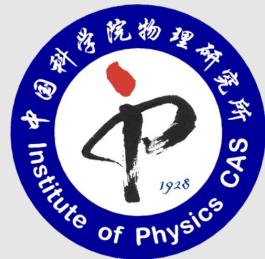


Flow model for computational physics

Qi Zhang (张琦)

Co-Supervisor: Lei Wang (王磊)

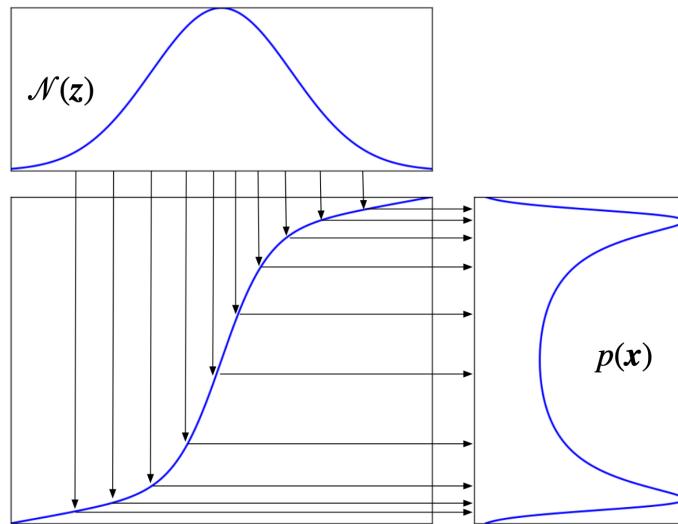
Institute of Physics, CAS



Outline

Normalizing flow

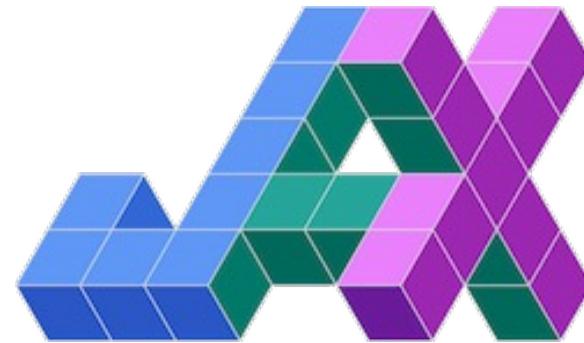
$$p(\mathbf{x}) = \mathcal{N}(\mathbf{z}) \left| \det \left(\frac{\partial \mathbf{z}}{\partial \mathbf{x}} \right) \right|$$



<https://wangleiphy.github.io/lectures/PILtutorial.pdf>
<https://github.com/wangleiphy/ml4p>

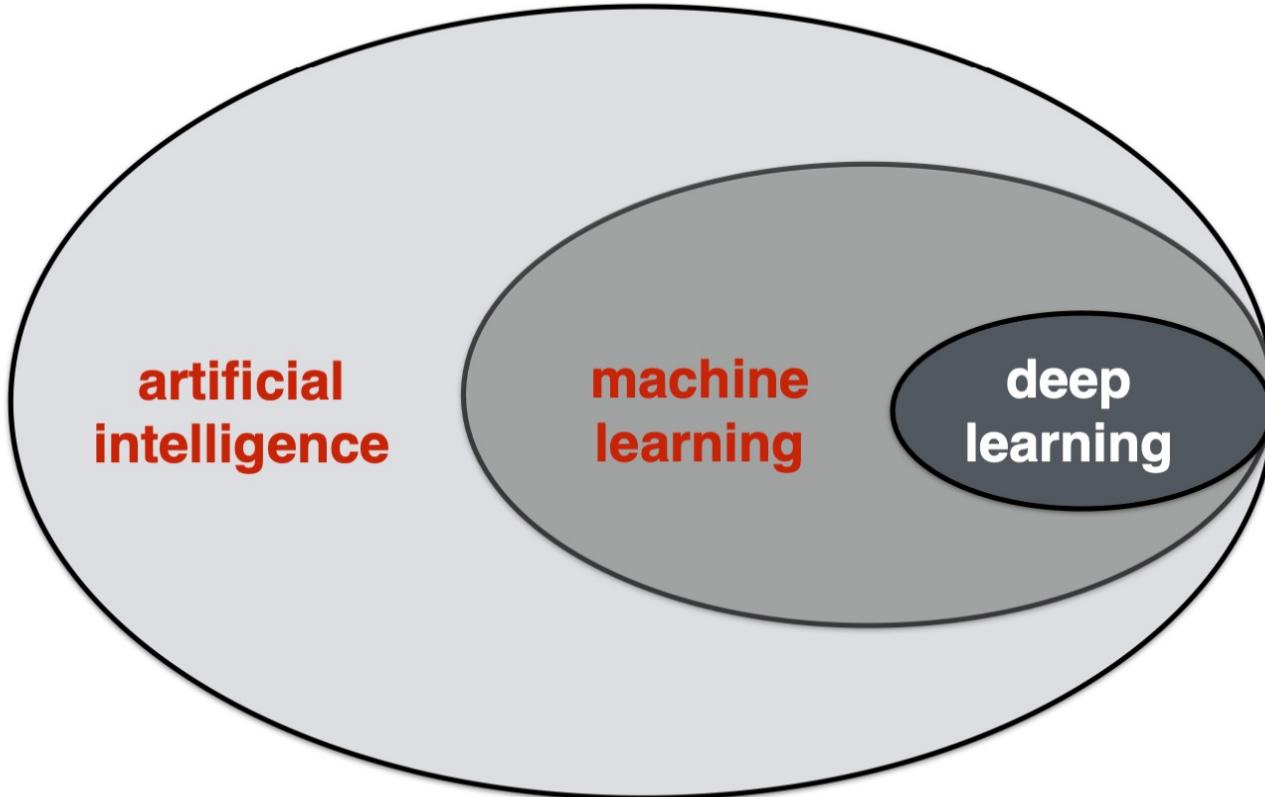
JAX

`grad`, `jit`, `vmap`, and `pmap`.



<https://github.com/google/jax>
<https://jax.readthedocs.io/en/latest/index.html>

Machine learning



Game-changing technology for human life and scientific research.

Machine learning



Discriminative learning

$$y = f(x)$$

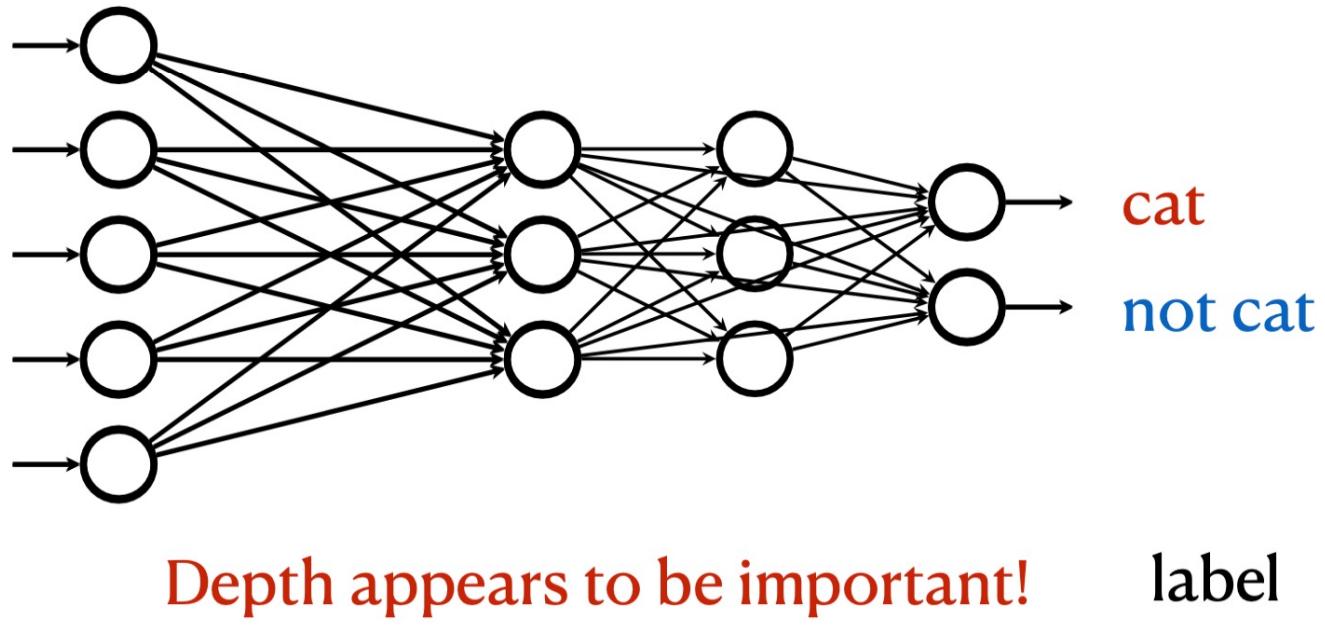
or $p(y | x)$



Generative learning

$$p(x, y)$$

Deep learning



Q: Why does deep learning work?

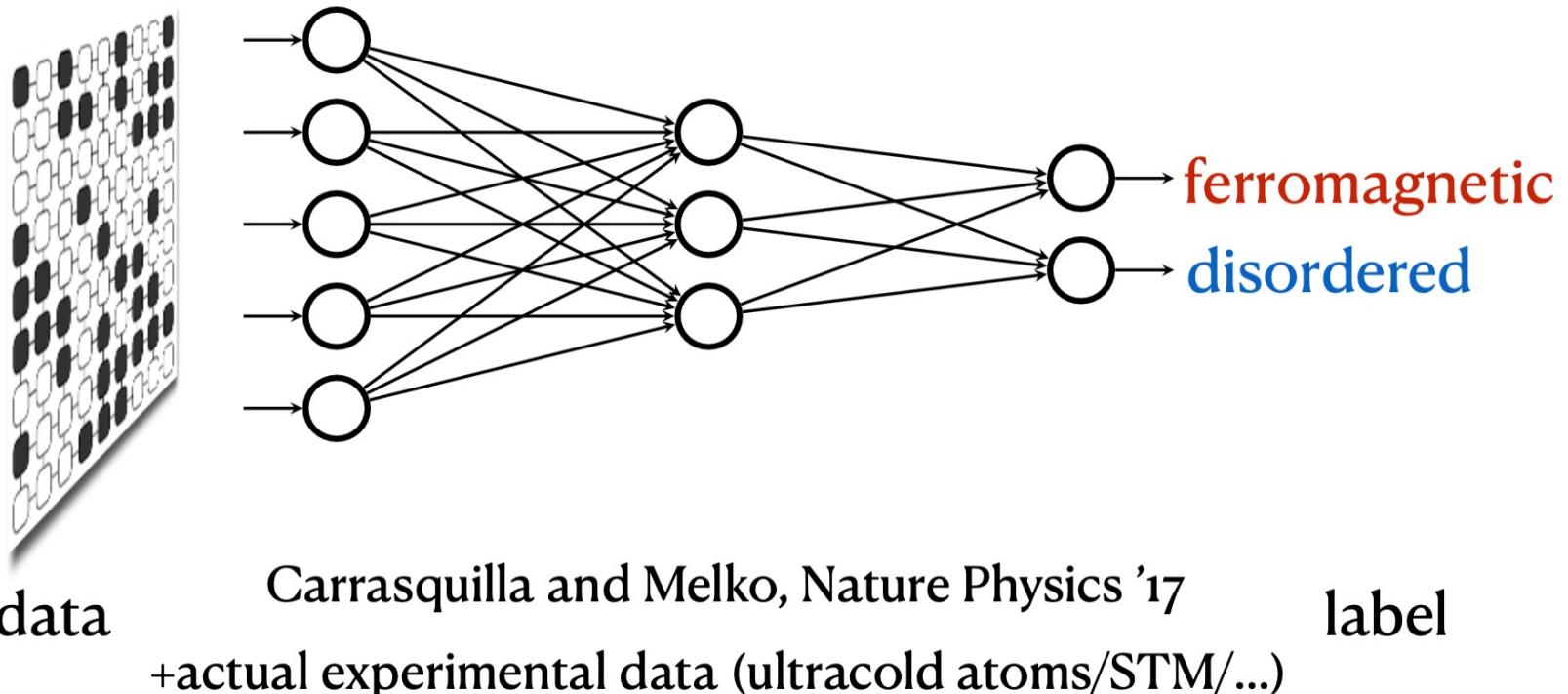
Universal Function Approximator

A: Law of physics: symmetry, locality, composability...

Cybenko 1989
Lin, Tegmark, Rolnick 1608.08225
Hornik, Stinchcombe, White 1989

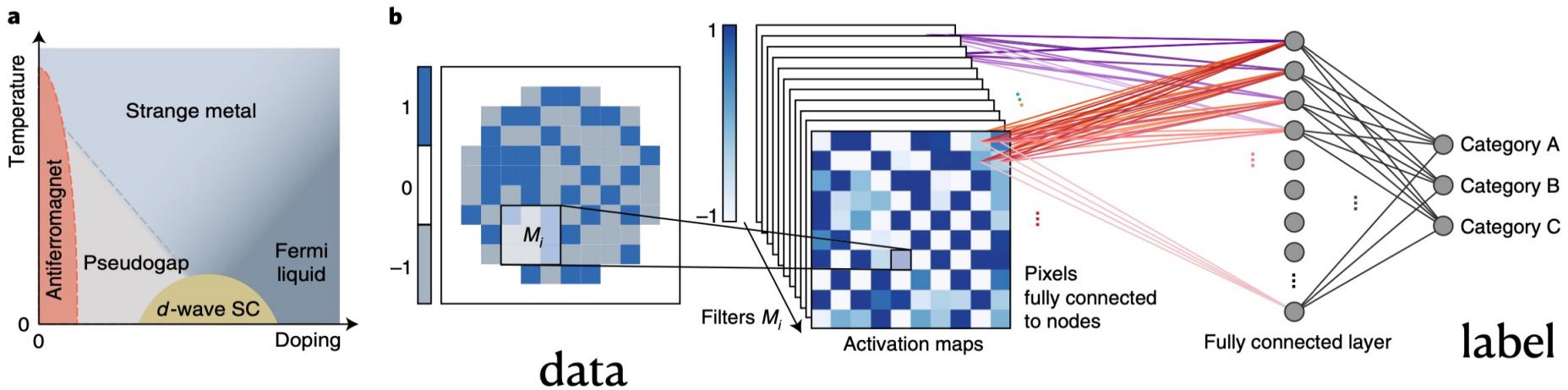
Deep learning

Ising spins



Discovering Phase Transitions with Unsupervised Learning, LW, PRB, 16'

Deep learning



Bohrdt et al, Nature Physics 2019

Generative Pre-Training

Using ChatGPT:

<https://chatgpt.com>

ChatGPT 4o ▾

How to solve 2D Ising model?



The 2D Ising model is a mathematical model used in statistical mechanics to describe ferromagnetism in materials. Solving the 2D Ising model involves finding the partition function, magnetization, and other thermodynamic properties. Here's an overview of the steps involved in solving the 2D Ising model:

1. Model Definition

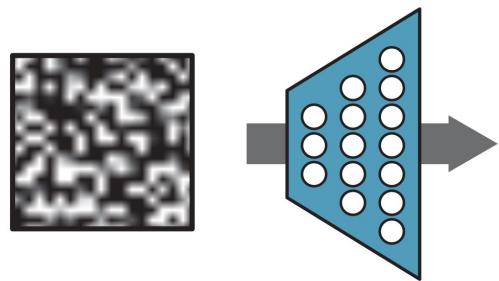
The 2D Ising model consists of a grid (lattice) of spins, where each spin can be in one of two states: up (+1) or down (-1). The Hamiltonian (energy function) of the system is given by:

$$H = -J \sum_{\langle i,j \rangle} s_i s_j - h \sum_i s_i$$

where:

- s_i is the spin at site i .
- J is the interaction energy between neighboring spins.
- $\langle i, j \rangle$ denotes the sum over nearest-neighbor pairs.
- h is an external magnetic field.

Generative arts

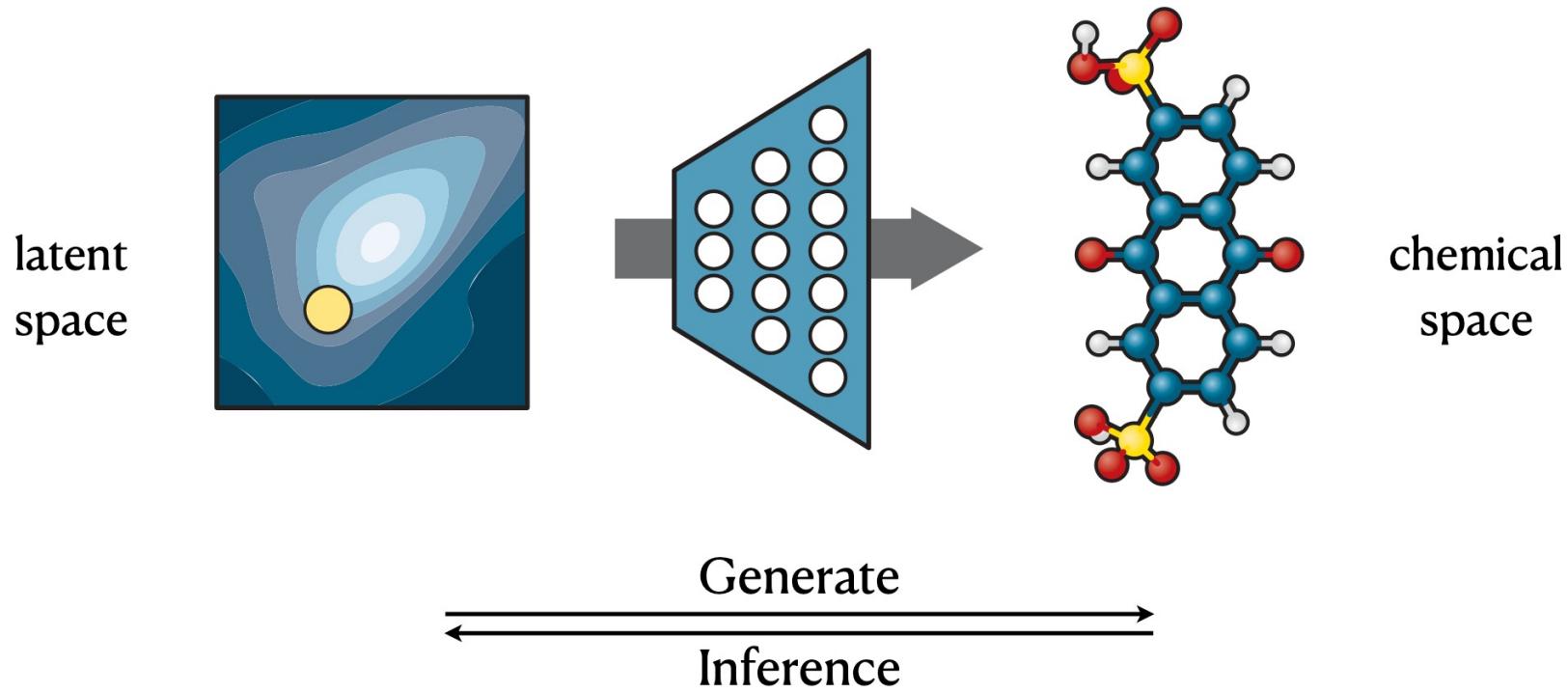


Gaussian Noise Generative Network



\$432,500
25 October 2018
Christie's New York

Generative molecules

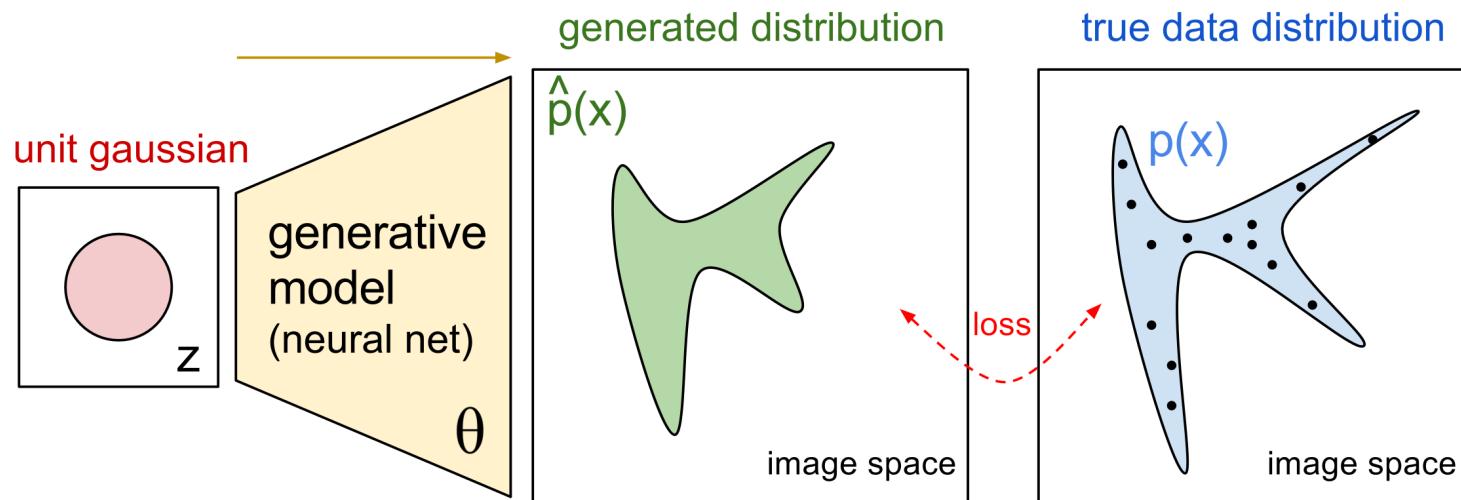


Review: “Inverse molecular design using machine learning”, Sanchez-Lengeling & Aspuru-Guzik, Science ’18

Probabilistic Generative Modeling

$$p(x)$$

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



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

References

Lecture Note: <https://wangleiphy.github.io/lectures/PILtutorial.pdf>

Generative Models for Physicists

Lei Wang*

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

November 5, 2018

Abstract

Generative models generate unseen samples according to a learned joint probability distribution in the high-dimensional space. They find wide applications in density estimation, variational inference, representation learning and more. Deep generative models and associated techniques (such as differentiable programing and representation learning) are cutting-edge technologies physicists can learn from deep learning.

This note introduces the concept and principles of generative modeling, together with applications of modern generative models (autoregressive models, normalizing flows, variational autoencoders etc) as well as the old ones (Boltzmann machines) to physics problems. As a bonus, this note puts some emphasize on physics-inspired generative models which take insights from statistical, quantum, and fluid mechanics.

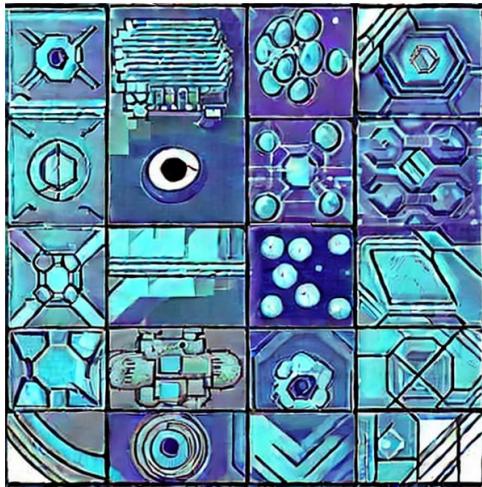
The latest version of the note is at <http://wangleiphy.github.io/>. Please send comments, suggestions and corrections to the email address in below.

CONTENTS

1	GENERATIVE MODELING	2
1.1	Probabilistic Generative Modeling	2
1.2	Generative Model Zoo	4
1.2.1	Boltzmann Machines	4
1.2.2	Autoregressive Models	8
1.2.3	Normalizing Flow	9
1.2.4	Variational Autoencoders	12
1.2.5	Tensor Networks	15
1.2.6	Generative Adversarial Networks	17
1.2.7	Generative Moment Matching Networks	17
1.3	Summary	20
2	PHYSICS APPLICATIONS	21
2.1	Variational Ansatz	21
2.2	Renormalization Group	22
2.3	Monte Carlo Update Proposals	22
2.4	Chemical and Material Design	23
2.5	Quantum Information Science and Beyond	24
3	RESOURCES	25
	BIBLIOGRAPHY	26

References

Machine learning for physicists: <https://github.com/wangleiphy/ml4p>



Machine learning for physicists

1. [Overview](#)
 - Scientific machine learning with and without data
2. [Machine learning practices](#)
 - Hello world example: Hand-written digits recognition
 - Programming frameworks, hardware, and workflow
3. [A hitchhiker's guide to deep learning](#)
 - The four pillars: data, model, loss function, and optimization
 - Deep learning primitives: CNN, GNN, and transformer
4. [Symmetries in machine learning](#)
 - Invariant and equivariant neural networks
 - DeepMD, Euclidean equivariant GNN, Tensor field networks
 - Permutation symmetry and quantum wavefunctions
5. [Differentiable programming](#)
 - The engine of deep learning: automatic differentiation on computation graphs
 - Differentiable DFT/MD/Tensor networks/..., and why they are useful
6. [Generative models-I](#)
 - A dictionary of generative models and statistical physics
 - Boltzmann machines
 - Autoregressive models
 - Variational autoencoders
7. [Generative models-II](#)
 - Normalizing flows
 - Diffusions models
 - Applications of generative models to many-body problems
 - The Universe as a generative model
8. [Wrap up](#)
 - AI for science: why now?

Title image generated by stable diffusion with the prompt: "a tile image for the course on 'Machine learning for physicists', eye-catching, artist style with sci-fi feeling". (Yes, "tile" instead of "title" :P)

Generative models for physicists

Generative modeling	Statistical physics
Negative log-likelihood	Energy function
Score function	Force
Latent variables	Collective variables/coarse graining/renormalization group
Partition function	Free energy calculation
Sample diversity	Enhanced sampling

Generative models for physicists

Generative modeling



Known: samples

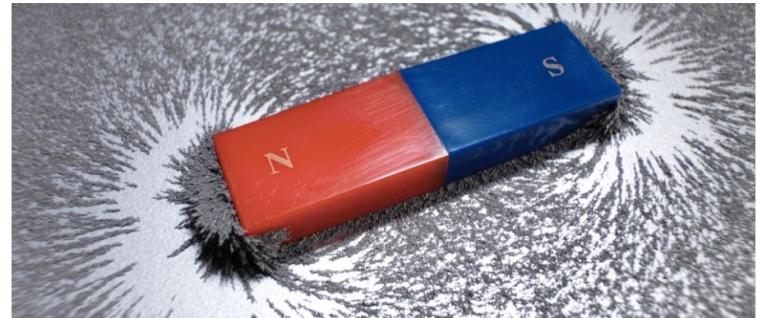
Unknown: generating distribution

Maximum likelihood estimation

“learn from data”

$$\mathcal{L} = - \mathbb{E}_{\mathbf{x} \sim \text{dataset}} [\ln p(\mathbf{x})]$$

Statistical physics



Known: energy function

Unknown: partition function, samples

Variational free energy

“learn from Hamiltonian”

$$F = \mathbb{E}_{\mathbf{x} \sim p(\mathbf{x})} [k_B T \ln p(\mathbf{x}) + H(\mathbf{x})]$$

Kullback-Leibler divergence

How to measure one probability distribution diverges from a reference distribution?

Kullback-Leibler divergence!

$$\text{KL}(p||q) = \sum_{\mathbf{x}} p(\mathbf{x}) \ln \frac{p(\mathbf{x})}{q(\mathbf{x})}.$$

$$\mathbb{KL}(\pi \parallel p) \equiv \int d\mathbf{x} \pi(\mathbf{x}) [\ln \pi(\mathbf{x}) - \ln p(\mathbf{x})]$$

$$\mathbb{KL}(\pi \parallel p) \geq 0$$

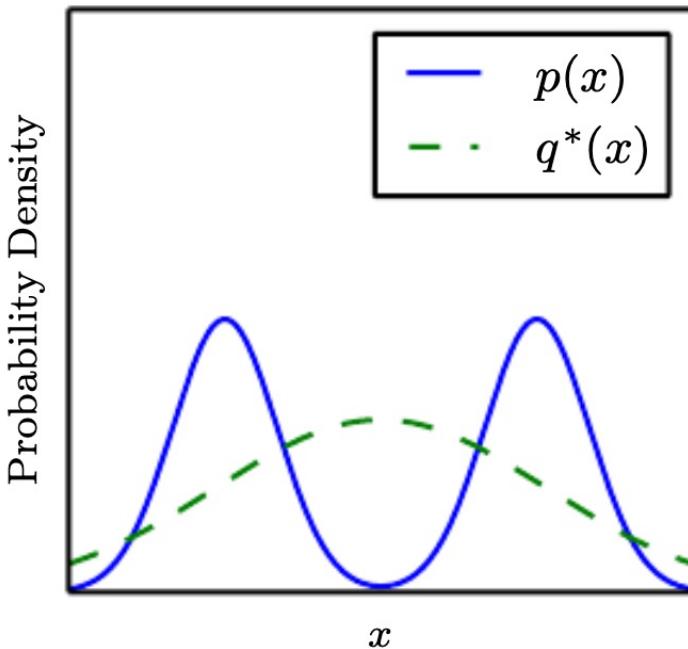
$$\mathbb{KL}(\pi \parallel p) = 0 \iff \pi(\mathbf{x}) = p(\mathbf{x})$$

$$\mathbb{KL}(\pi \parallel p) \neq \mathbb{KL}(p \parallel \pi)$$

Forward KL or Reverse KL?

Maximum likelihood estimation

$$q^* = \operatorname{argmin}_q D_{\text{KL}}(p \| q)$$



Variational free energy

$$q^* = \operatorname{argmin}_q D_{\text{KL}}(q \| p)$$

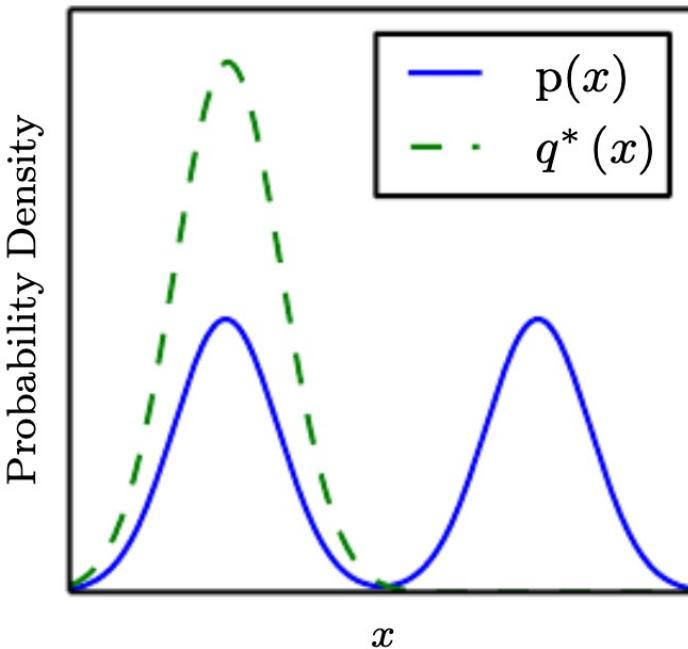


Fig. 3.6, Goodfellow, Bengio, Courville, <http://www.deeplearningbook.org/>

$$\text{KL}(q \| p) = \text{const.} - \mathbb{E}_{x \sim q(x)} [\ln p(x, \theta)]$$

$$\text{KL}(p \| q) = \mathbb{E}_{x \sim p(x)} [\ln p(x) - \ln q(x)]$$

Training approaches

Density estimation

“learn from data”

$$\mathcal{L} = - \mathbb{E}_{\mathbf{x} \sim \text{dataset}} [\ln p(\mathbf{x})]$$

$$\mathbb{KL}(\pi || p) = \sum_{\mathbf{x}} \pi \ln \pi - \underbrace{\sum_{\mathbf{x}} \pi \ln p}_{\mathcal{L}}$$

Sample from dataset in the physical space

Variational calculation

“learn from Hamiltonian”

$$\mathcal{L} = \int d\mathbf{x} p(\mathbf{x}) [\ln p(\mathbf{x}) + \beta H(\mathbf{x})]$$

$$\mathcal{L} + \ln Z = \mathbb{KL}\left(p || \frac{e^{-\beta H}}{Z}\right) \geq 0$$

Sample in the latent space

Learn from Hamiltonian

Target distribution:

$$\pi(\mathbf{x}) \propto e^{-H/k_B T}$$

Boltzmann distribution!

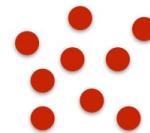
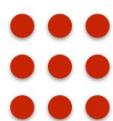
$$\min_{\theta} \text{KL}(p_{\theta} \parallel \pi) \iff \min_{\theta} \left\{ \mathbb{E}_{\mathbf{x} \sim p_{\theta}(\mathbf{x})} [H(\mathbf{x}) + k_B T \ln p_{\theta}(\mathbf{x})] \right\}$$

↑ ↑
model target

Variational free energy

$$F = E - TS$$

energy entropy



F is a **cost function** given by Nature

Deep variational free energy approach

Use deep generative models as the variational density

$$F[p] = \mathbb{E}_{\mathbf{x} \sim p(\mathbf{x})} [H(\mathbf{x}) + k_B T \ln p(\mathbf{x})]$$

↓ ↓
energy entropy 😊

Li and LW, PRL '18
Wu, LW, Zhang, PRL '19

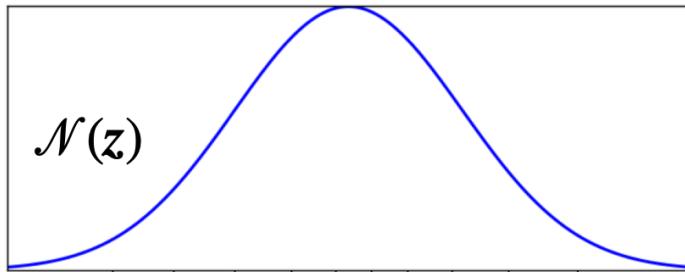
with normalizing flow &
autoregressive models

- ✓ Tractable entropy
- ✓ Direct sampling
- ✓ Turning a sampling problem to an optimization problem
better leverages the deep learning engine:

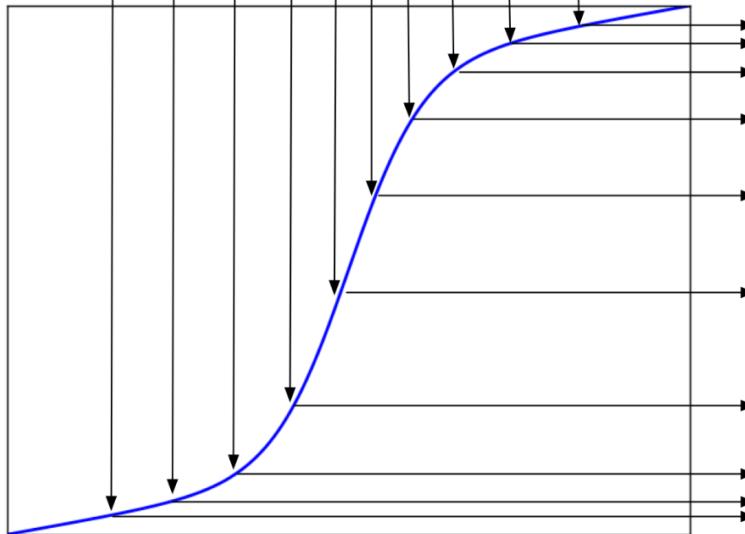


Normalizing flow

Base density



“neural net”
with 1 neuron



$$p(\mathbf{x}) = \mathcal{N}(\mathbf{z}) \left| \det \left(\frac{\partial \mathbf{z}}{\partial \mathbf{x}} \right) \right|$$

Review article 1912.02762

Target
density

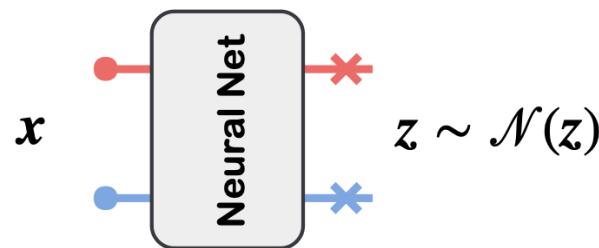
Normalizing flow

Change of variables $x \leftrightarrow z$ **with deep neural nets**

$$p(x) = \mathcal{N}(z) \left| \det \left(\frac{\partial z}{\partial x} \right) \right|$$

Review article 1912.02762
Tutorial https://iclr.cc/virtual_2020/speaker_4.html

composable, differentiable, and invertible mapping between manifolds



Learn probability transformations with normalizing flows

Got this name in Tabak & Vanden-Eijnden, Commun. Math. Sci. '10

Monte Carlo Gradient Estimators

$$F[p] = \mathbb{E}_{\mathbf{x} \sim p(\mathbf{x})} [H(\mathbf{x}) + k_B T \ln p(\mathbf{x})]$$

↓ ↓
energy entropy 😊

$$p(\mathbf{x}) = \mathcal{N}(\mathbf{z}) \left| \det \left(\frac{\partial \mathbf{z}}{\partial \mathbf{x}} \right) \right|$$

Review: 1906.10652

$$\nabla_{\boldsymbol{\theta}} \mathbb{E}_{\mathbf{x} \sim p_{\boldsymbol{\theta}}} [f(\mathbf{x})]$$

Reinforcement learning
Variational inference
Variational Monte Carlo
Variational quantum algorithms
...

Score function estimator (REINFORCE)

$$\nabla_{\boldsymbol{\theta}} \mathbb{E}_{\mathbf{x} \sim p_{\boldsymbol{\theta}}} [f(\mathbf{x})] = \mathbb{E}_{\mathbf{x} \sim p_{\boldsymbol{\theta}}} [f(\mathbf{x}) \nabla_{\boldsymbol{\theta}} \ln p_{\boldsymbol{\theta}}(\mathbf{x})]$$

Pathwise estimator (Reparametrization trick) $\mathbf{x} = g_{\boldsymbol{\theta}}(\mathbf{z})$

$$\nabla_{\boldsymbol{\theta}} \mathbb{E}_{\mathbf{x} \sim p_{\boldsymbol{\theta}}} [f(\mathbf{x})] = \mathbb{E}_{\mathbf{z} \sim \mathcal{N}(\mathbf{z})} [\nabla_{\boldsymbol{\theta}} f(g_{\boldsymbol{\theta}}(\mathbf{z}))]$$

Choose the one with the lowest variance

Design principles



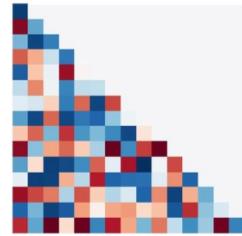
Composability

$$z = \mathcal{T}(x)$$

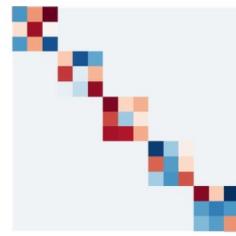
$$\mathcal{T} = \mathcal{T}_1 \circ \mathcal{T}_2 \circ \mathcal{T}_3 \circ \dots$$

Balanced
efficiency &
inductive bias

$$\left| \det \left(\frac{\partial z}{\partial x} \right) \right|$$



Autoregressive



Neural RG

$$\frac{\partial \rho(x, t)}{\partial t} + \nabla \cdot [\rho(x, t)v] = 0$$

Continuous flow

Example of a building block

Forward

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

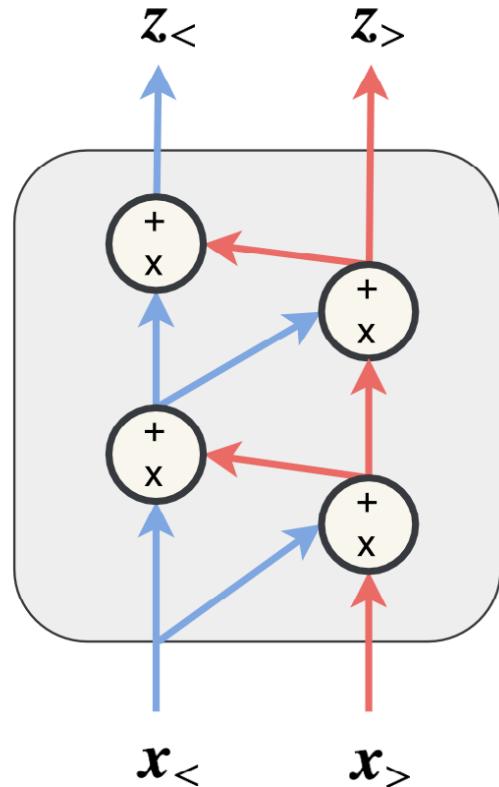
arbitrary
neural nets

Inverse

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

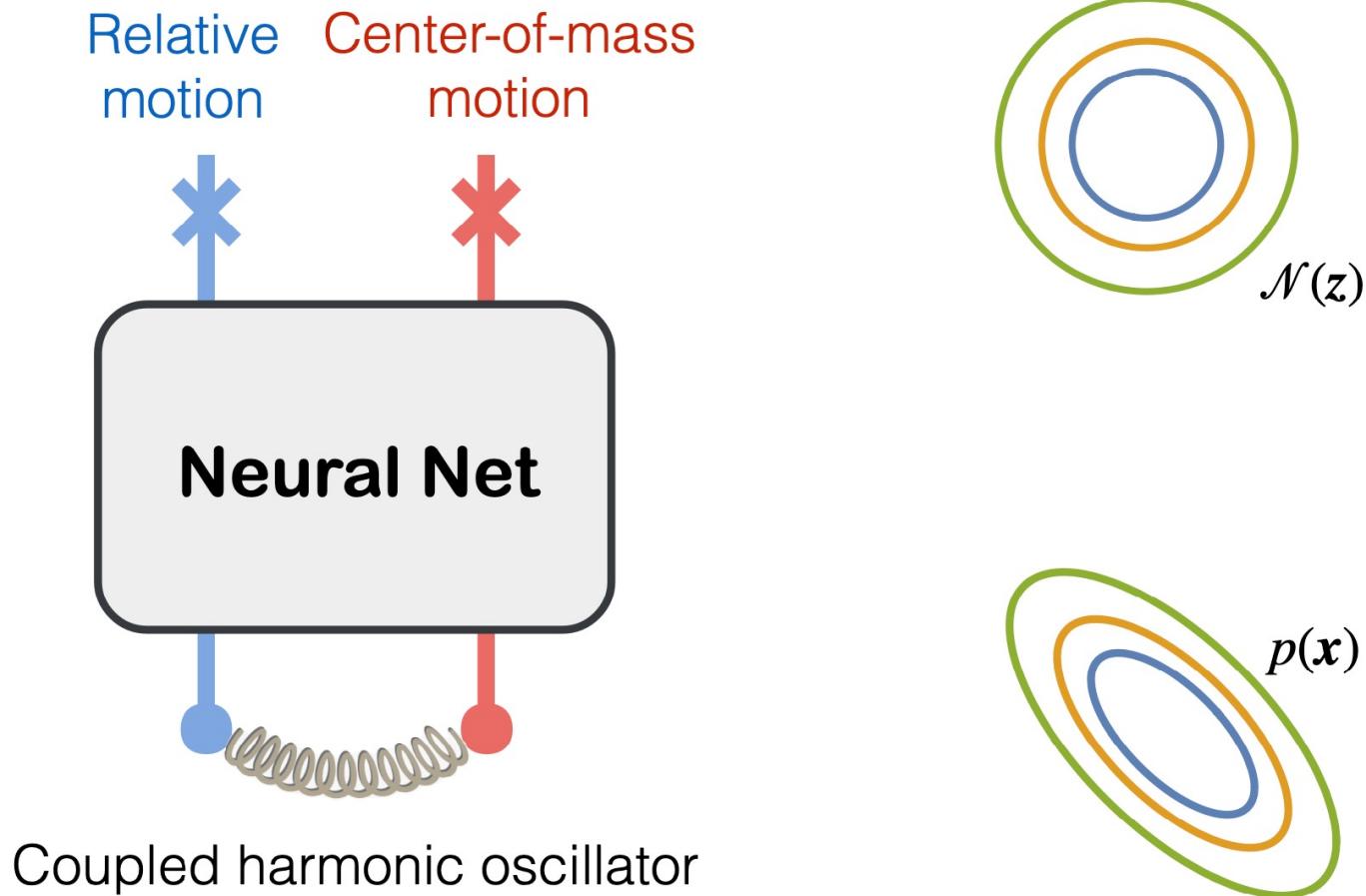
Log-Abs-Jacobian-Det

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



Real NVP, Dinh et al, 1605.08803

How it can be useful in physics?



Continuous normalizing flows

$$\ln p(\mathbf{x}) = \ln \mathcal{N}(\mathbf{z}) - \ln \left| \det \left(\frac{\partial \mathbf{x}}{\partial \mathbf{z}} \right) \right|$$

Consider infinitesimal change-of-variables Chen et al 1806.07366

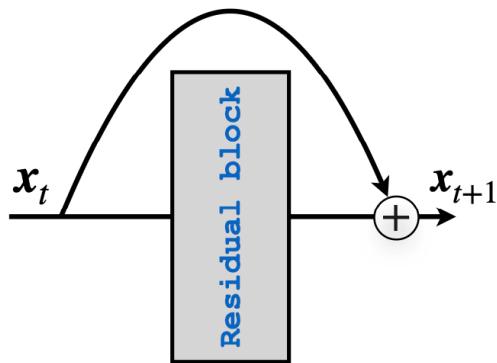
$$\mathbf{x} = \mathbf{z} + \varepsilon \mathbf{v} \quad \ln p(\mathbf{x}) - \ln \mathcal{N}(\mathbf{z}) = -\ln \left| \det \left(1 + \varepsilon \frac{\partial \mathbf{v}}{\partial \mathbf{z}} \right) \right|$$

$$\varepsilon \rightarrow 0$$

$$\frac{d\mathbf{x}}{dt} = \mathbf{v} \quad \frac{d \ln \rho(\mathbf{x}, t)}{dt} = -\nabla \cdot \mathbf{v}$$

Neural Ordinary Differential Equations

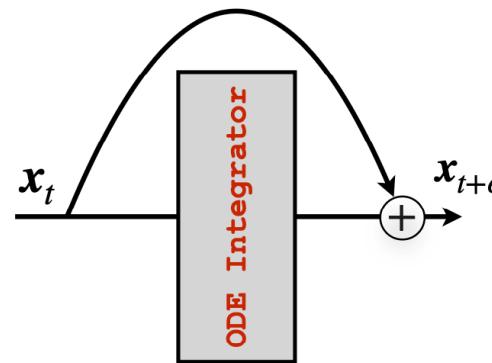
Residual network



$$x_{t+1} = x_t + f(x_t)$$

Chen et al, 1806.07366

ODE integration

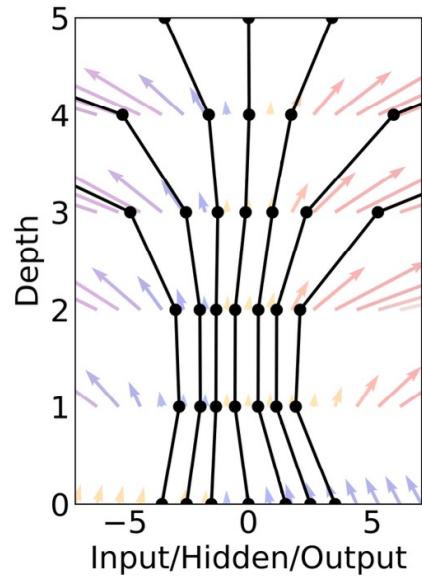


$$dx/dt = f(x)$$

Harbor el al 1705.03341
Lu et al 1710.10121,
E Commun. Math. Stat 17'...

Neural Ordinary Differential Equations

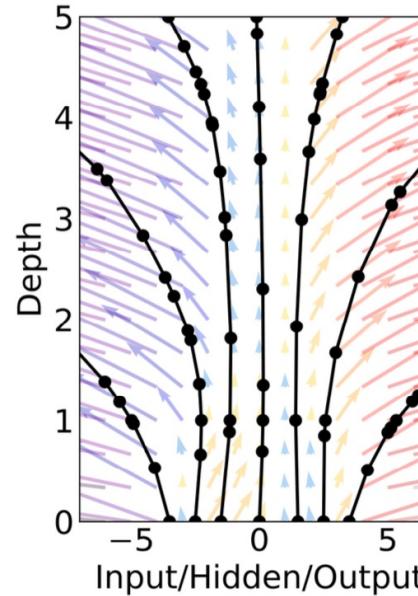
Residual network



$$\mathbf{x}_{t+1} = \mathbf{x}_t + f(\mathbf{x}_t)$$

Chen et al, 1806.07366

ODE integration

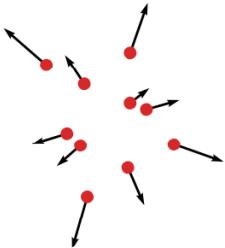


$$d\mathbf{x}/dt = f(\mathbf{x})$$

Harbor et al 1705.03341
Lu et al 1710.10121,
E Commun. Math. Stat 17'...

Fluid physics behind flows

$$\frac{dx}{dt} = v$$



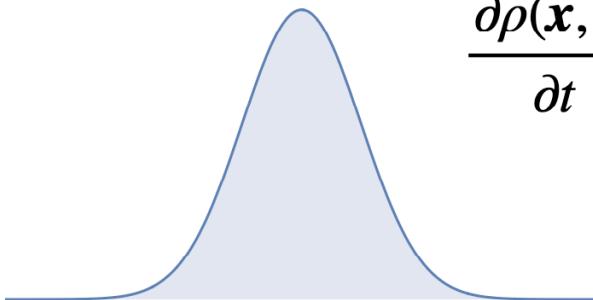
$$\frac{d \ln \rho(\mathbf{x}, t)}{dt} = - \nabla \cdot \mathbf{v}$$

$$\frac{d}{dt} = \frac{\partial}{\partial t} + \mathbf{v} \cdot \nabla$$

“material derivative”

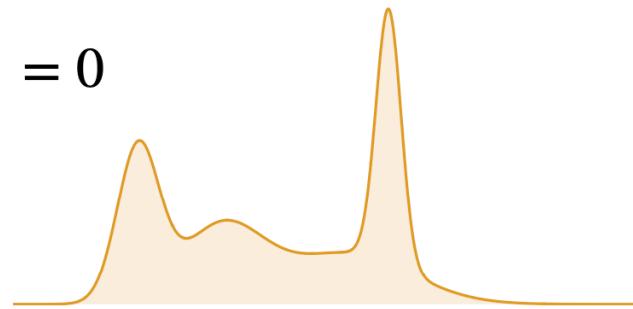


Zhang, E, LW 1809.10188
[wangleiphy/MongeAmpereFlow](https://github.com/wangleiphy/MongeAmpereFlow)



Simple density

$$\frac{\partial \rho(\mathbf{x}, t)}{\partial t} + \nabla \cdot [\rho(\mathbf{x}, t) \mathbf{v}] = 0$$

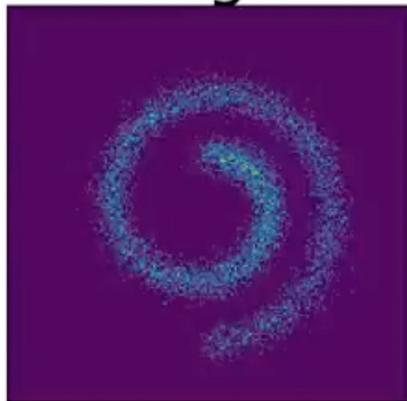


Complex density

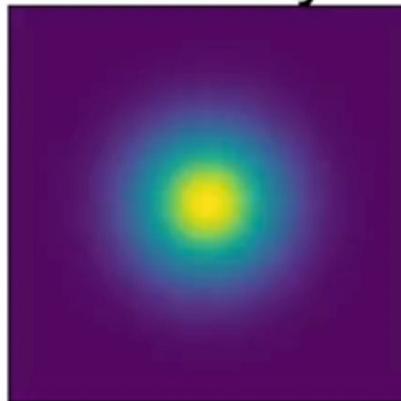
Neural Ordinary Differential Equations

Chen et al, 1806.07366, Grathwohl et al 1810.01367

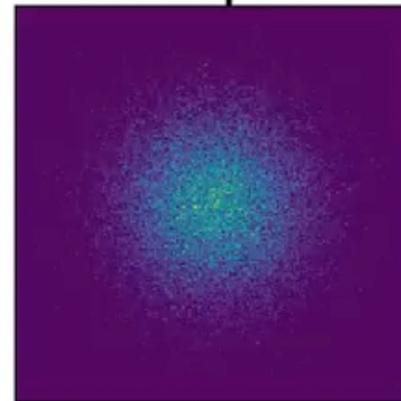
Target



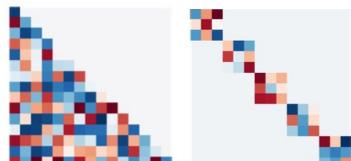
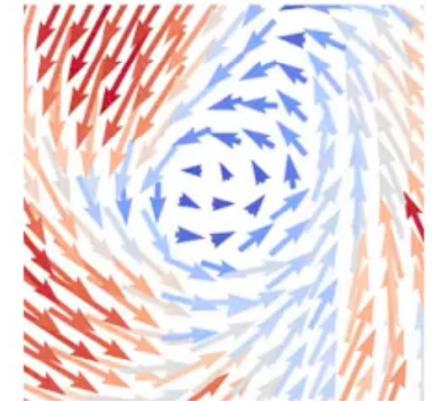
Density



Samples



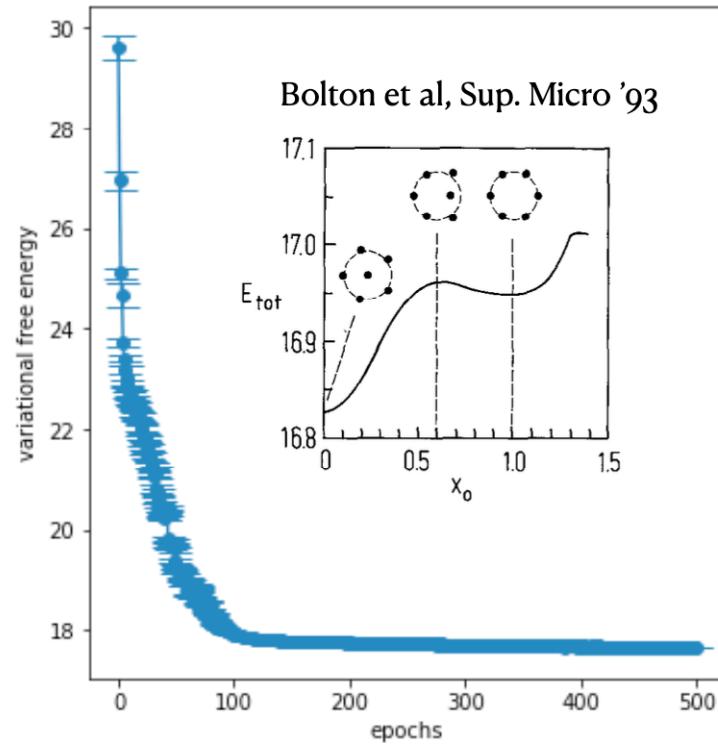
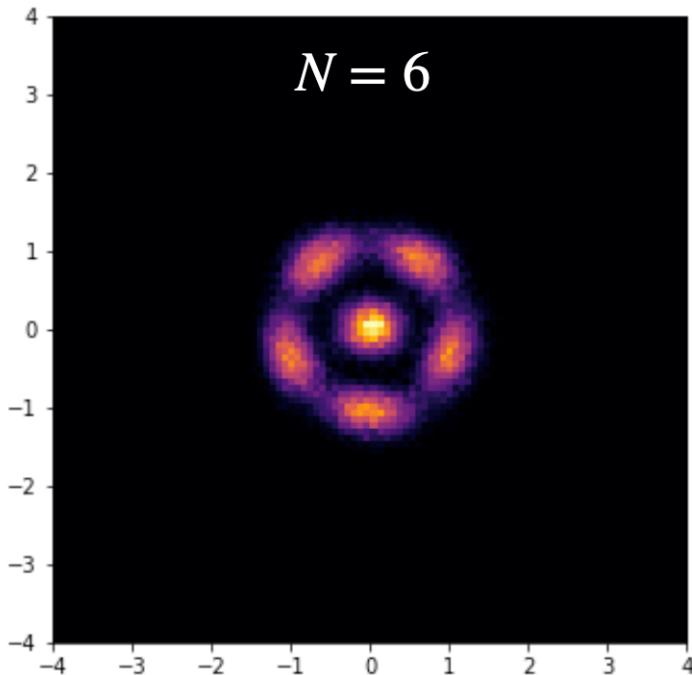
Vector Field



Continuous normalizing flow have no structural constraints on the transformation Jacobian

Demo: Classical Coulomb gas

$$H = \sum_{i < j} \frac{1}{|x_i - x_j|} + \sum_i x_i^2$$

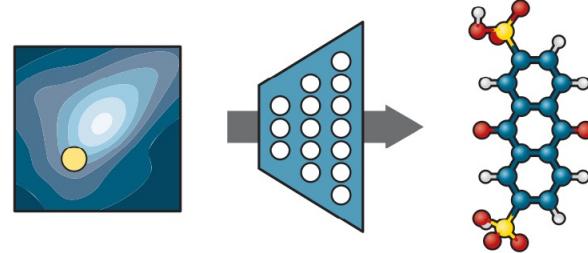


①

How to Build a GPT-3 for Science

Scientific language model

②



Matter inverse design

③

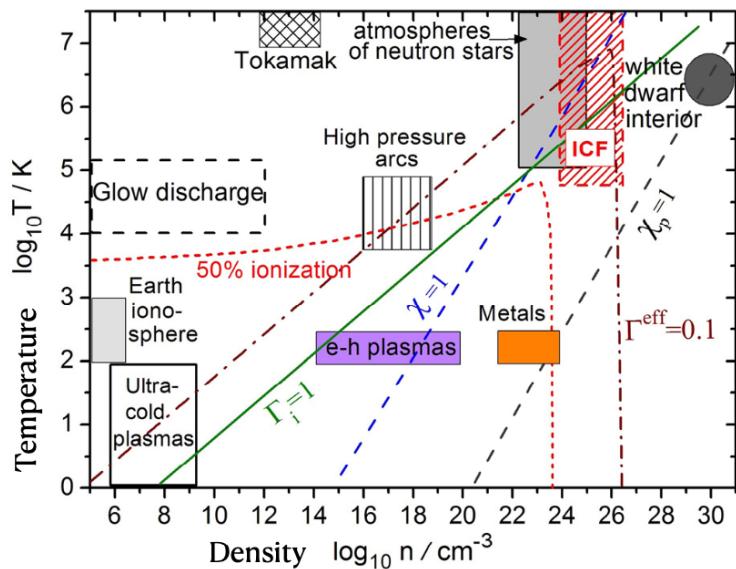
$$F = E - TS$$

Nature's cost function

Ab-initio study of quantum matters at T>0

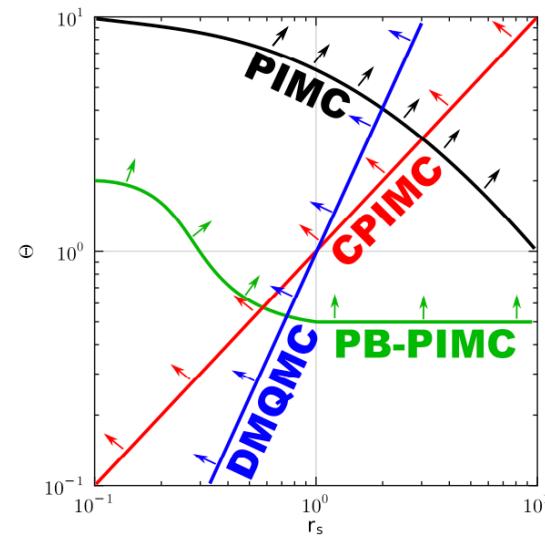
$$H = - \sum_i \frac{\hbar^2}{2m_e} \nabla_i^2 - \sum_I \frac{\hbar^2}{2m_I} \nabla_I^2 - \sum_{I,i} \frac{Z_I e^2}{|R_I - r_i|} + \frac{1}{2} \sum_{i \neq j} \frac{e^2}{|r_i - r_j|} + \frac{1}{2} \sum_{I \neq J} \frac{Z_I Z_J e^2}{|R_I - R_J|}$$

$$Z = \text{Tr}(e^{-H/k_B T})$$



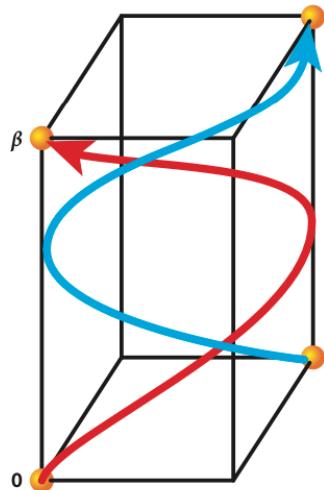
Bonitz et al, Phys. Plasmas '20

Application range
of quantum Monte Carlo



Dornheim et al, Phys. Plasmas '17

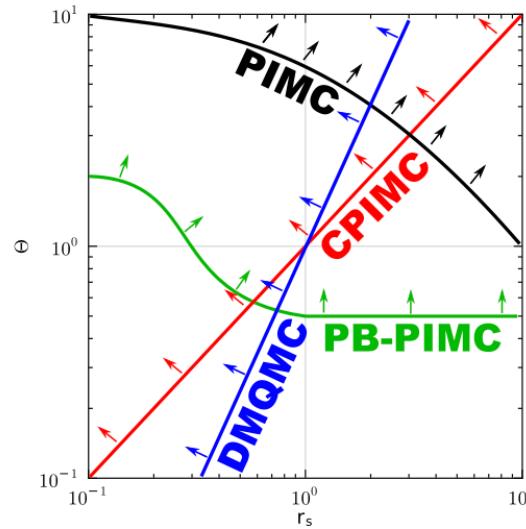
How to solve quantum many-body systems?



Quantum-to-classical mapping via path integral, then we are done

$$Z = \int d\tau dx \dots$$

However, the **sign problem** strikes again: the “weights” may not be positive 😂



Classical to Quantum

Classical world

Probability density p

Kullback-Leibler divergence

$$\mathbb{KL}(p || q)$$

Variational free-energy

$$F = \int dx \left[\frac{1}{\beta} p(x) \ln p(x) + p(x) H(x) \right]$$

Quantum world

Density matrix ρ

Quantum relative entropy

$$S(\rho || \sigma)$$

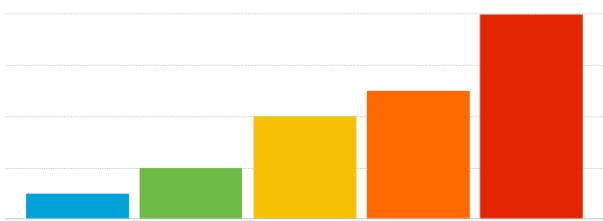
Variational free-energy

$$F = \frac{1}{\beta} \text{Tr}(\rho \ln \rho) + \text{Tr}(\rho H)$$

Density matrix

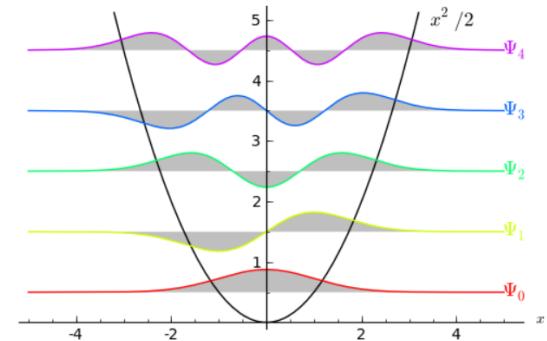
$$\rho = \sum_n p_n |\Psi_n\rangle\langle\Psi_n|$$

Classical probability $0 < p_n < 1$



$$\sum_n p_n = 1$$

Quantum states $\Psi_n(x) = \langle x | \Psi_n \rangle$



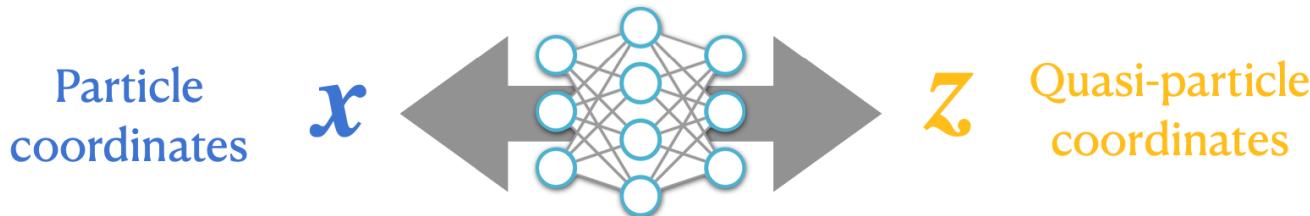
$$\langle \Psi_m | \Psi_n \rangle = \delta_{mn}$$

How to represent them ??

Use TWO deep generative models !!

Square root of normalizing flow

$\sqrt{\text{Normalizing flow}}$



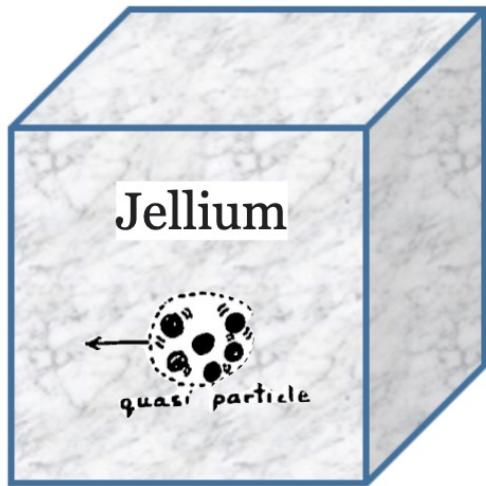
$$\Psi_n(x) = \Phi_n(z) \cdot \left| \det \left(\frac{\partial z}{\partial x} \right) \right|^{\frac{1}{2}}$$

Target states Base states Jacobian of the flow

The flow implements a *learnable* many-body unitary transformation
hence the name “neural canonical transformation” a classical generalization of Li, Dong, Zhang, LW, PRX ‘20

Applications

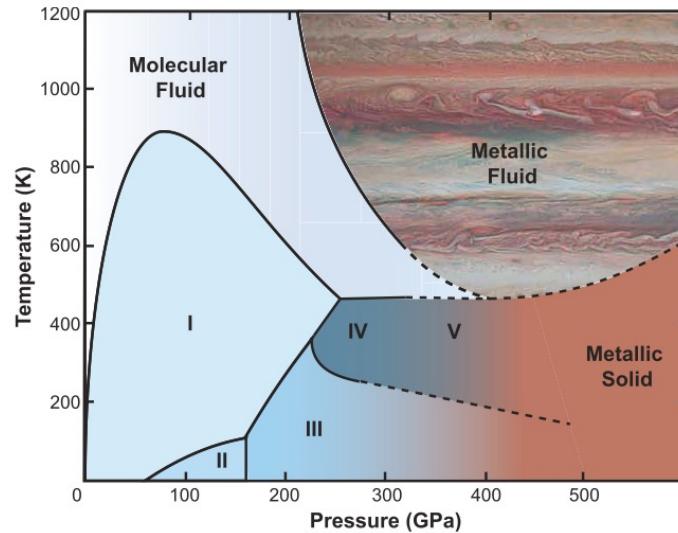
Uniform electron gas



Xie, Zhang, LW, 2201.03156

Dense hydrogen

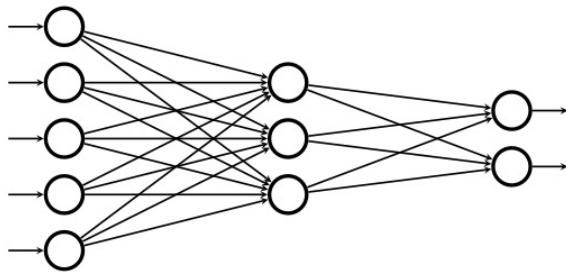
Gregoryanz et al, Matter Radiat. Extremes, 2020



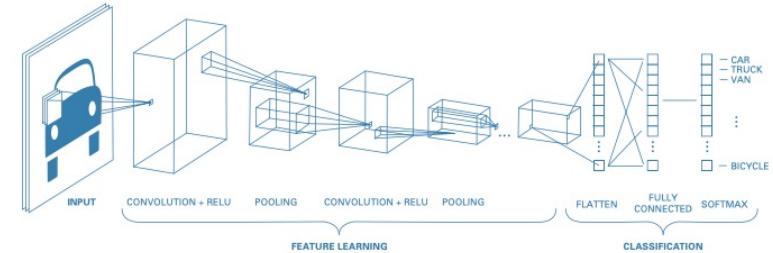
Xie, Li, Wang, Zhang, LW, 2209.06095

Models

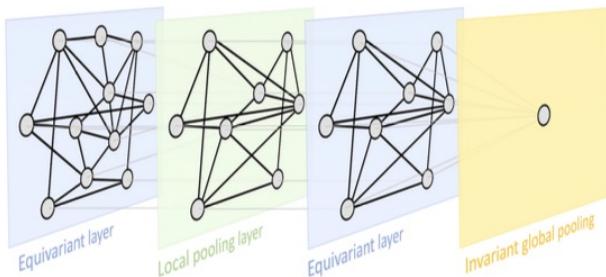
Multi-layer perceptron



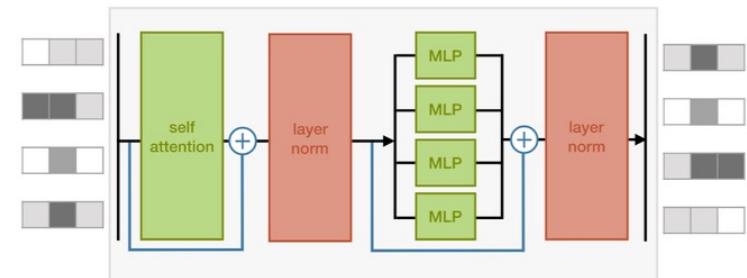
Convolutional neural network



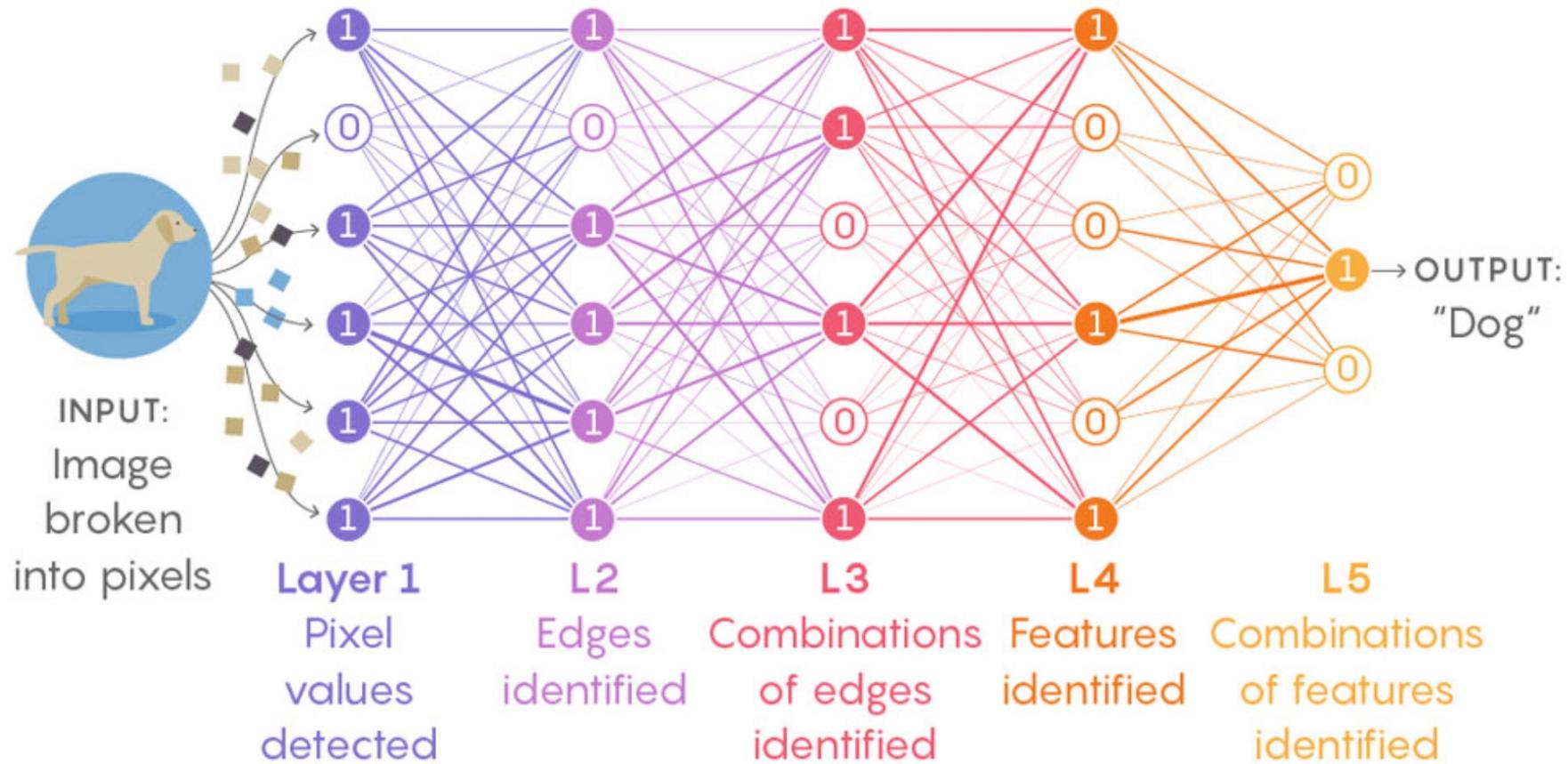
Graph neural network



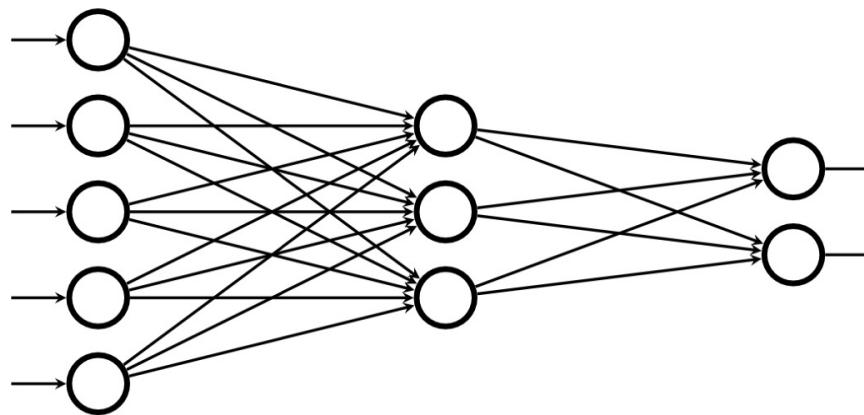
Transformer



Multiple-layer perceptrons



Multiple-layer perceptrons

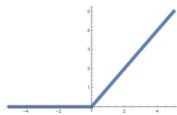


$$\mathbf{y} = \sigma(\mathbf{W}\mathbf{x} + \mathbf{b})$$

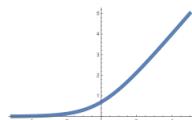
$$\mathbf{x} \in \mathbb{R}^{F_x} \quad \mathbf{W} \in \mathbb{R}^{F_y \times F_x} \quad \mathbf{b} \in \mathbb{R}^{F_y}$$

σ : elementwise activation

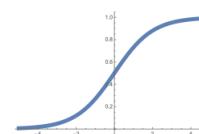
$$\text{ReLU} = \max(0, x)$$



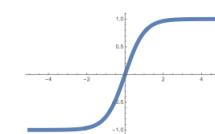
$$\text{Softplus} = \ln(1 + e^x)$$



$$\text{Sigmoid} = 1/(1 + e^{-x})$$



$$\tanh$$

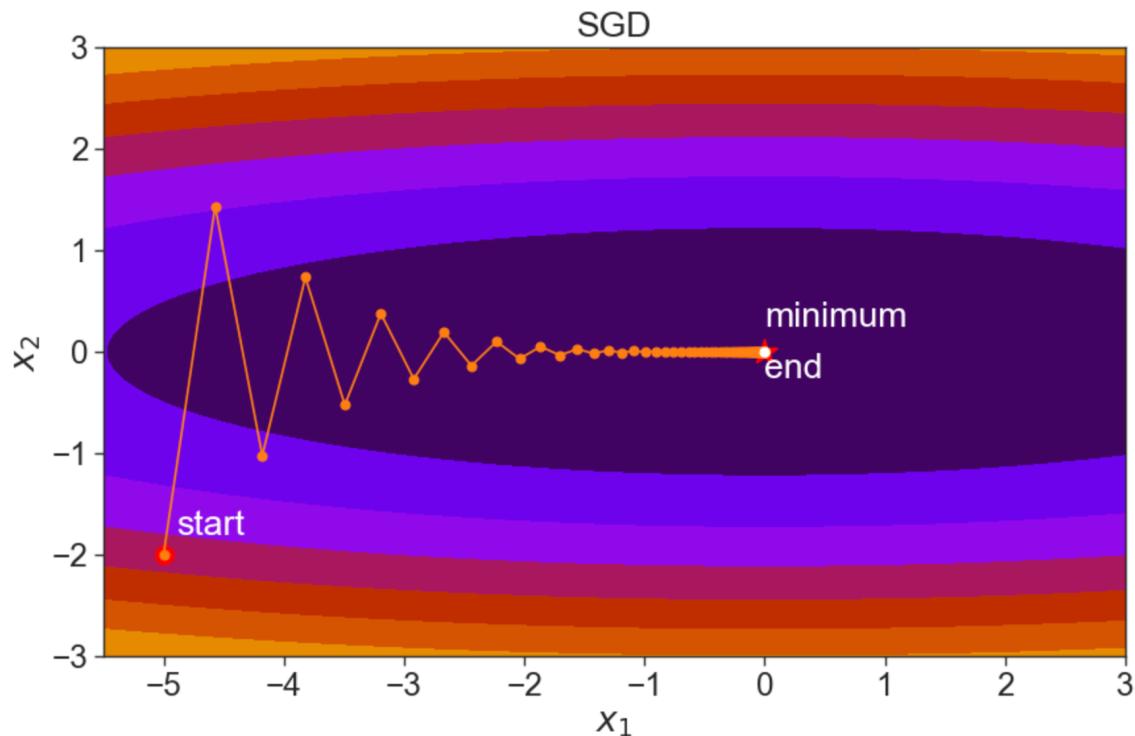


Optimization

Stochastic gradient descent

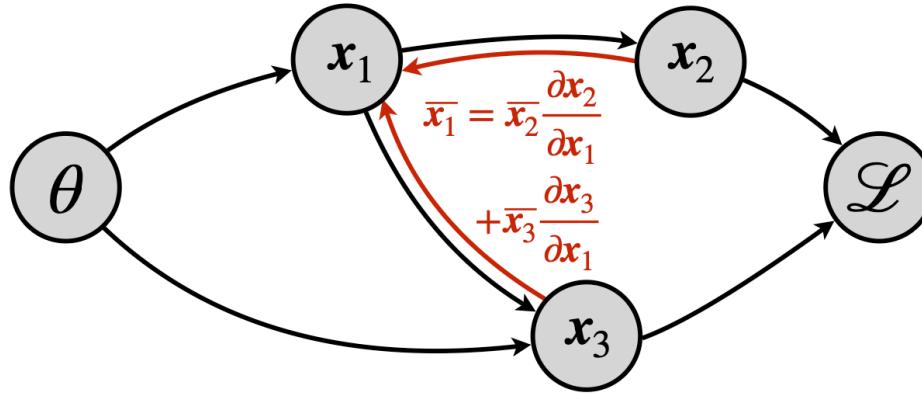
$$\theta \leftarrow \theta - \eta \frac{\partial \mathcal{L}}{\partial \theta}$$

$$\mathcal{L} = \mathbb{E}_x [\ell(x)]$$



Automatic differentiation

directed
acyclic graph



$$\bar{x}_i = \sum_{j: \text{child of } i} \bar{x}_j \frac{\partial x_j}{\partial x_i} \quad \text{with} \quad \overline{\mathcal{L}} = 1$$

Message passing for the adjoint at each node

Reverse mode AD

$$\frac{\partial \mathcal{L}}{\partial \theta} = \underbrace{\frac{\partial \mathcal{L}}{\partial x_n} \frac{\partial x_n}{\partial x_{n-1}} \dots \frac{\partial x_2}{\partial x_1} \frac{\partial x_1}{\partial \theta}}$$

Reverse mode AD: **Vector-Jacobian Product of primitives**

- Backtrace the computation graph
- Needs to store intermediate results
- Efficient for graphs with large fan-in

Backpropagation = Reverse mode AD applied to neural networks

Forward mode AD

$$\frac{\partial \mathcal{L}}{\partial \theta} = \underbrace{\frac{\partial \mathcal{L}}{\partial x_n} \frac{\partial x_n}{\partial x_{n-1}} \cdots \frac{\partial x_2}{\partial x_1} \frac{\partial x_1}{\partial \theta}}$$

Forward mode AD: Jacobian-Vector Product of primitives

- Same order with the function evaluation
- No storage overhead
- Efficient for graph with large fan-out

Less efficient for scalar output, but useful for higher-order derivatives

Differentiable programming tools

HIPS/autograd

PyTorch



SciML



NiLang



Hardware accelerators

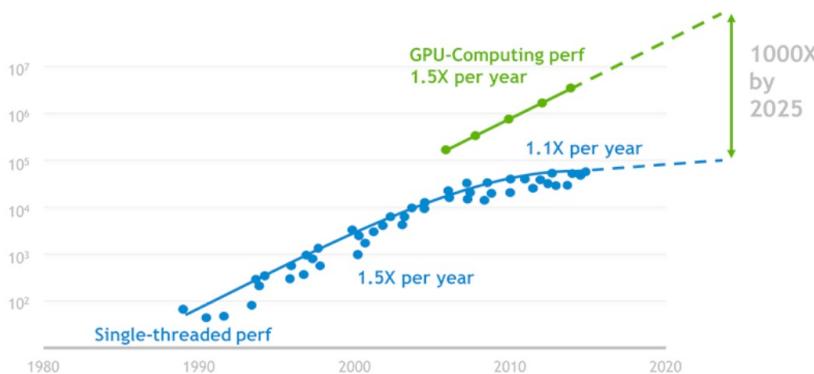
for massive homogeneous computations (e.g. matrix multiplication)



GPU



TPU



The Hardware Lottery

Sara Hooker

Google Research, Brain Team
shooker@google.com

- NumPy and SciPy compatible
- Automatic differentiation (**grad**)
- Just-in-time compilation (**jit**)
- Automatic vectorization (**vmap**)
- Code transformations are composable
- Actively developed by Google
- Gaining a lot of popularity among ML and science researchers



Monte Carlo sampling

算法 1 Metropolis-Hastings sampling algorithm

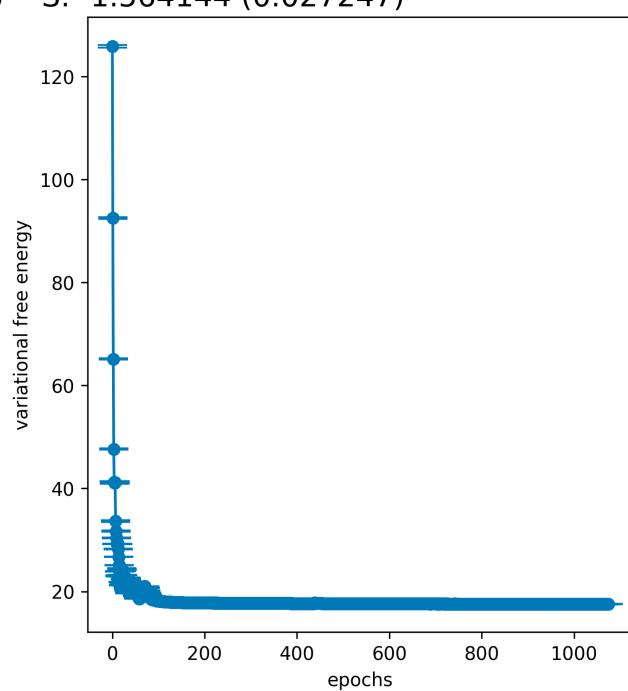
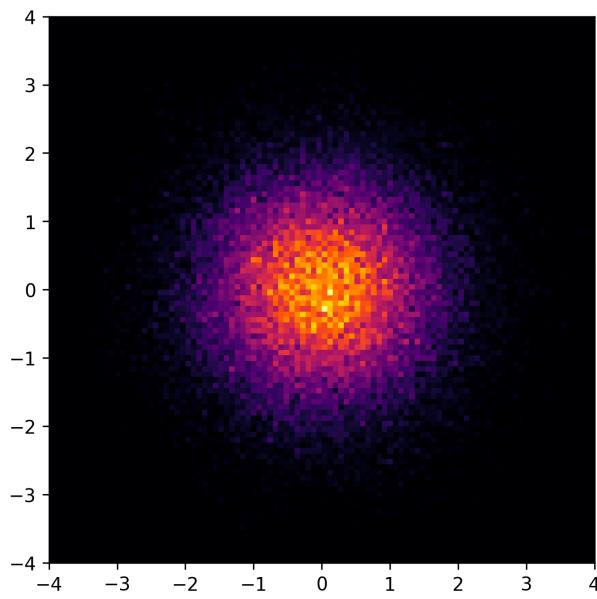
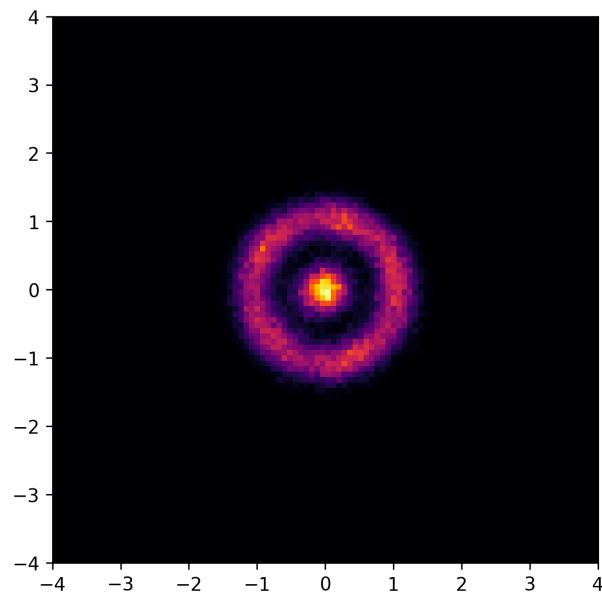
Input: Initial state \mathbf{x}_{init} , total steps N .

Output: Final state $\mathbf{x}_{\text{final}}$.

```
1:  $\mathbf{x} = \mathbf{x}_{\text{init}}$ 
2: for  $i = 1, \dots, N$  do
3:   Sample new state  $\mathbf{x}'$  from the proposal probability  $g(\mathbf{x} \rightarrow \mathbf{x}')$ 
4:    $u \sim \text{Uniform}(0, 1)$ 
5:   if  $u < \min \left( 1, \frac{p(\mathbf{x}')g(\mathbf{x}' \rightarrow \mathbf{x})}{p(\mathbf{x})g(\mathbf{x} \rightarrow \mathbf{x}')} \right)$  then
6:      $\mathbf{x} = \mathbf{x}'$                                  $\triangleright$  accept new sample
7:   end if
8: end for
9: return  $\mathbf{x}$ 
```

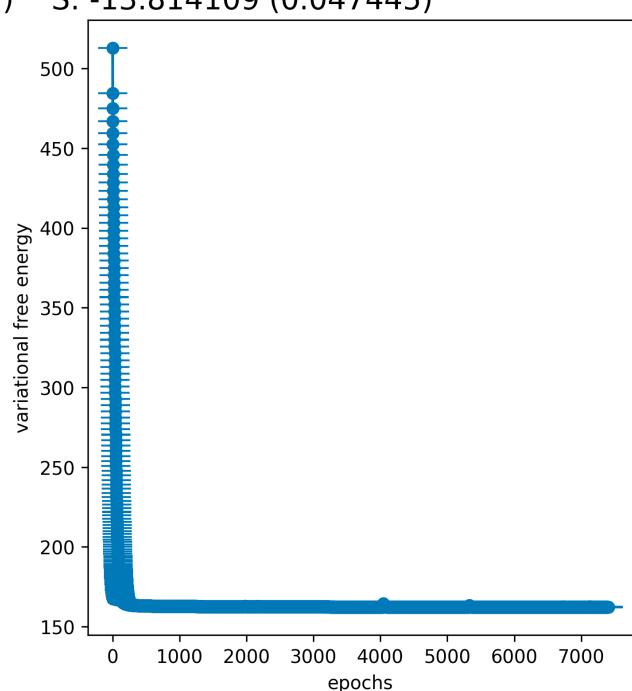
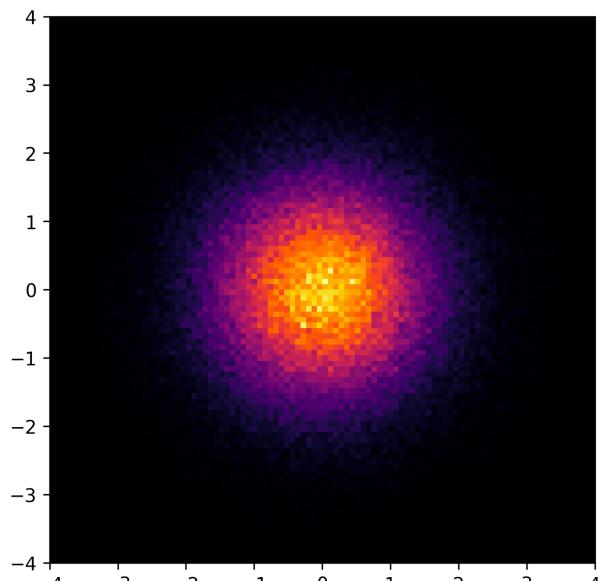
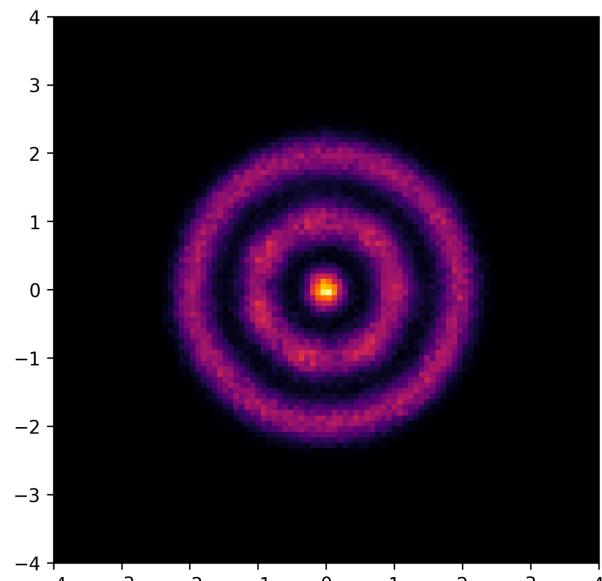
Some questions

epoch: 1074 F: 17.534191 (0.001373) E: 17.377776 (0.003067) S: -1.564144 (0.027247)



Some questions

epoch: 7404 F: 162.236320 (0.003012) E: 160.854909 (0.005717) S: -13.814109 (0.047445)



Thanks for your attention!