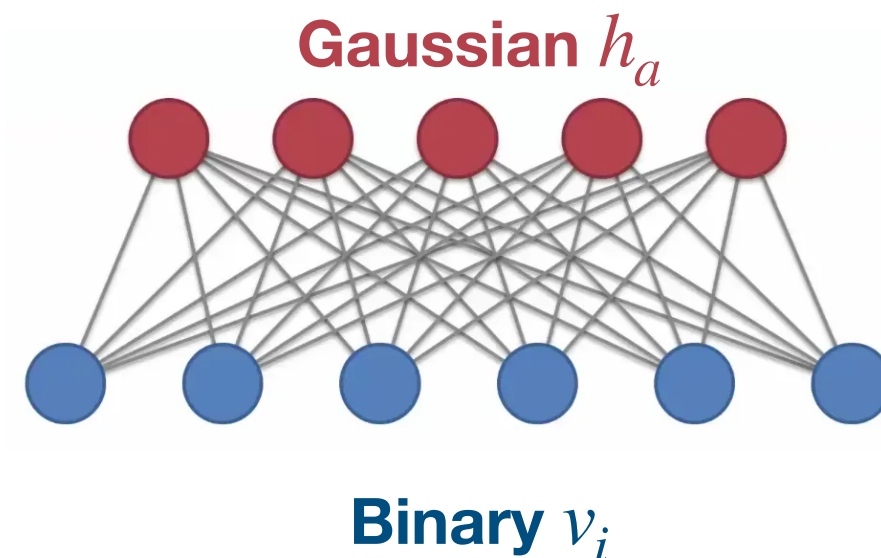
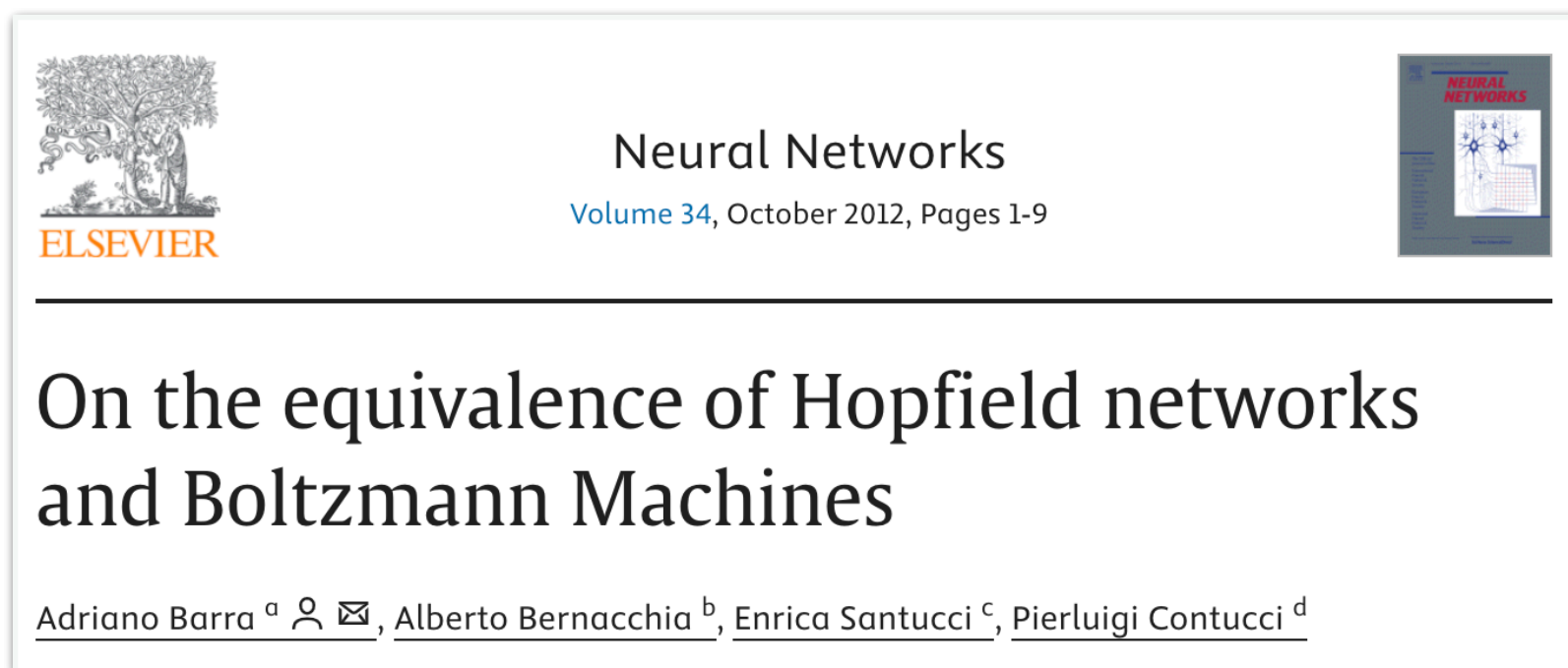
$$P(v) = \frac{1}{Z} \sum_h e^{\sum_{ia} W_{ia} v_i h_a}$$



- RBM是二分图上的玻尔兹曼机 (*Hinton, Sejnowski, 1986*)
 - ➔ 只有隐变量和显变量之间的连接
 - ➔ 数据中的高阶关联通过隐变量诱导
 - ➔ Contrastive Divergence算法可有效计算梯度 (*Hinton 2002*)

$$\frac{\partial \log P}{\partial W_{ia}} = \underbrace{\langle v_i h_a \rangle_{\text{data+model}}}_{\text{减小数据能量}} - \underbrace{\langle v_i h_a \rangle_{\text{model}}}_{\text{增加其他构型能量}}$$

奇怪的知识：Hopfield model等价于Gaussian RBM

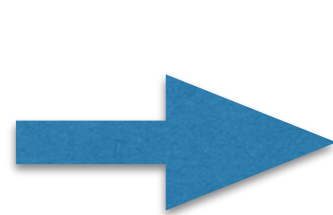
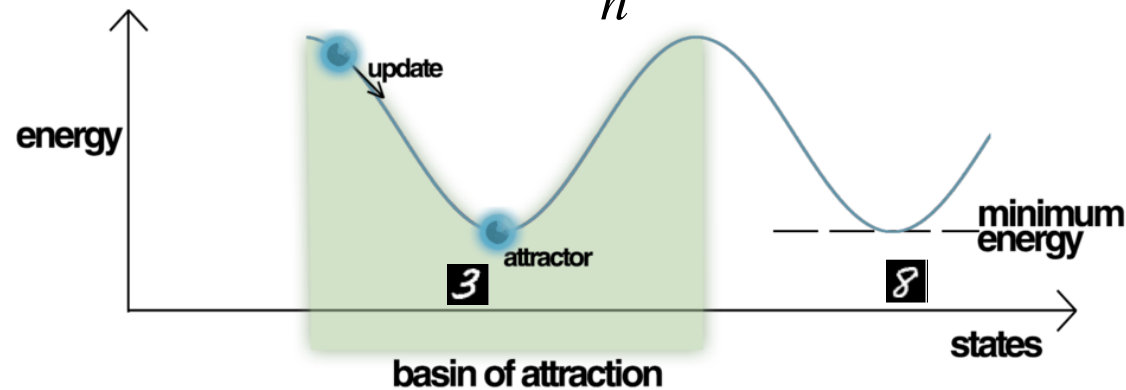


$$Z = \sum_s \int \prod_{\mu} \frac{d\lambda_{\mu}}{\sqrt{2\pi/\beta}} \times \exp \left[-\frac{\beta}{2} \sum_{\mu} \lambda_{\mu}^2 + \beta \sum_{\mu,i} \frac{\xi_i^{\mu}}{\sqrt{N}} s_i \lambda_{\mu} \right].$$

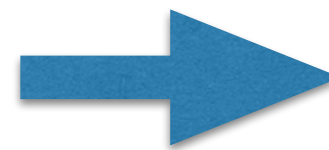
Hubbard Stratonovich 变换

用RBM学习数据分布

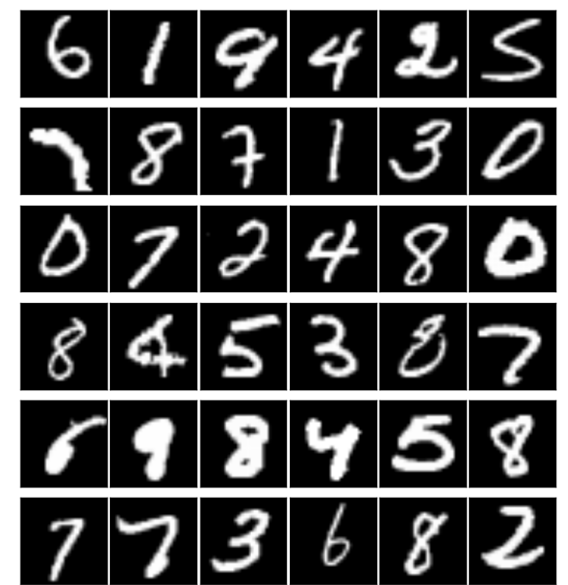
$$P(v) = \frac{1}{Z} \sum_h e^{\sum_{ia} W_{ia} v_i h_a}$$



RBM

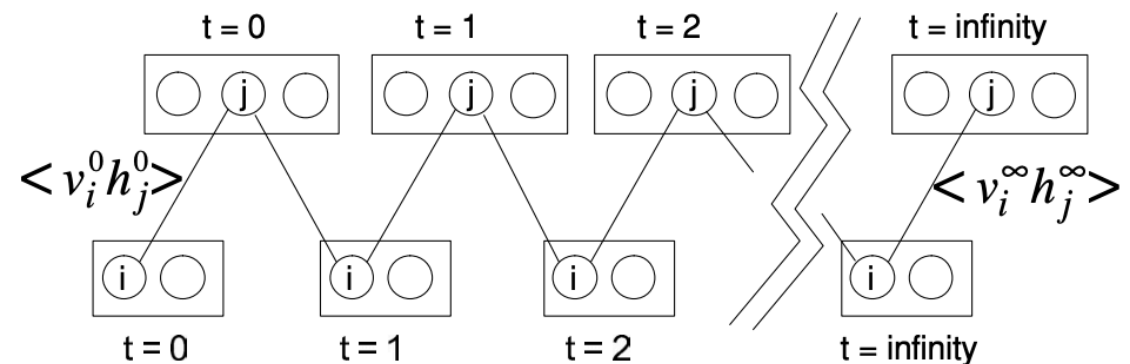


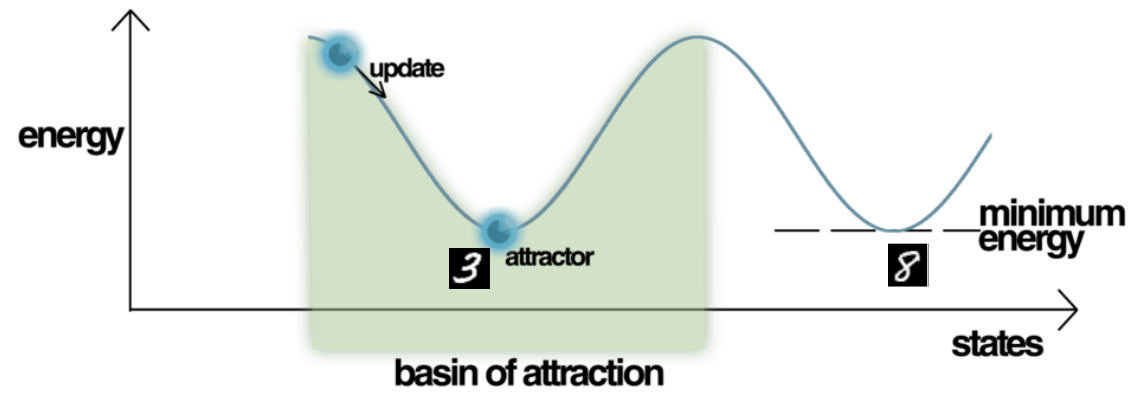
采样



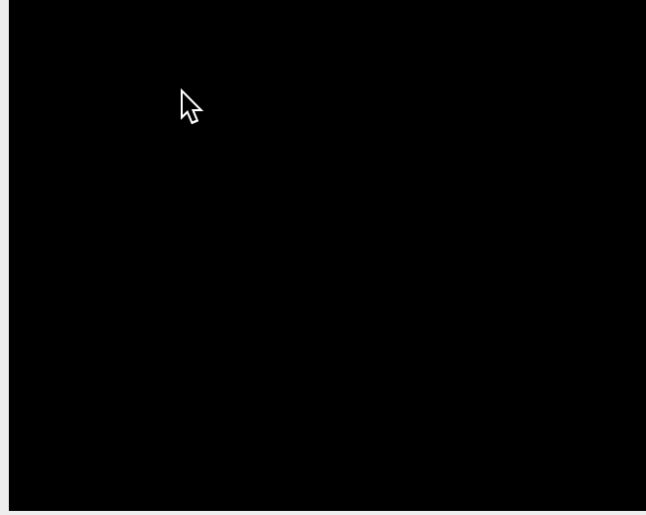
改变 W_{ia}

$$\frac{\partial \log P}{\partial W_{ia}} = \underbrace{\langle v_i h_a \rangle_{\text{data+model}}}_{\text{减小数据能量}} - \underbrace{\langle v_i h_a \rangle_{\text{model}}}_{\text{增加其他构型能量}}$$

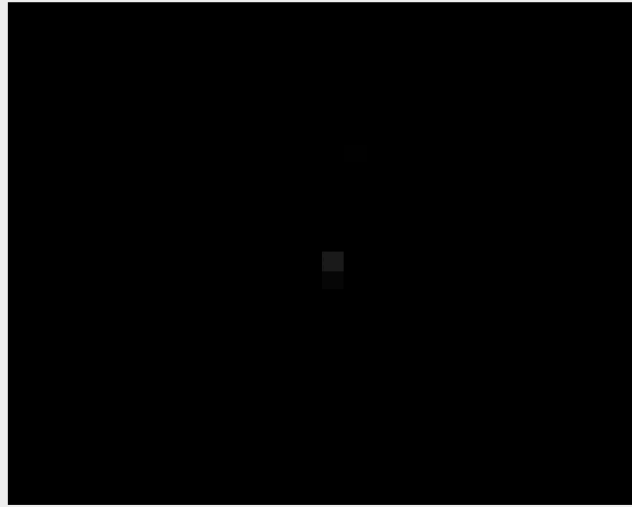




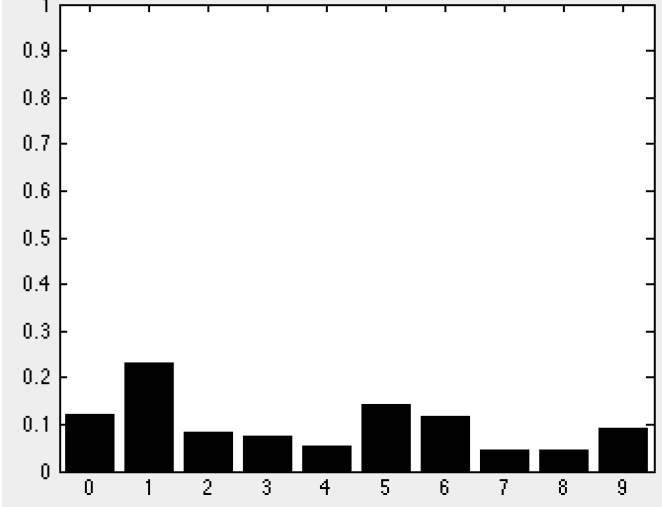
Input Image



Mean-Field Reconstruction



Classification



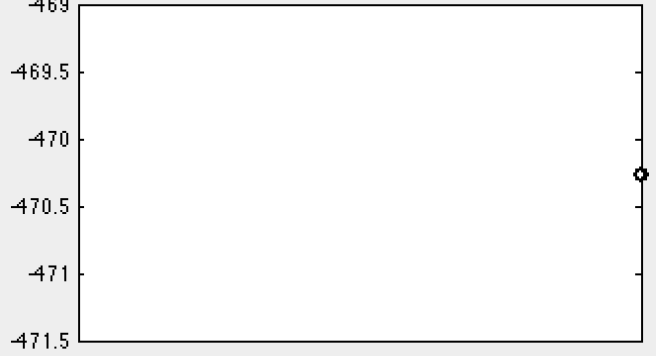
Input Tool

Pencil
 Brush
 Noise
 Eraser

Inductive Principle

SML
 CD
 PL
 RM

Free Energy History

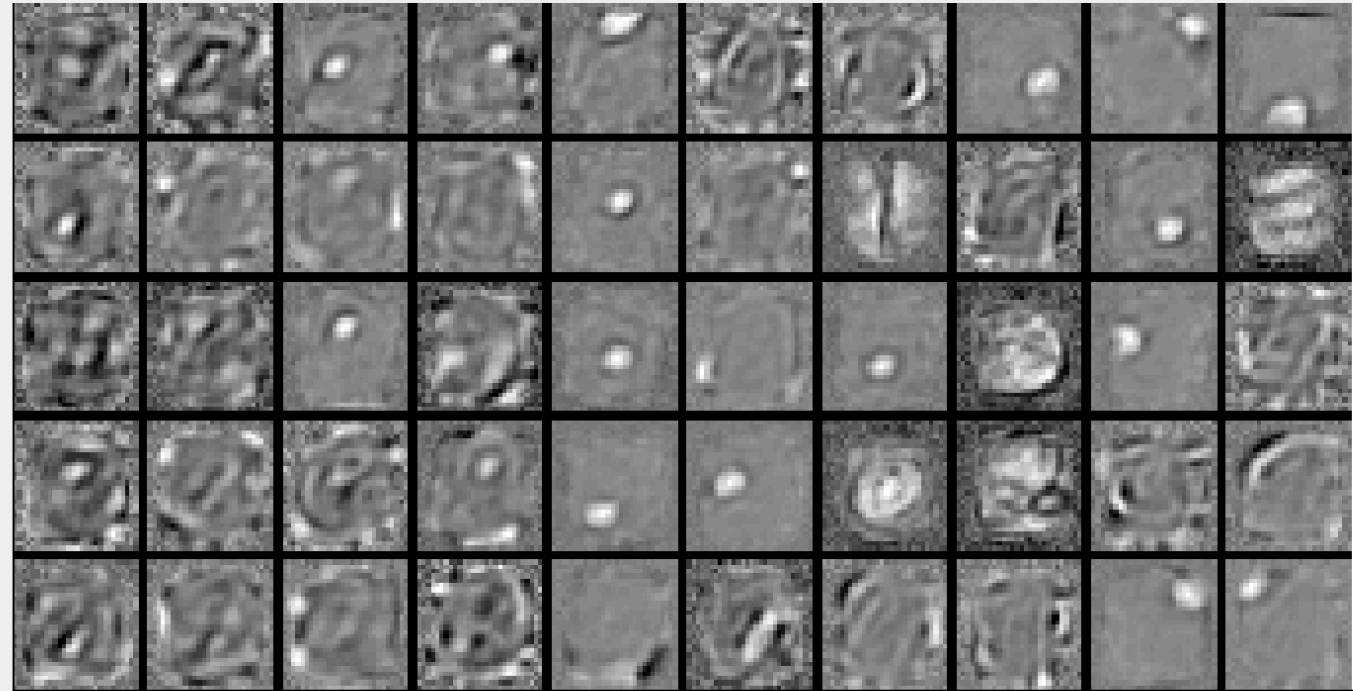


Start Sample

Stop Sample

Reset

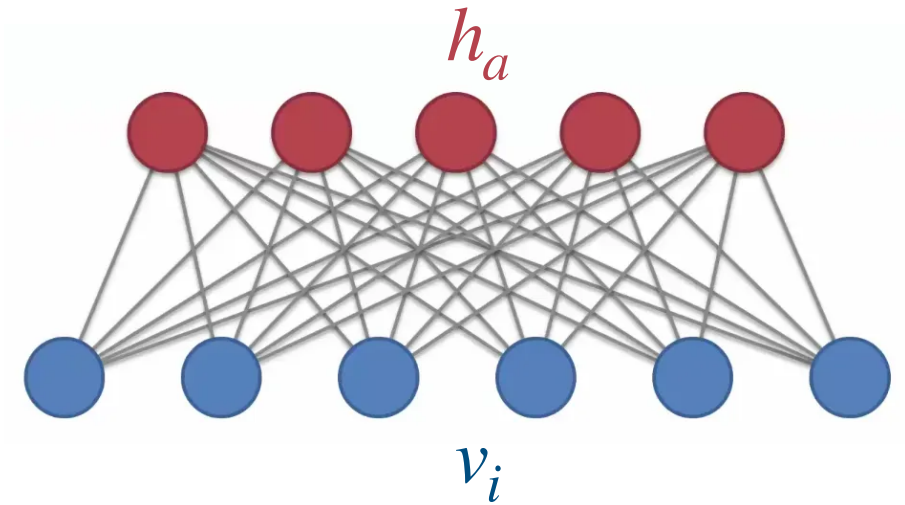
Top Filters



受限玻尔兹曼机 (RBM): 表示学习

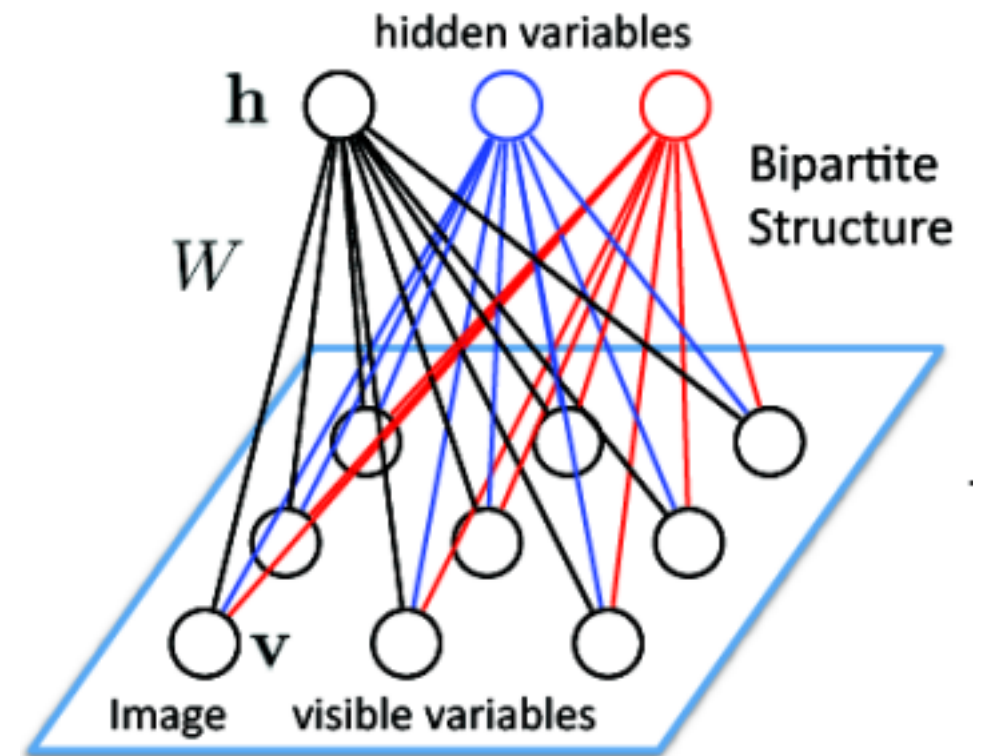
$$P(v, h) = \frac{1}{Z} e^{\sum_{ia} W_{ia} v_i h_a}$$

$$P(v) = \frac{1}{Z} \sum_h e^{\sum_{ia} W_{ia} v_i h_a}$$



- RBM隐变量可以给出数据的表示 (representation)

- ➔ 显变量(数据)分布到隐变量分布
- ➔ 分布维度缩小
- ➔ 类似物理中的重整化



表示学习、预训练、Fine-tune

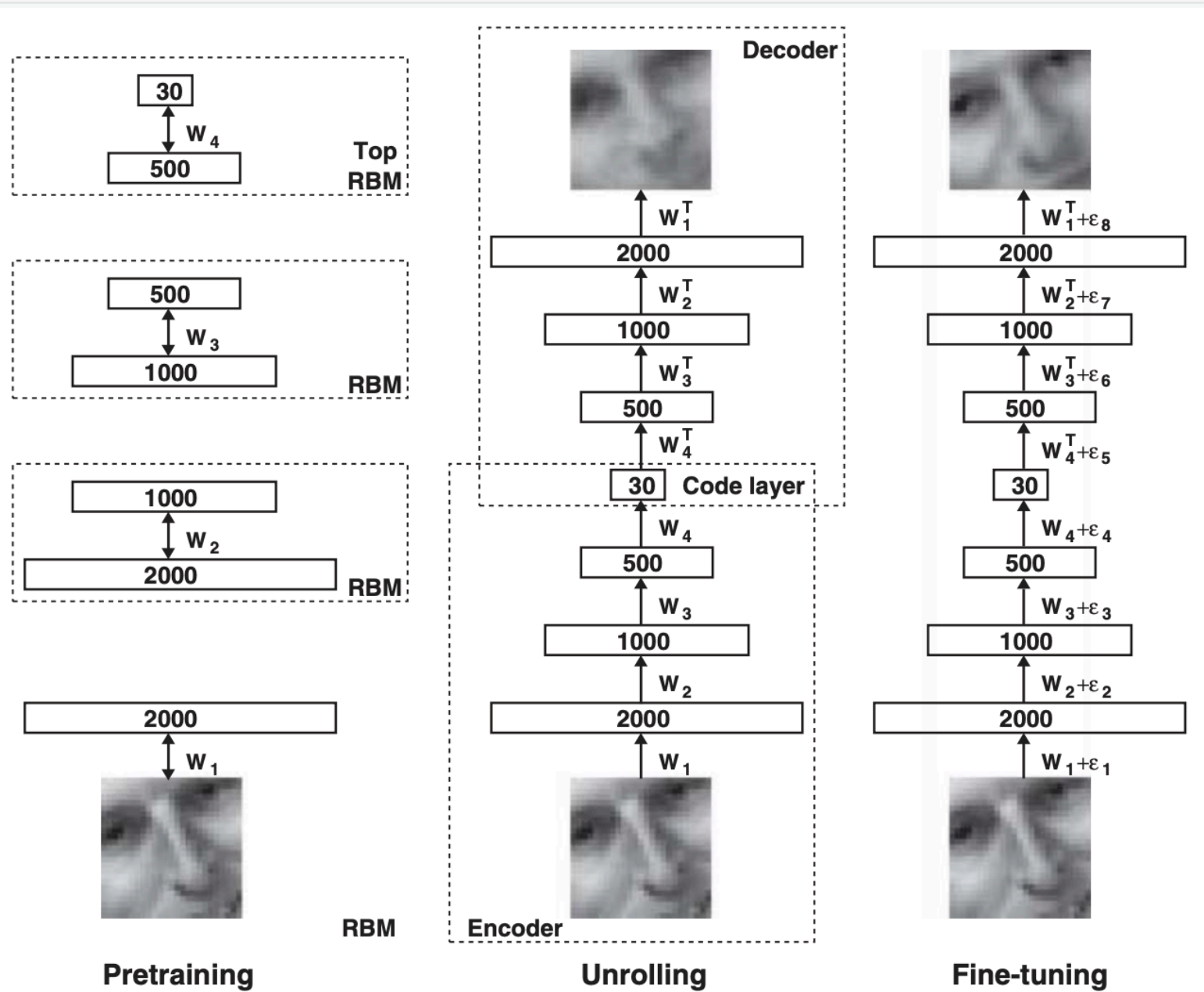


Fig. 1. Pretraining consists of learning a stack of restricted Boltzmann machines (RBMs), each having only one layer of feature detectors. The learned feature activations of one RBM are used as the “data” for training the next RBM in the stack. After the pretraining, the RBMs are “unrolled” to create a deep autoencoder, which is then fine-tuned using backpropagation of error derivatives.

Reducing the Dimensionality of Data with Neural Networks

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

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

Dimensionality reduction facilitates the classification, visualization, communication, and storage of high-dimensional data. A simple and widely used method is principal components analysis (PCA), which finds the directions of greatest variance in a data set and represents each data point in terms of a few coordinates along each of these directions. We describe a nonlinear generalization of PCA that uses an adaptive, multilayer “encoder” network to learn low-dimensional codes that work better than PCA as a tool to reduce the dimensionality of data.

2006 VOL 313 SCIENCE www.sciencemag.org

Learning representations by back-propagating errors

David E. Rumelhart*, Geoffrey E. Hinton† & Ronald J. Williams*

* Institute for Cognitive Science, C-015, University of California, San Diego, La Jolla, California 92093, USA

† Department of Computer Science, Carnegie-Mellon University, Pittsburgh, Philadelphia 15213, USA

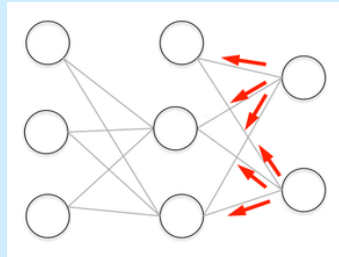
Rumelhart, Hinton, Williams, 1986

“My main contribution was to show how you can use it for learning distributed representations” - Hinton

从Hinton的主要工作看神经网络学习发展

生成学习萌芽

基于RBM的
非监督学习
发展缓慢



Back-propagation
1986

Deep belief network
Autoencoder
Pre-training
Fine-tune

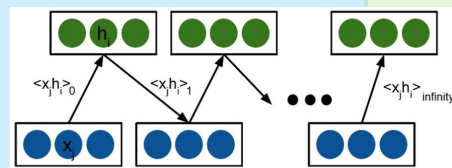
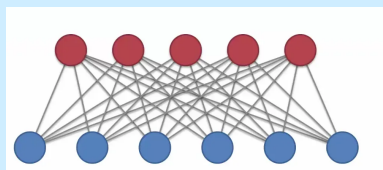
2006

2012

1985
Boltzmann
Machine

1995
Helmholtz
Machine

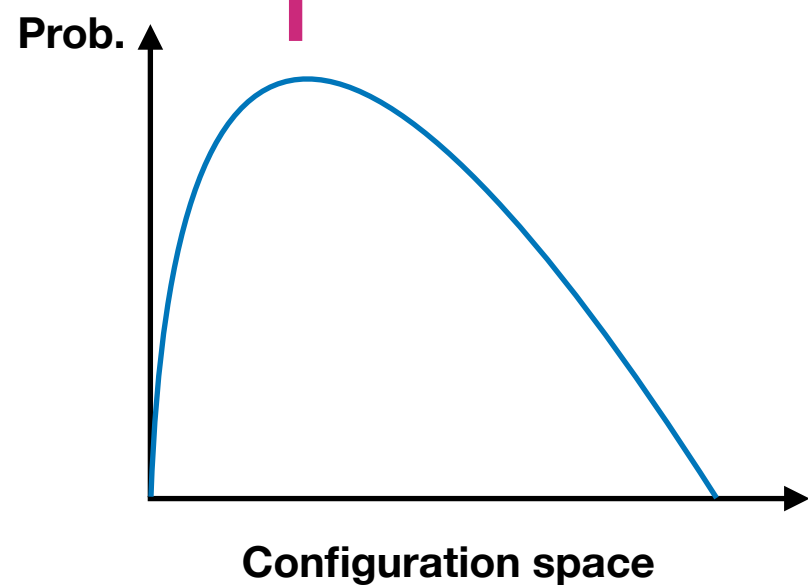
2002
Con. Div.
Product of Experts



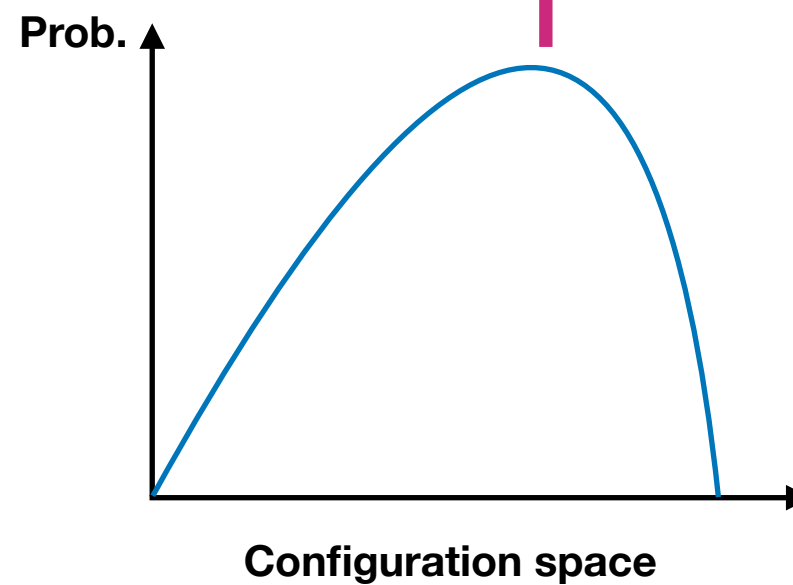
玻尔兹曼分布：样本生成困难,配分函数难以计算



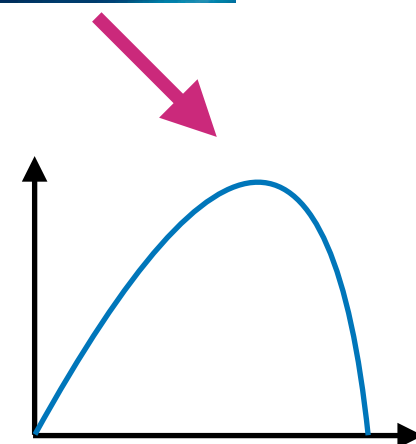
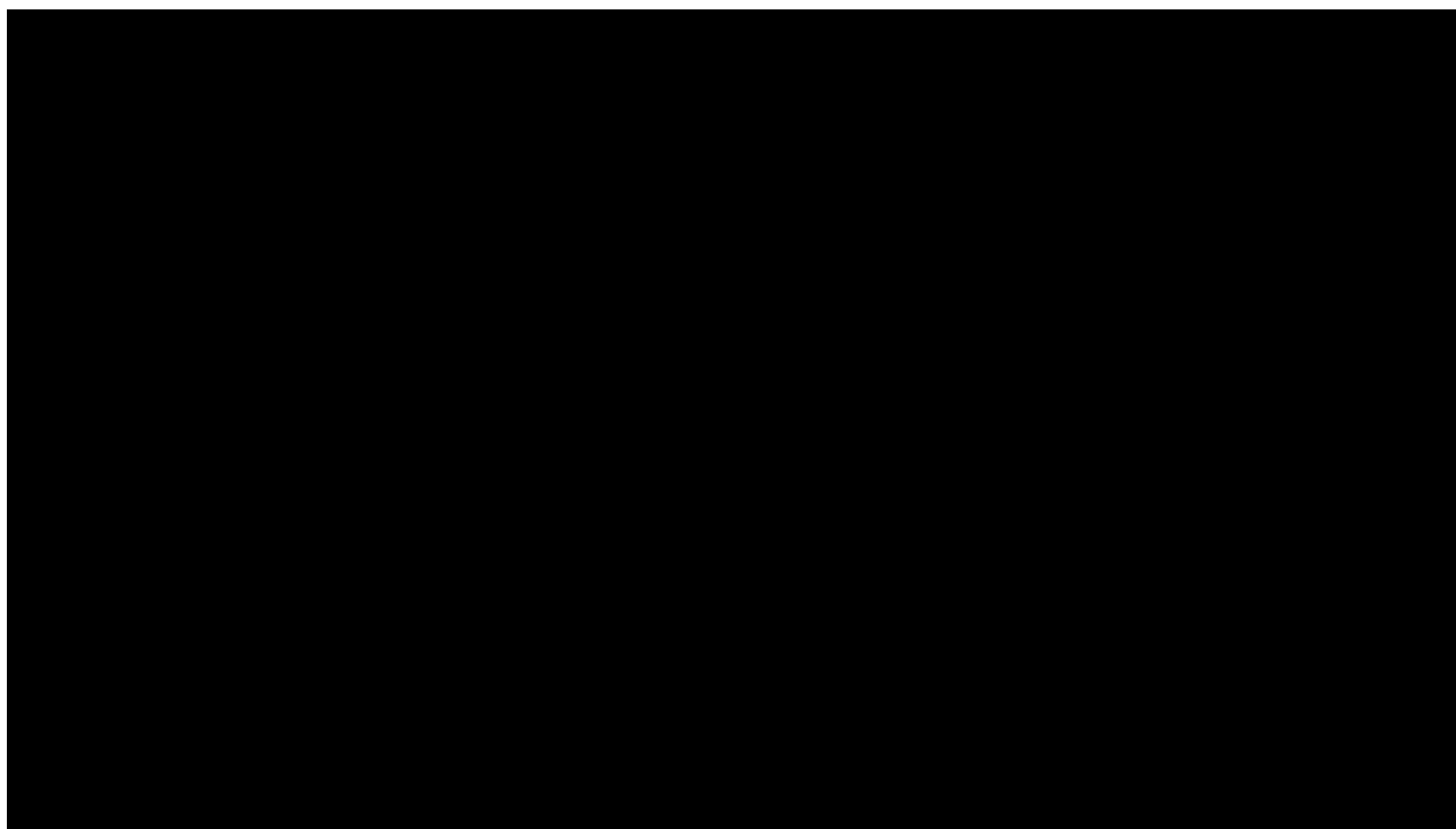
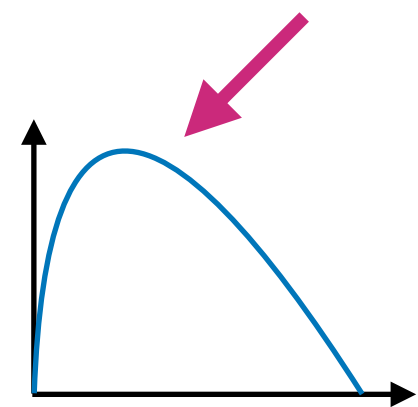
Sampling



Sampling



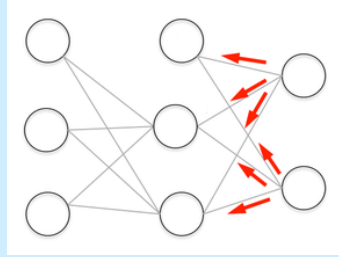
玻尔兹曼分布：样本生成困难,配分函数难以计算



从Hinton的主要工作看神经网络学习发展

生成学习萌芽

酝酿期



Back-propagation
1986

Deep belief network
Autoencoder
Pre-training
Fine-tune

Alex net
2012

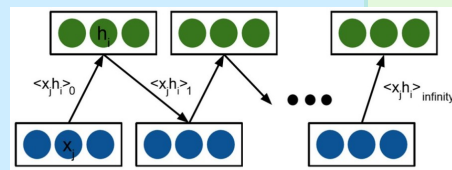
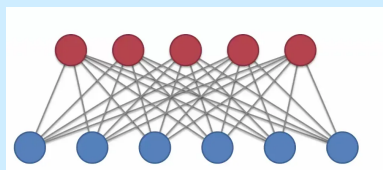


1985
Boltzmann
Machine

1995
Helmholtz
Machine

2002
Con. Div.
Product of Experts

2009
ImageNet



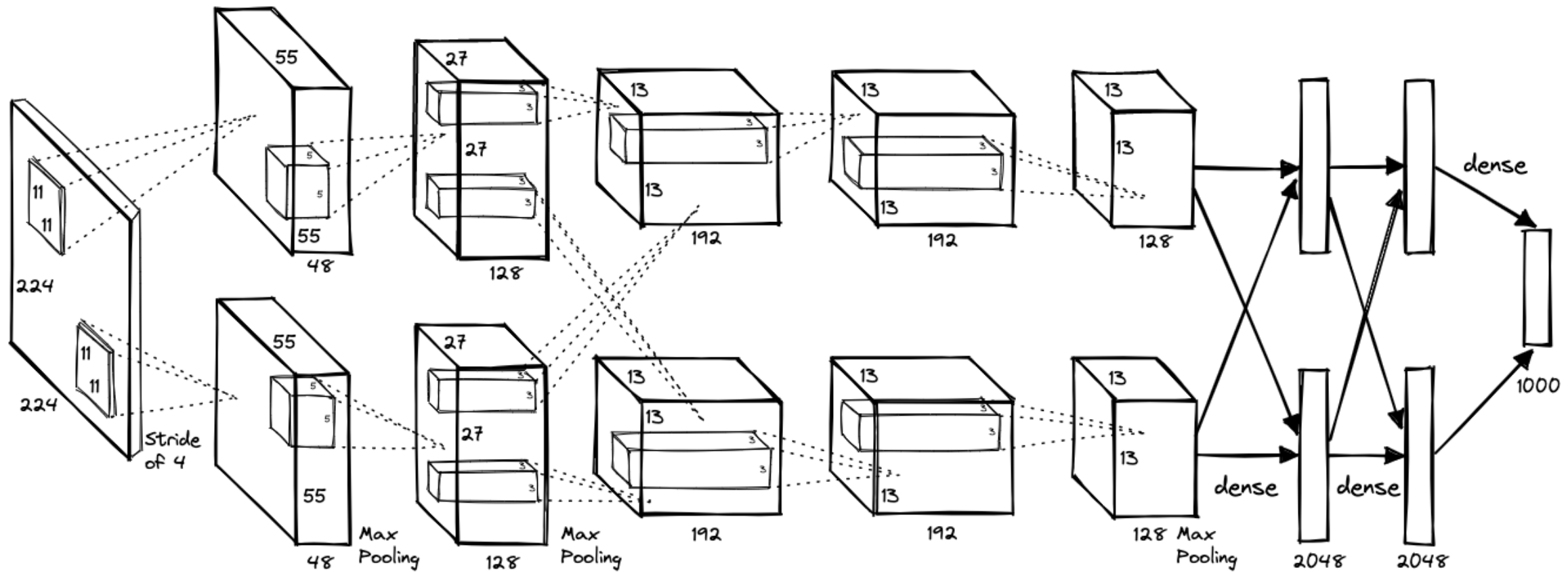
ImageNet



“... while a lot of people are paying attention to models, let’s pay attention to data. Data will redefine how we think about models.” – *Fei-Fei Li*

- WordNet + 人工标签 (Amazon众包服务 + 标签错误矫正)
- 2009: 12 million images across 22,000 categories
- 2010: 第一届 ImageNet Challenge

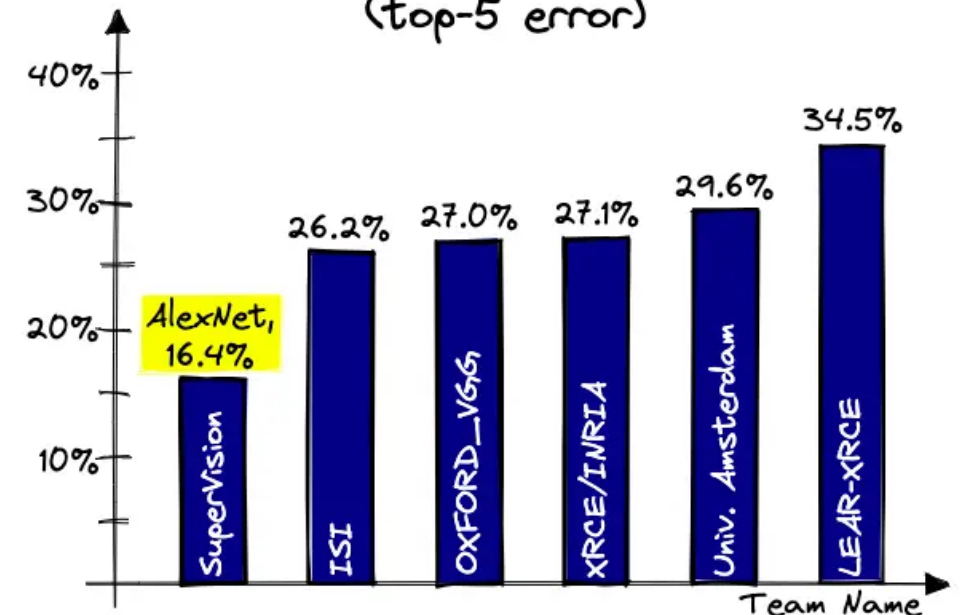
AlexNet: The birth of Deep Learning



Alex Krizhevsky, Ilya Sutskever, Geoffrey Hinton, 2012

- Convolution network + SGD + ImageNet
- Trained across 2 GPUs (Nvidia GTX 580)
- ReLu activation function
- Data Augmentation, Dropout

2012 ImageNet Challenge (top-5 error)

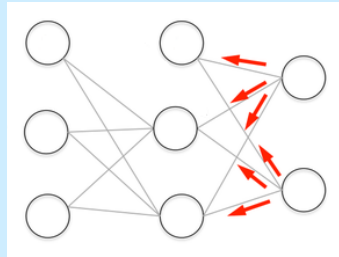


Era of Deep Learning

生成学习萌芽

酝酿期

Deep Learning



Back-propagation
1986

Deep belief network
Autoencoder
Pre-training
Fine-tune

2006

Alex net
2012

ResNet Model
2015

Diffusion

2015

Transformer
2017



ChatGPT
2022



1985
Boltzmann
Machine

1995
Helmholtz
Machine

2002
Con. Div.
Product of Experts

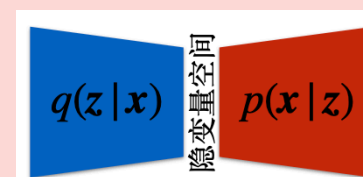
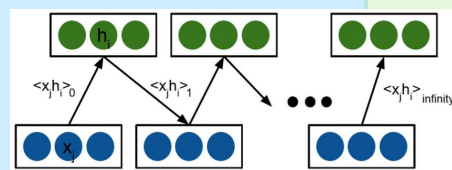
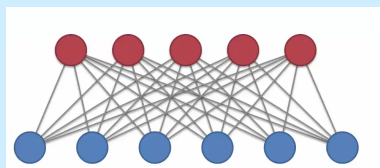
2009
ImageNet

2014
GAN
VAE models

2015
Flow
models

2016
AlphaGo

2018-2020
AlphaFold



Hinton的insight和信心(坚持)

The Forward-Forward Algorithm: Some Preliminary Investigations

Geoffrey Hinton
Google Brain
geoffhinton@google.com

2022年NeurIPS
Hinton 75岁

Abstract

The aim of this paper is to introduce a new learning procedure for neural networks and to demonstrate that it works well enough on a few small problems to be worth further investigation. The Forward-Forward algorithm replaces the forward and backward passes of backpropagation by two forward passes, one with positive (*i.e.* real) data and the other with negative data which could be generated by the network itself. Each layer has its own objective function which is simply to have high goodness for positive data and low goodness for negative data. The sum of the squared activities in a layer can be used as the goodness but there are many other possibilities, including minus the sum of the squared activities. If the positive and negative passes could be separated in time, the negative passes could be done offline, which would make the learning much simpler in the positive pass and allow video to be pipelined through the network without ever storing activities or stopping to propagate derivatives.

1 What is wrong with backpropagation

The astonishing success of deep learning over the last decade has established the effectiveness of performing stochastic gradient descent with a large number of parameters and a lot of data. The

Hinton的insight和信心(坚持)

Recent Papers

Hinton, G. E. (2022)

The Forward-Forward Algorithm: Some Preliminary Investigations

arXiv:2212.13345

[[pdf of final version](#)]

[[ffcode.zip matlab code for the supervised version of FF](#)]

[[load mnistdata.mat in matlab to create the data](#)]

[[README.txt explains what to do to run FF](#)]

Sindy Loewe's translation to python code is available at [https://github.com/sindyloewe/ff](#)

Chen, T., Zhang, R., & Hinton, G. (2022)

Analog bits: Generating discrete data using diffusion models

arXiv preprint arXiv:2208.04202 [[pdf](#)]

Ren, M., Kornblith, S., Liao, R., & Hinton, G. (2022)

Scaling Forward Gradient With Local Losses

arXiv preprint arXiv:2210.03310 [[pdf](#)]

MATLAB CODE

.... [Matlab for Science paper](#)

.... [t-SNE software](#)

.... [trajectory from motor program](#)

.... [ink from trajectory](#)

.... [introduction to python](#)

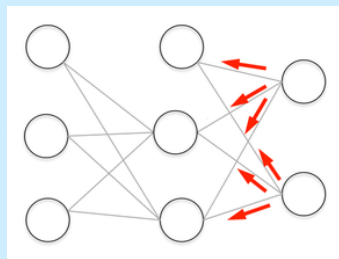


Neural Networks and Statistical Physics ???

生成学习萌芽

酝酿期

Deep Learning



Back-propagation
1986

Deep belief network
Autoencoder
Pre-training
Fine-tune
2006

Alex net
2012

ResNet Model
2015

Diffusion
2015

Transformer
2017

ChatGPT
2022



1985
Boltzmann
Machine

1995
Helmholtz
Machine

2002
Con. Div.
Product of Experts

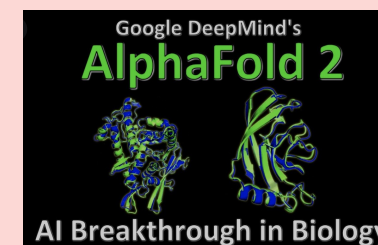
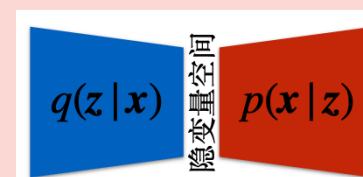
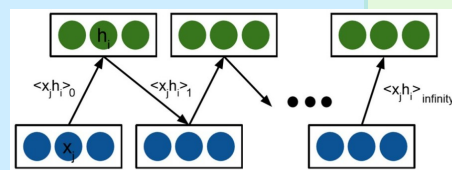
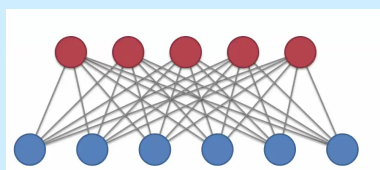
2009
ImageNet

2014
GAN
VAE models

2015
Flow
models

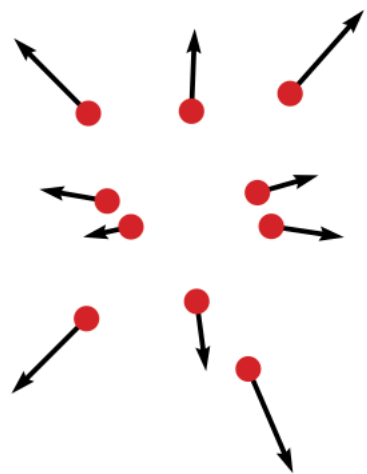
2016
AlphaGo

2018-2020
AlphaFold



现代生成模型有统计物理基因，但直接关系不强

(a) $x \leftarrow x + \tau \nabla \ln p(x) + \sqrt{2\tau} \epsilon$



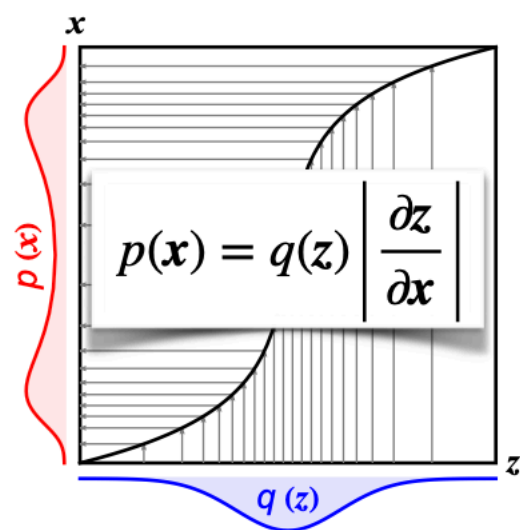
Diffusion models

(b)



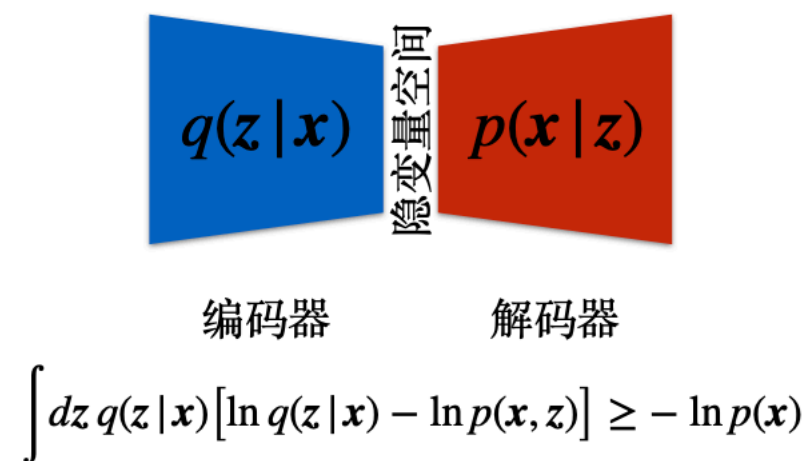
Autoregressive models

(c)



Flow models

(d)



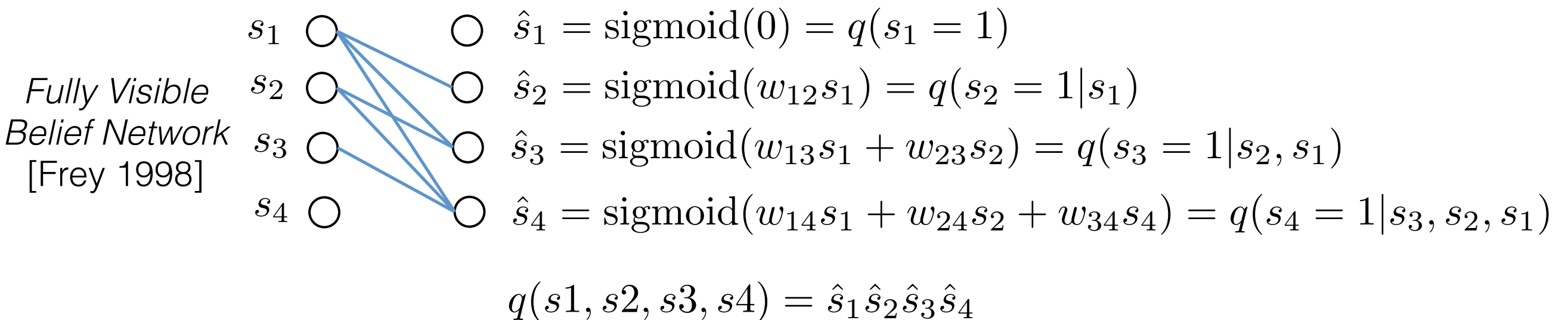
Variational autoencoder

Auto-regressive distribution

- Representing joint distribution using chain rule of conditional probabilities.

$$q(\mathbf{s}) = \prod_i q(s_i | \mathbf{s}_{j < i})$$

$$\begin{aligned} q(s_1, s_2, s_3, s_4) &= q(s_4 | s_3, s_2, s_1) q(s_3, s_2, s_1) \\ &= q(s_4 | s_3, s_2, s_1) q(s_3 | s_2, s_1) q(s_2, s_1) \\ &= q(s_4 | s_3, s_2, s_1) q(s_3 | s_2, s_1) q(s_2 | s_1) q(s_1) \end{aligned}$$



- conditional probabilities* \implies Directed Sampling
Known as **ancestral sampling** [Bishop 2006]

统计物理可否解释机器学习？

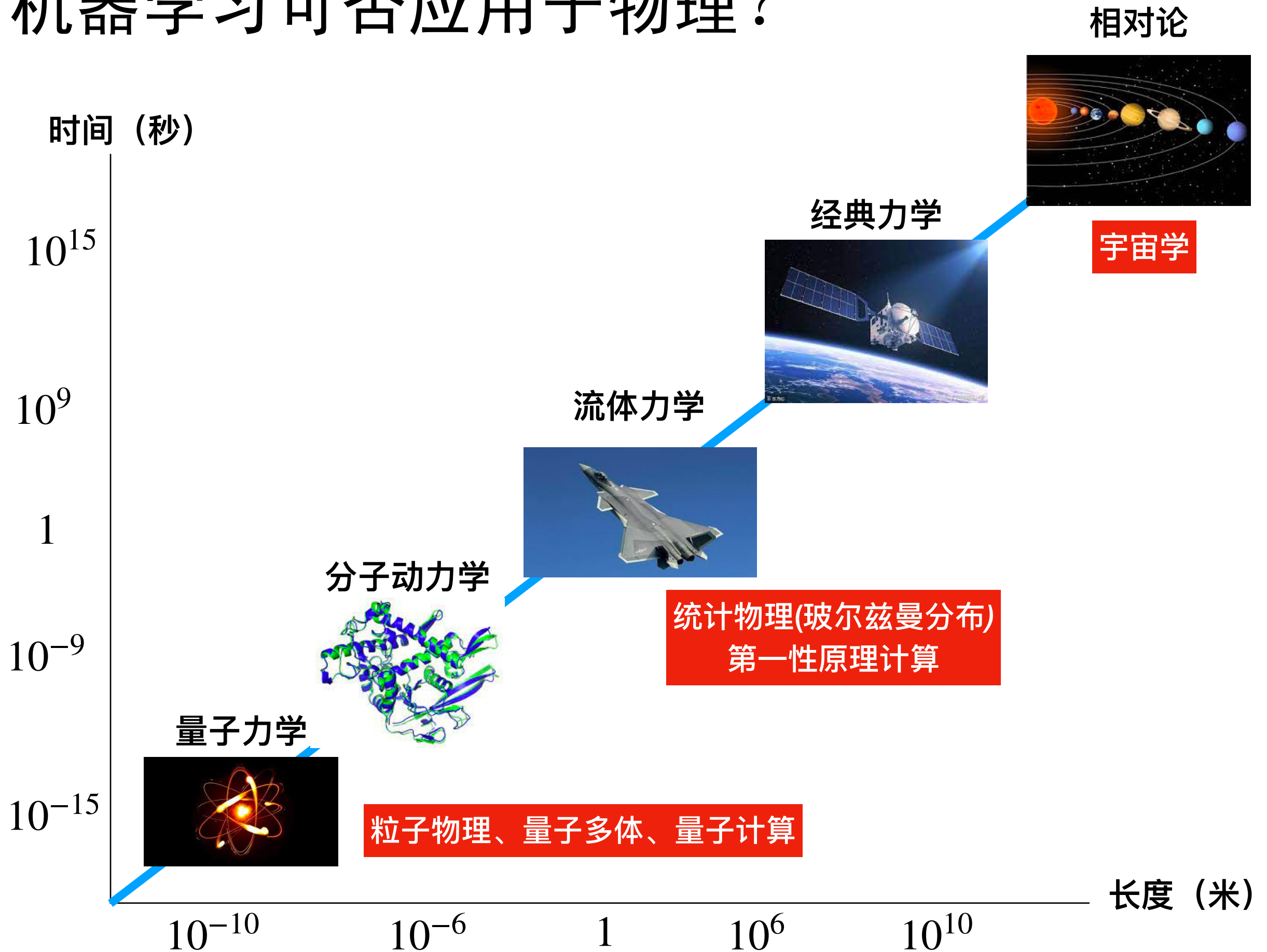
Strengths (Spin-Glass theory, Replica method, mean-field, message passing):

- Hopfield模型相图, 动力学 [*Amit et al 1985, Coolen 2001*]
- Perceptron and generalized linear model [*Krauth, Mezard 1987, Barbier 2019*]
- 浅层、随机、线性网络相图, error landscape [*Saxe et al 2013, Gabrié et al 2019*]
- Information Bottleneck [*Tishby et al, 2000*]
- Glassiness, overparameterization [*Baity-Jesi et al., 2018, Baldassi et al., 2016*]

Weaknesses:

- 简单模型、随机模型、无限宽模型
- 难以考虑复杂数据
- 难以定量描述泛化性

机器学习可否应用于物理？



机器学习在统计物理中的应用

- 识别物质相，相变 (*Wang 2016, Carrasquilla, Mello 2017*)
 - 监督学习、非监督学习，序参量
- 统计力学，重整化 (*Li, Wang 2018, Wu et al 2019*)
 - 变分自由能计算，强化学习，采样，帮助MCMC?
- 非平衡系统 (*Tang et al 2023*)
 - 时间演化，动力学相变
- 非线性动力学系统 (*Pathak et al 2017*)
 - 预测，控制非线性系统，Reservoir 网络

机器学习在量子多体中的应用

- 多体态分类
 - 多体局域化, 哈密顿量、纠缠谱 (*Hsu et al 2018*)
 - non-local序参量, 拓扑不变形 (*Zhang, Kim 2017*)
- 神经网络量子态 (*Carleo, Troyer 2017*)
 - 用RBM、MLP、Autoregressive网络表达波函数
 - 变分蒙特卡洛、强化学习
 - 内禀对称性 (平移、交换, backflow *Luo, Clark 2018*)
- 自动微分赋能张量网络 (*Liao et al 2019*)
 - 时间演化, 动力学相变
- 张量网络与监督、非监督学习
 - 线性分类器 (*Stoudenmire, Schwab 2016*)
 - MPS玻恩机 Born Machine (*Han et al, 2018*)

机器学习与第一性原理计算、物质生成

分子结构、能量面：分子动力学模拟

- 从DFT数据学习能量和力场 (*Zhang et al 2018*)
- 自由能surface, 集体变量 (*Noe 2019*)

材料性质预测

电子密度与DFT：创造新的density functional (*Nagi et al 2018*)

物质生成：从数据生成到原子、分子生成

- 利用对称性, CrystalFormer (*Cao, Luo, Lv, Wang 2024*)

机器学习在粒子物理、宇宙学中的应用

- LHC, LSST, LIGO等大科学装置有大量数据需要处理
- 需要用量子场论、微扰方法、广义相对论得到大量模拟数据
 - 学习神经网络用于快速模拟, 生成数据
 - 快速判断是否存在(引力波)信号、LHC的Trigger系统
 - 从数据学习神经网络, 用于推断模型参数
- Jet 物理: jet标记、flavor 标记、jet聚类、spectral density estimation
- 中微子物理: CNN用于信号处理, 寻找中微子作用位置
- 引力波物理: 信号分类, 广义相对论模型参数估计
- LatticeQCD: Hamilton MCMC获取组态代价极高
 - 利用flow model等生成模型也许可以提供好的proposal

机器学习在量子计算中的应用

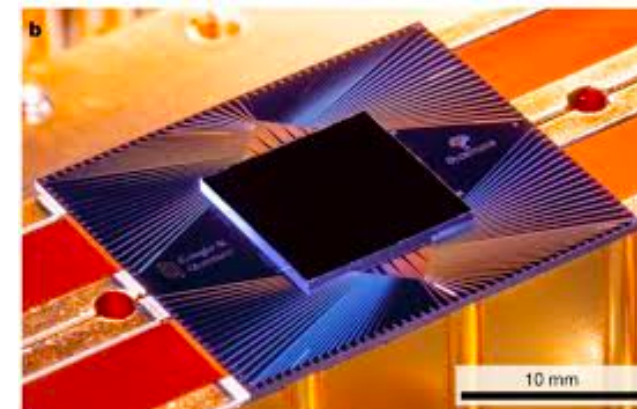
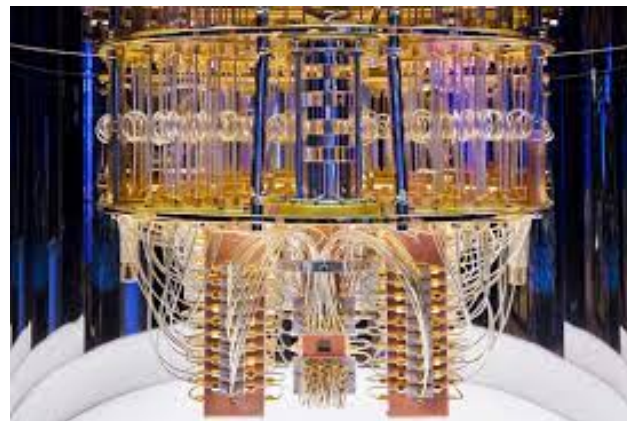
- 量子机器学习:

- 量子数据表示

- 量子算法设计

- 量子优势

- 混合量子-经典算法



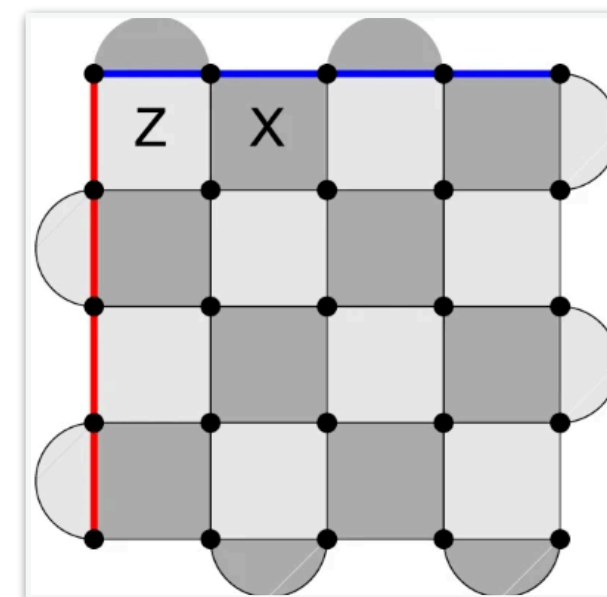
- 量子线路优化: 减少线路深度, 优化量子门参数

- Quantum State Tomography: 相比于张量网络QST, 可以表述纠缠更强的密度矩阵

- 量子纠错:

- 噪音建模 (beyond depolarizing、SI1000, MWPM的prior)

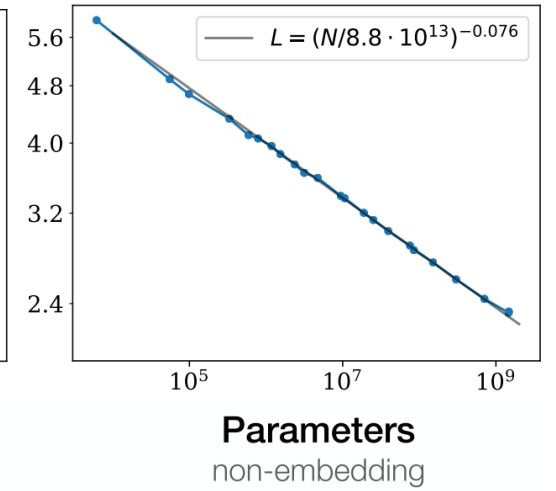
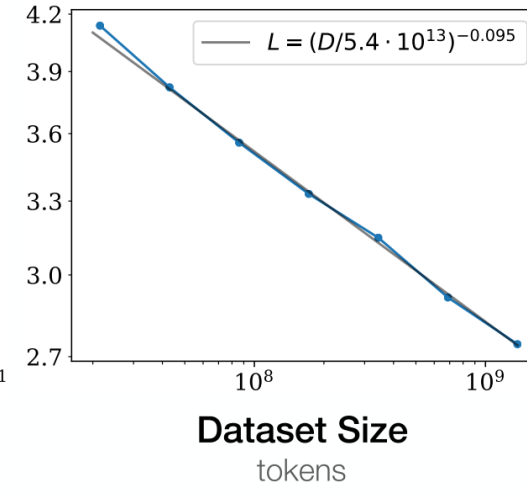
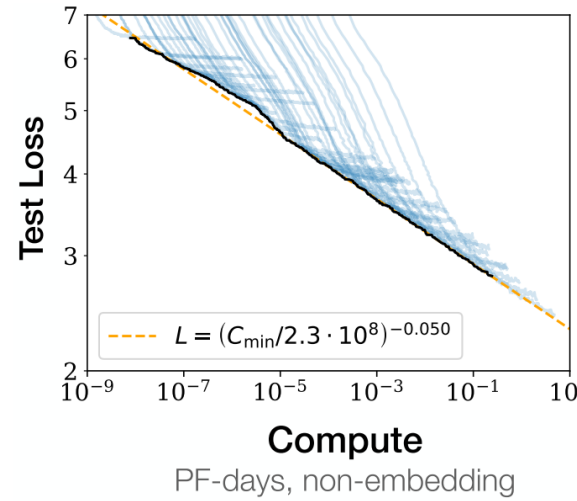
- 从数据中学习logical operator (*Google 2024, qecGPT*)



机器学习应用于物理的挑战与机遇

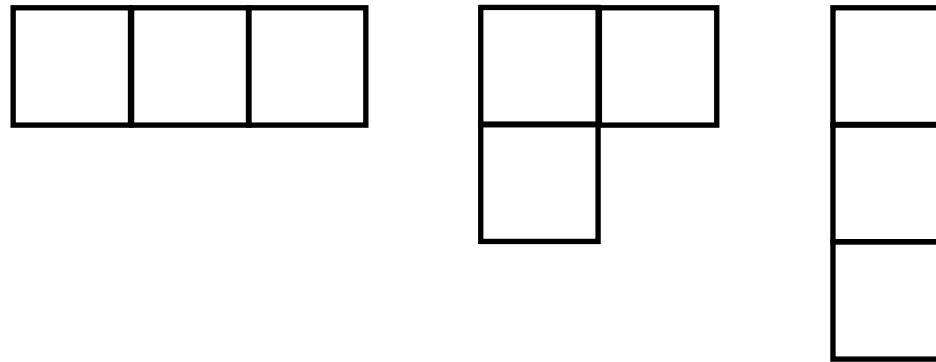
1. 数据量不够

- 高质量数据获取昂贵



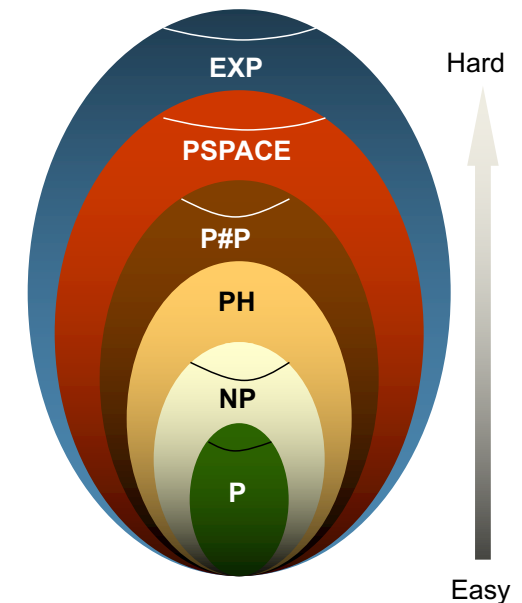
2. 尊重对称性等约束

- 晶体点群, 空间群
- 费米子交换反对称性
-



3. 需要贝叶斯主义, 自由能原理

- 玻尔兹曼分布
- 参数推断 (引力波拟合, 模型参数选择)



Machine Learning and Physics ?

