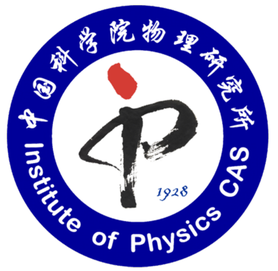
用生成模型解决物理问题

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



Outline

1. 物理学家眼中的生成模型

2. 生成模型四问: 哪种最好？多少先验？能推理吗？懂不懂物理？

3. 物质科学应用举例：晶体材料设计、变分自由能计算

What are the limits of classical systems?

Classical Turing Machines can do much more than we previously thought

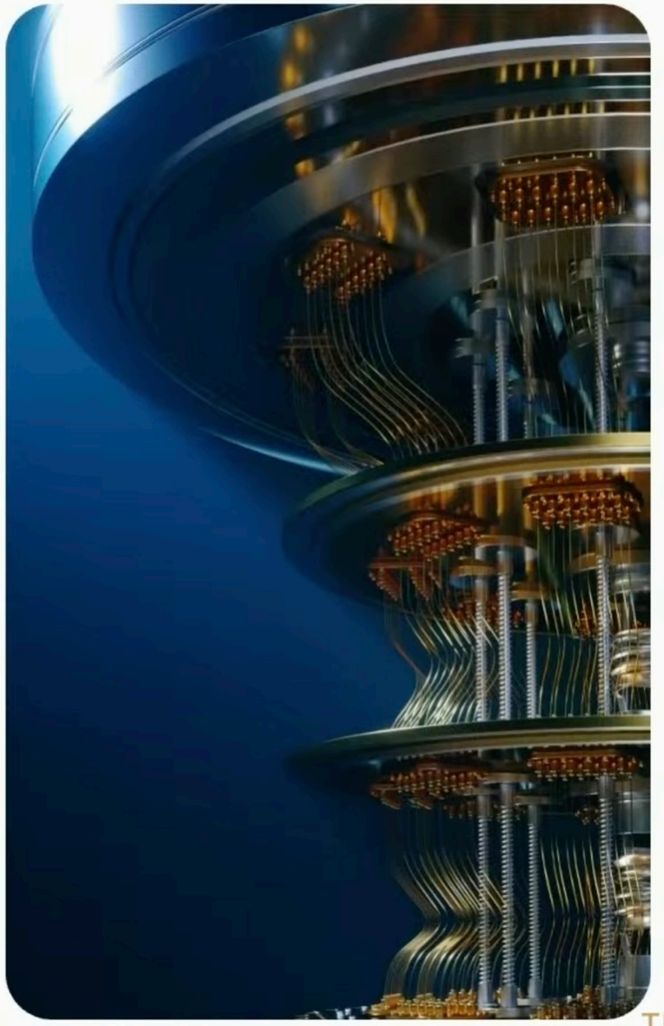
By doing a massive amount of pre-compute upfront to develop a good model

Then use the model to efficiently explore a solution space in polynomial time

My Proposed Conjecture:

"Any pattern that can be generated or found in nature can be efficiently discovered and modelled by a classical learning algorithm"

If it turns out that classical systems can model certain types of quantum systems, it could potentially have big implications for complexity theory including P=NP, and maybe even fundamental physics!



THE NOBEL PRIZE

Demis Hassabis, Nobel prize lecture, 2024 Dec

ChatGPT: Optimizing Language Models for Dialogue
November 30, 2022 — Announcements, Research

DALL·E API Now Available in Public Beta
November 3, 2022 — Announcements, API

DALL·E Now Available Without Waitlist
September 28, 2022 — Announcements

Introducing Whisper
September 21, 2022 — Research

DALL·E: Introducing Outpainting
August 31, 2022 — Announcements

Our Approach to Alignment Research
August 24, 2022 — Research

New and Improved Content Moderation Tooling
August 10, 2022 — Announcements

DALL·E Now Available in Beta
July 20, 2022 — Announcements

OpenAI Technical Goals
June 20, 2016 — Announcements

Generative Models
June 16, 2016 — Research, Milestones

Team Update
May 25, 2016 — Announcements

OpenAI Gym Beta
April 27, 2016 — Research

Welcome, Pieter and Shivon!
April 26, 2016 — Announcements

Team++
March 31, 2016 — Announcements

Introducing OpenAI
December 11, 2015 — Announcements

Generative **P**retrained **T**ransformer

$$\text{text} \sim p(\text{text} \mid \text{prompt})$$

https://openai.com/blog/

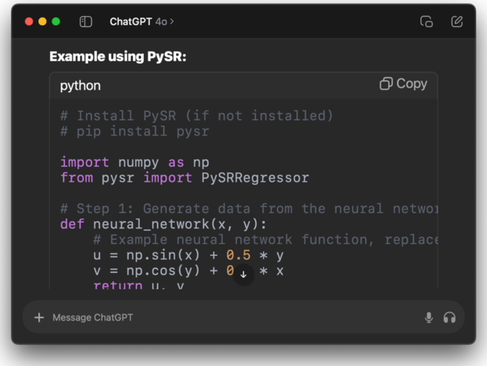
Probabilistic modeling with generative AI

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

pixels, words, atoms, ...

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

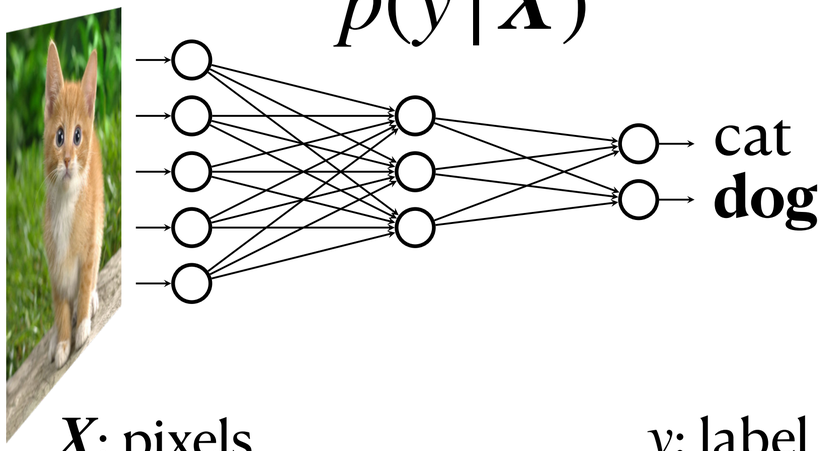
DaLL-E



ChatGPT

AlphaFold3

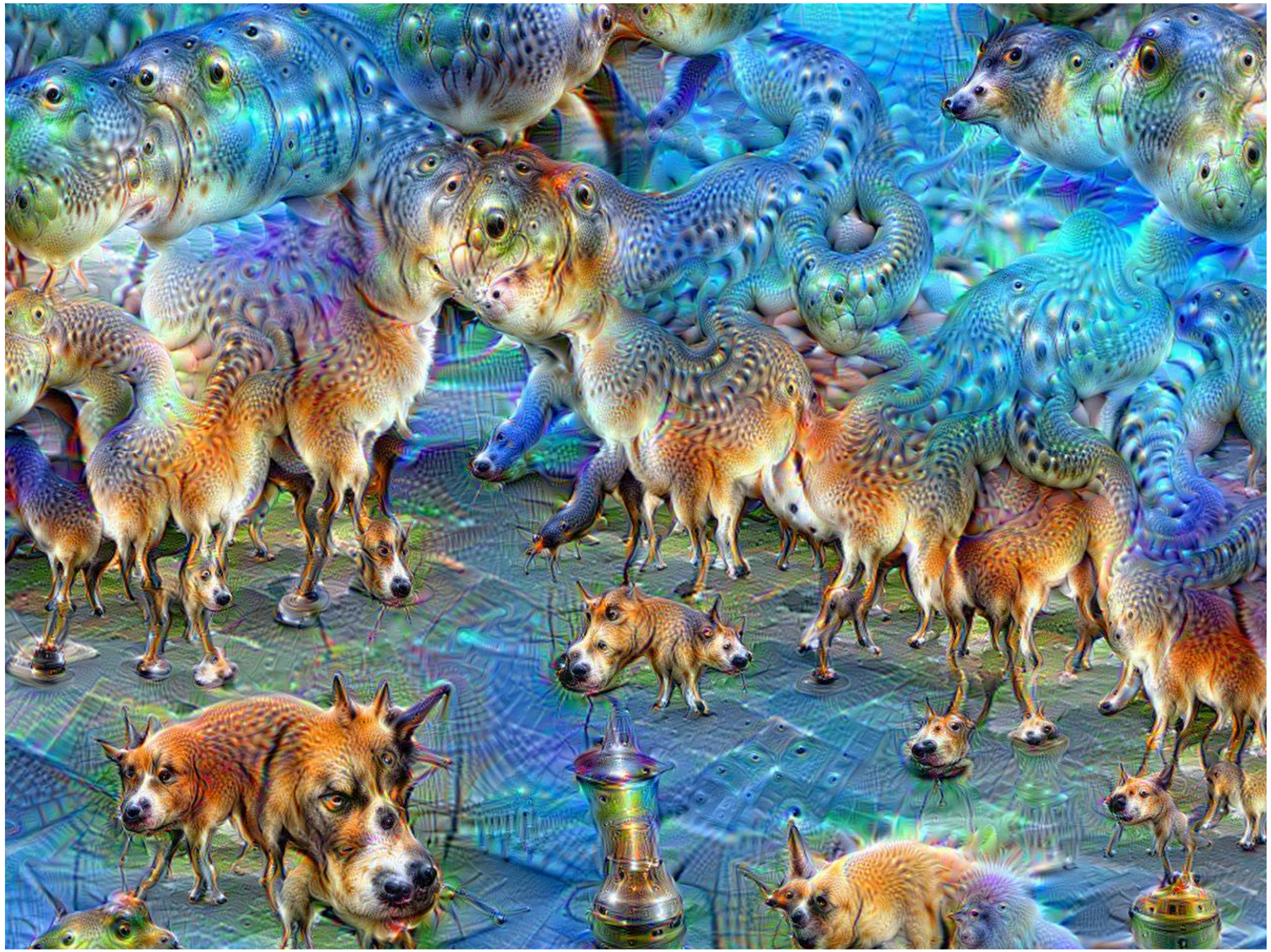
Discriminative AI is not enough



$\boldsymbol{X}$: pixels

y: label

$$\nabla_{\text{pixels}} p(\text{dog} \mid \text{pixels})$$



Deepdream: http://googleresearch.blogspot.ch/2015/06/inceptionism-going-deeper-into-neural.html

Bayes rule

$$p(\boldsymbol{X} \mid y) \propto p(\boldsymbol{X}) p(y \mid \boldsymbol{X})$$

posterior prior likelihood

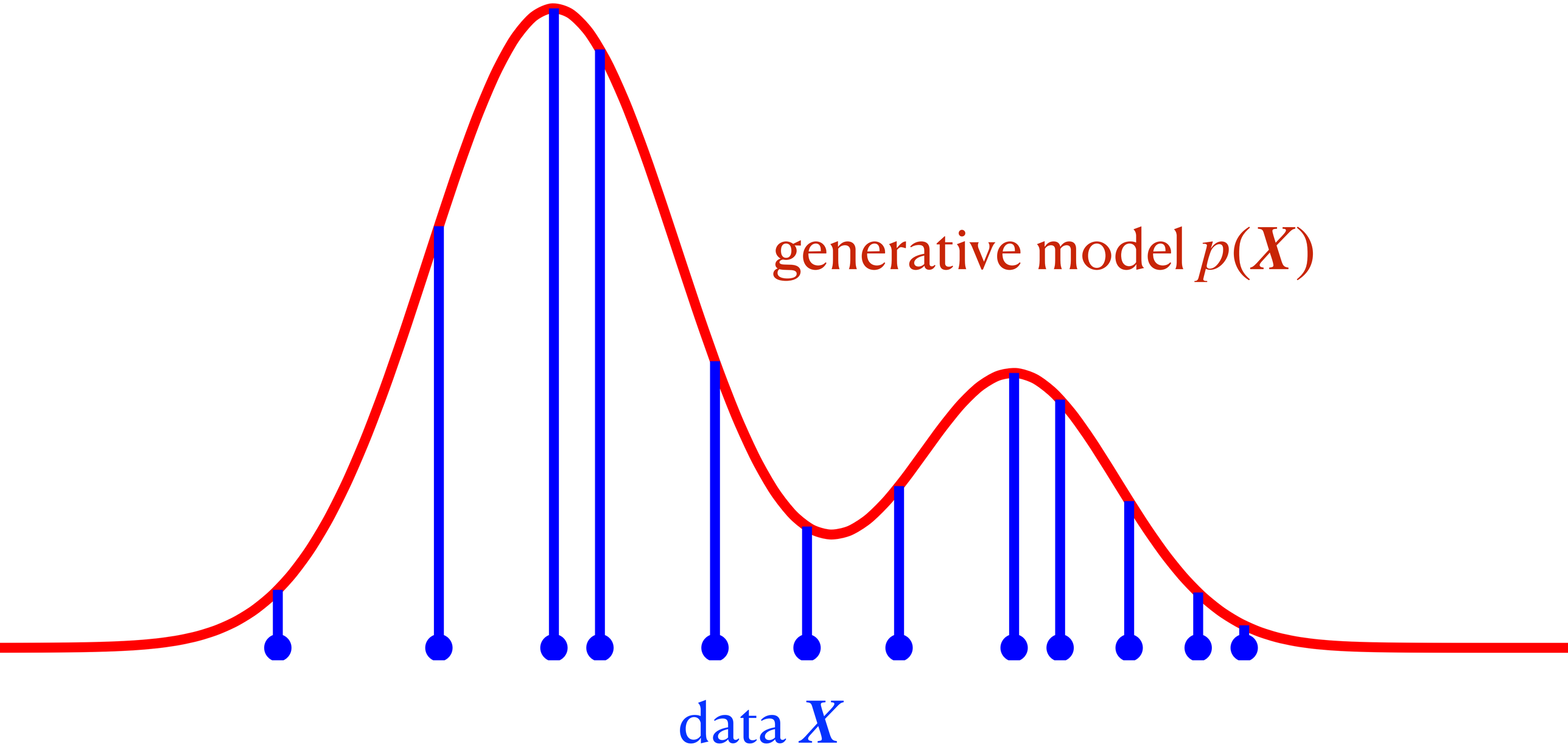
Probability theory 101

Conditional probability $p(y \mid \boldsymbol{X})$

Joint probability $p(\boldsymbol{X}, y)$

Product rule $p(\boldsymbol{X}, y) = p(y \mid \boldsymbol{X}) p(\boldsymbol{X})$

Sum rule $p(\boldsymbol{X}) = \sum_{y} p(\boldsymbol{X}, y)$

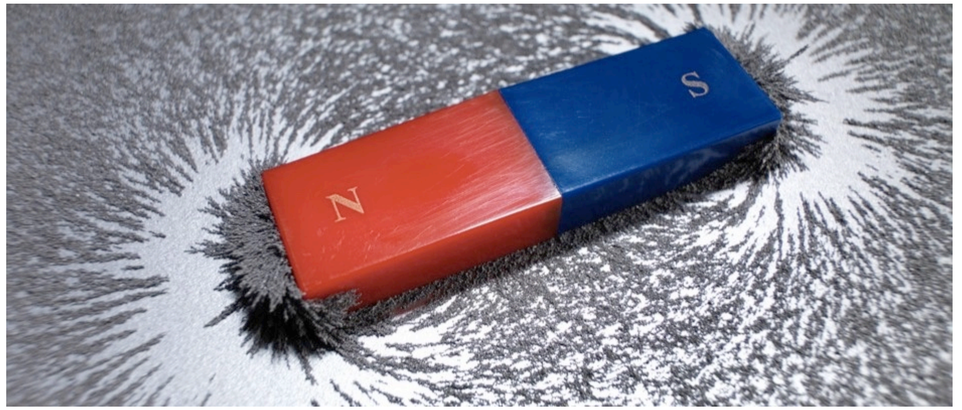

generative model $p(X)$
data X

Two sides of the same coin

Generative modeling

"learn from data"

Statistical physics



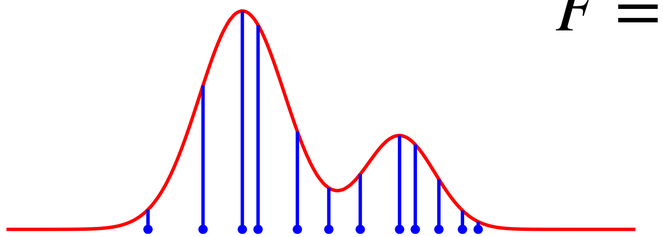
"learn from energy"

Mamixmum likelihood estimation

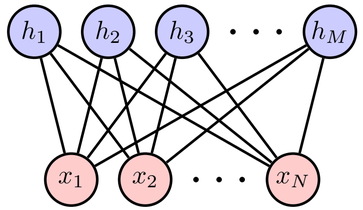
$$\mathscr{L} = -\mathbb{E}_{X\sim\text{data}}\left[\ln p(X)\right]$$

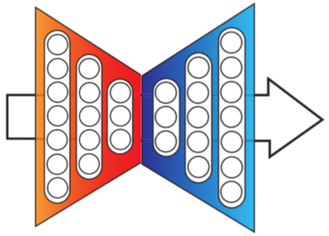
Variational free energy

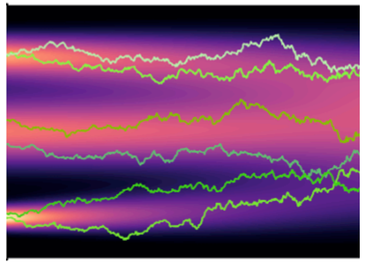
$$F = \mathop{\mathbb{E}}_{X\sim p(X)}\left[E(X) + k_B T \ln p(X)\right]$$

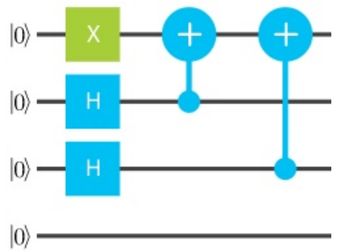


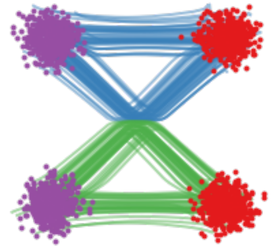
$\mathbb{KL}(\text{data} \parallel p)$ vs $\mathbb{KL}(p \parallel e^{-E/k_B T})$











Boltzmann Machine

1985

Variational Autoencoder

2013

Diffusion Model

2015

Born Machine

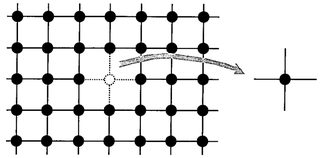
2017

Flow Matching

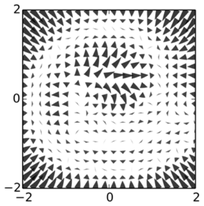
2022

Monte Carlo Ising model

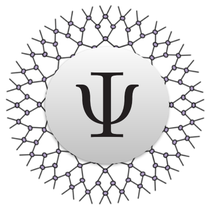
Variational mean field



Nonequilibrium thermodynamics



Tensor networks Quantum circuits



Fluid optimal transportation

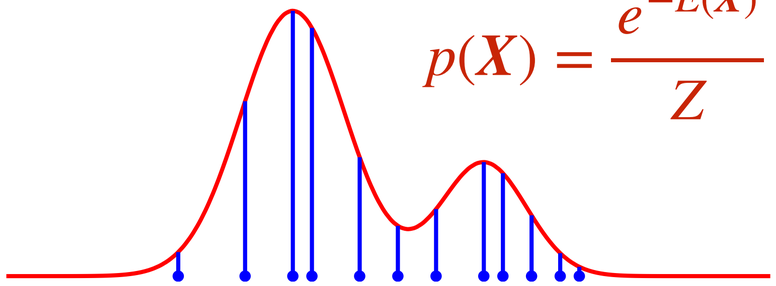
$$\frac{\partial p(\boldsymbol{X},t)}{\partial t} + \nabla \cdot \left[p(\boldsymbol{X},t)\boldsymbol{v}\right] = 0$$

Statistical, quantum, fluid, ... physics insights into generative models

Leverage the power of modern generative models for science

Boltzmann machines

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



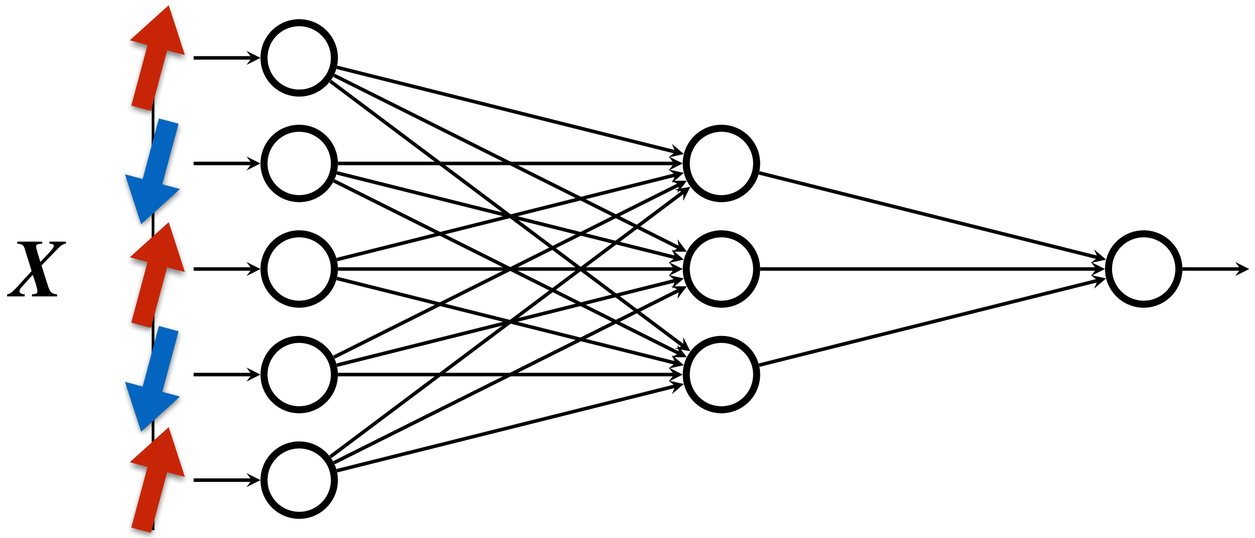
2210.10318

GAUSSIAN-BERNOULLI RBMS WITHOUT TEARS 😂

Renjie Liao[*1], **Simon Kornblith**[2], **Mengye Ren**[3], **David J. Fleet**[2,4,5], **Geoffrey Hinton**[2,4,5]

$$\nabla_\theta \mathscr{L} = \langle \nabla_\theta E \rangle_{\text{data}} - \langle \nabla_\theta E \rangle_{\text{model}}$$

So, why bother ?



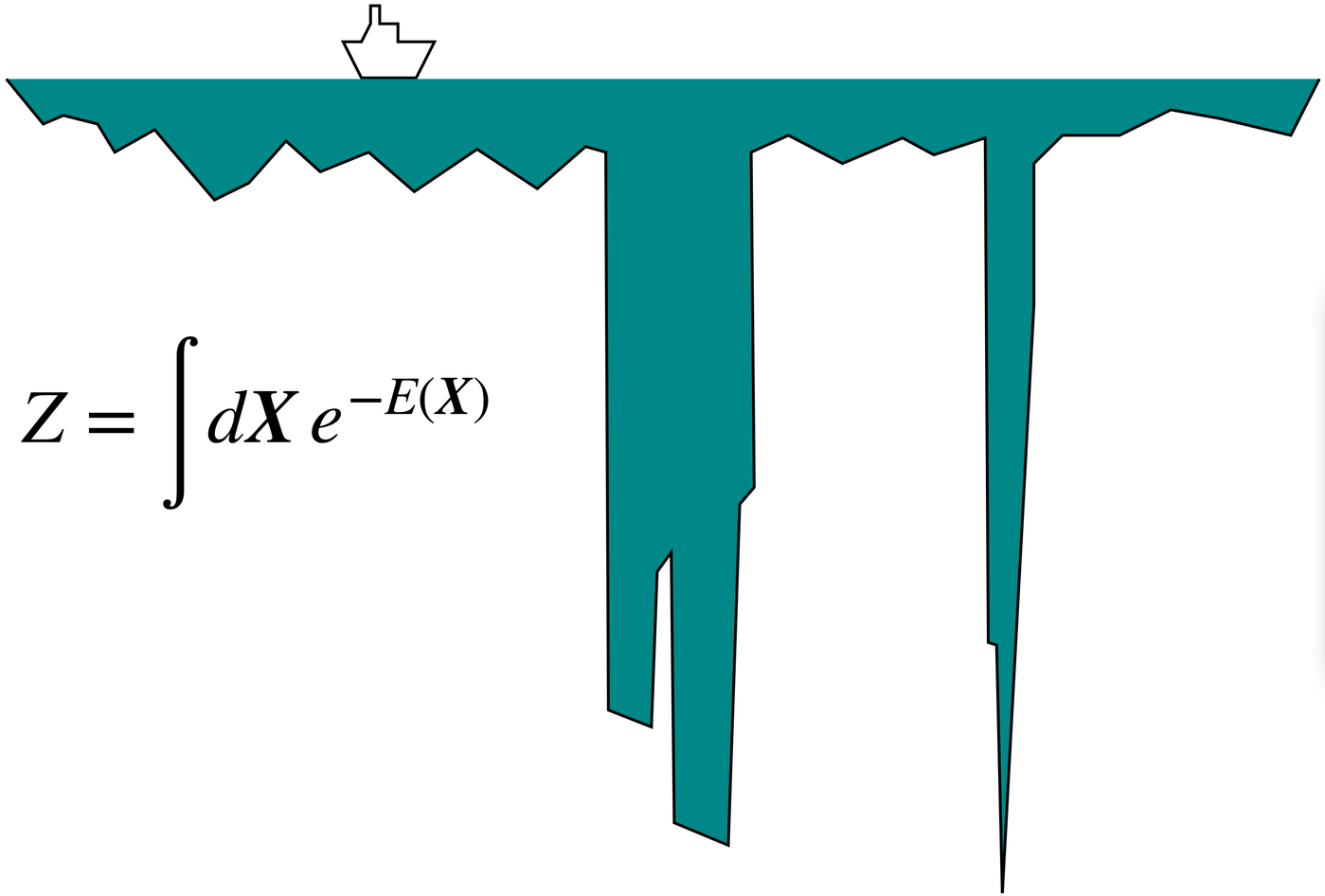
Normalization ?

$$\int dX\, p(X)$$

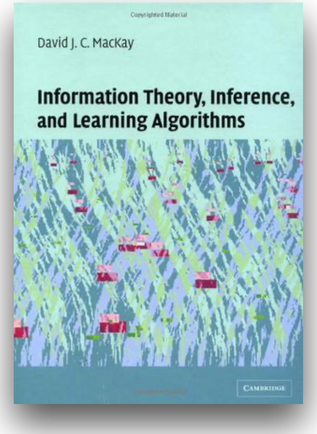
Sampling ?

$$X \sim p(X)$$

归一化难题



$$Z = \int dX\, e^{-E(X)}$$

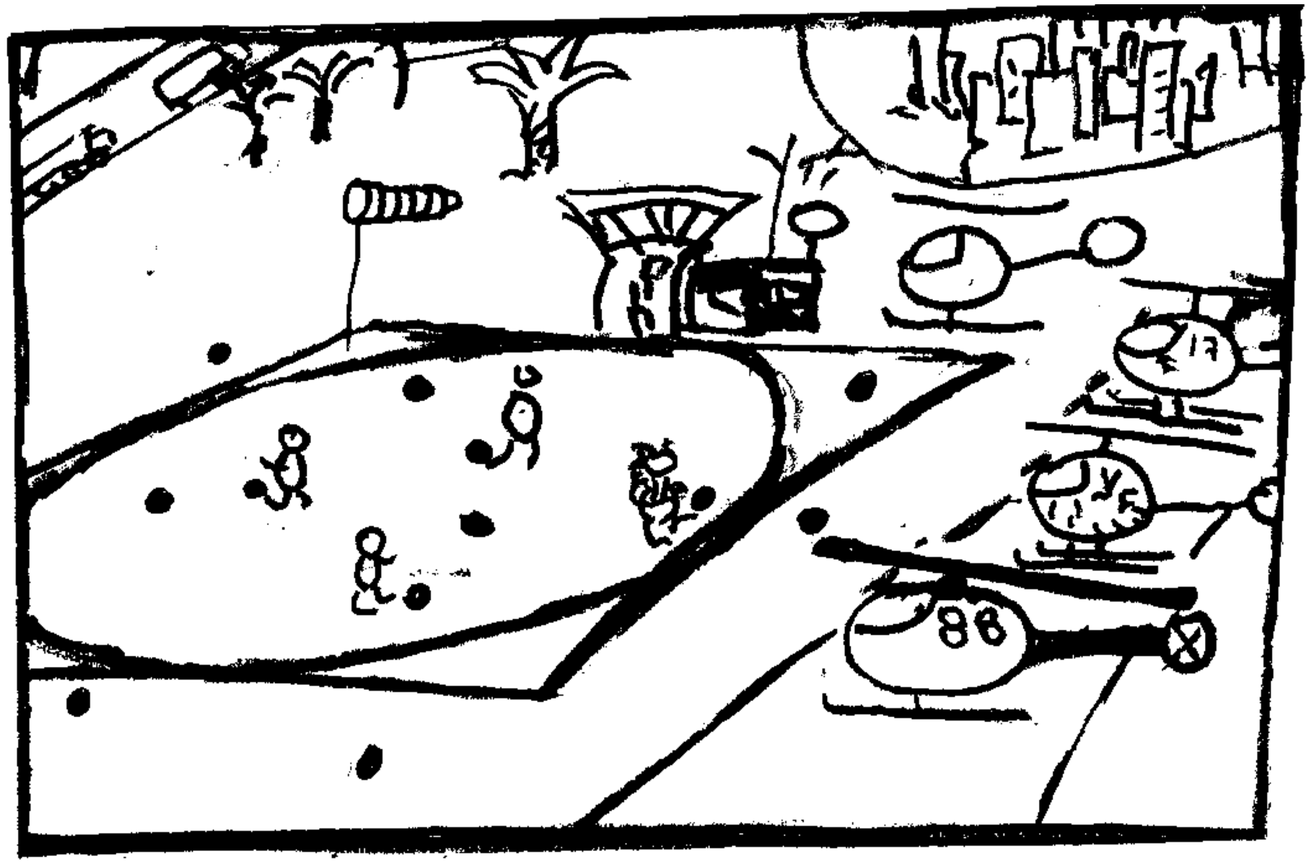


"Intractable" partition function Z

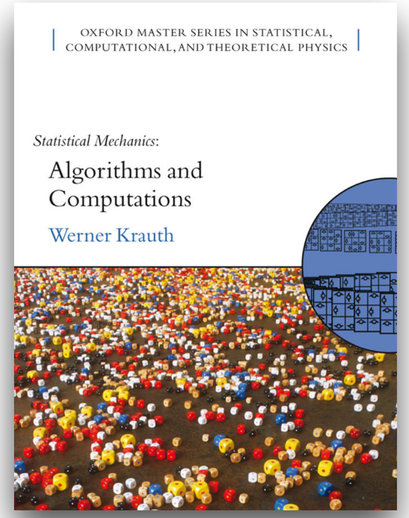
appears widely in machine learning and statistical physics (entropy and free energy calculation)

采样难题

$$X \sim p(X)$$



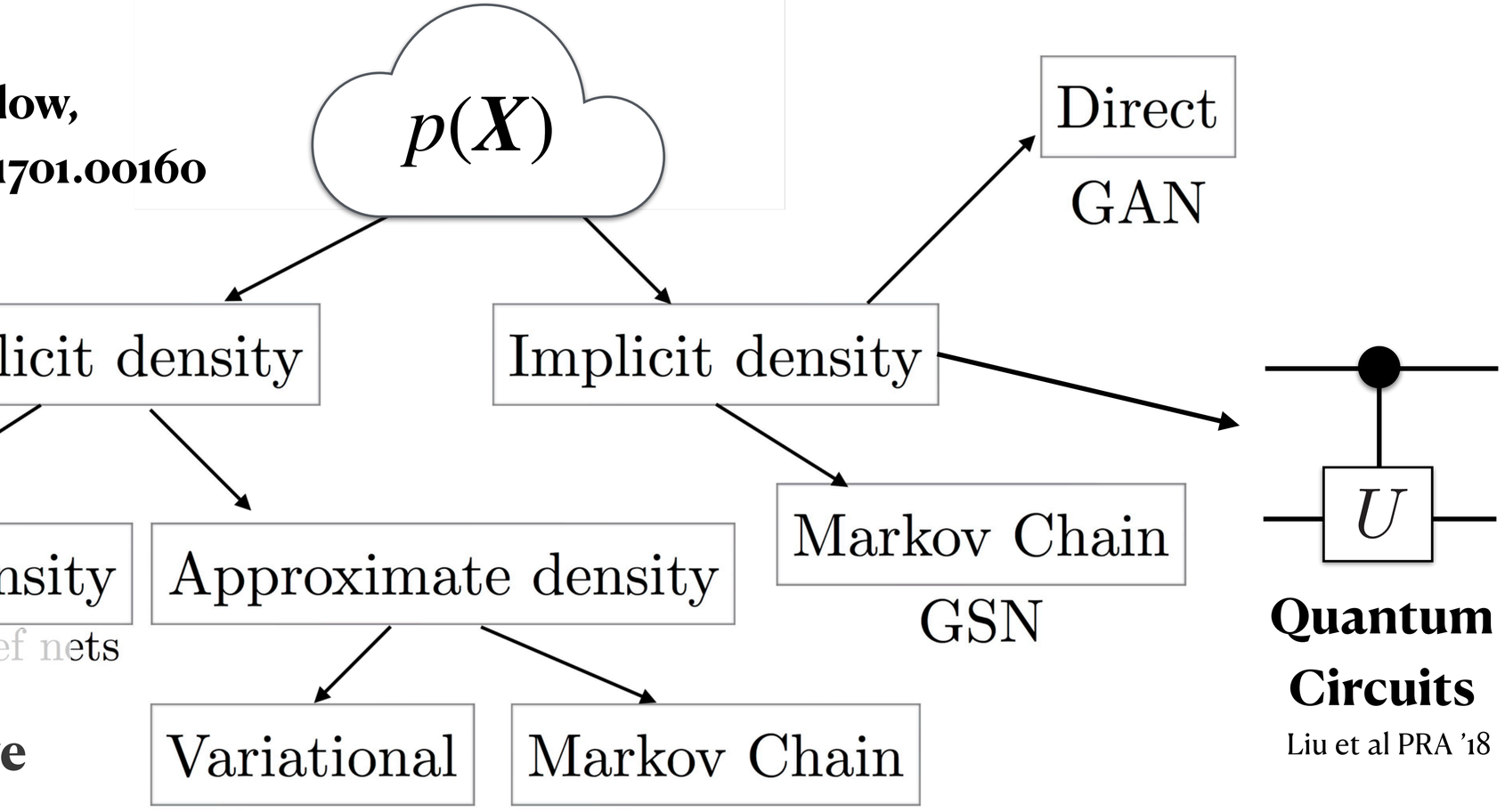
Adults computing the number π at the Monte Carlo heliport.

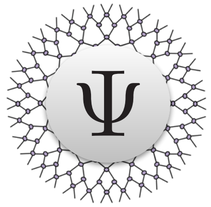


Direct sampling is generally difficult in high-dimensional space

Generative models and their physics genes

Goodfellow,
NIPS tutorial, 1701.00160





Tensor Networks

Han et al, PRX '18

Autoregressive model

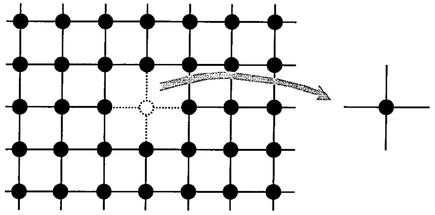
-Fully visible belief nets
-NADE
-MADE
-PixelRNN
-Change of variables models (nonlinear ICA)

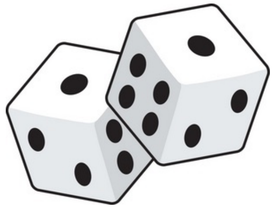
Flow model

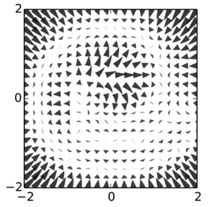
Quantum Circuits

Liu et al PRA '18

+**Diffusion models**





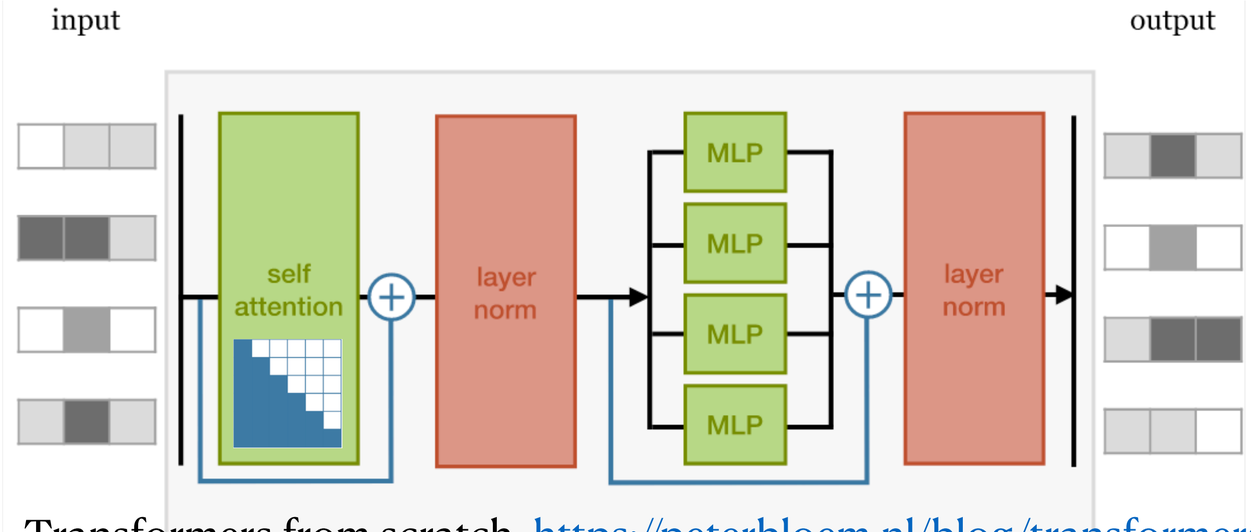


Autoregressive models

$$p(\boldsymbol{X}) = p(x_1)p(x_2 \mid x_1)p(x_3 \mid x_1, x_2)\cdots$$

"... the murderer is ___"

$p(\_ \mid \ldots)$



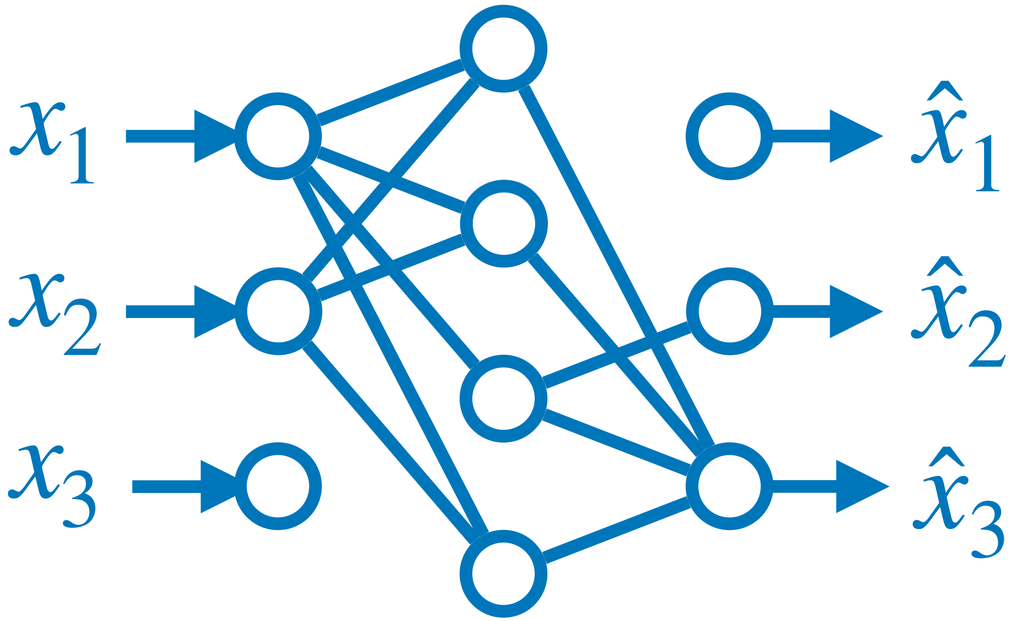
Transformers from scratch, https://peterbloem.nl/blog/transformers

Generative **P**retrained **T**ransformer

Implementation: autoregressive masks

Masked Autoencoder Germain et al, 1502.03509

Bit strings



Probabilities

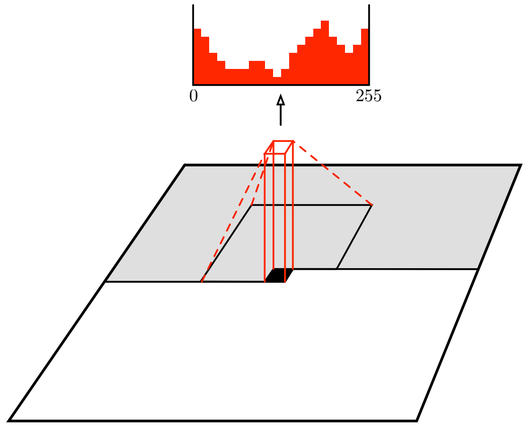
$$p(x_1) = \text{Bernoulli}(\hat{x}_1) \quad p(x_2 \mid x_1) = \text{Bernoulli}(\hat{x}_2) \quad p(x_3 \mid x_1, x_2) = \text{Bernoulli}(\hat{x}_3)$$

Other ways to implement autoregressive models: recurrent networks

Implementation: autoregressive masks

Mask convolutional kernel

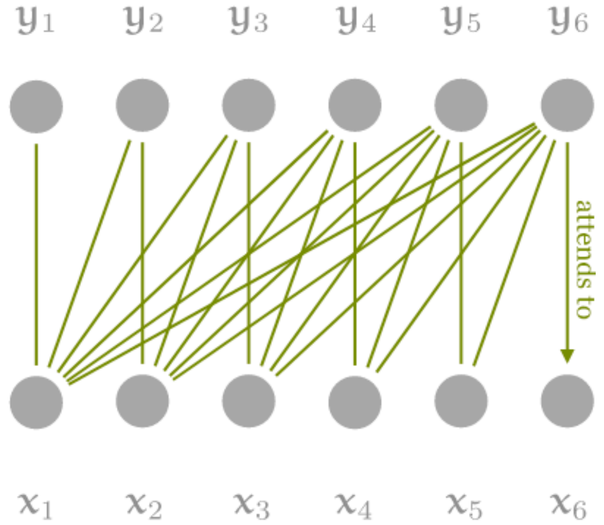
PixelCNN, van den Oord et al, 1601.06759



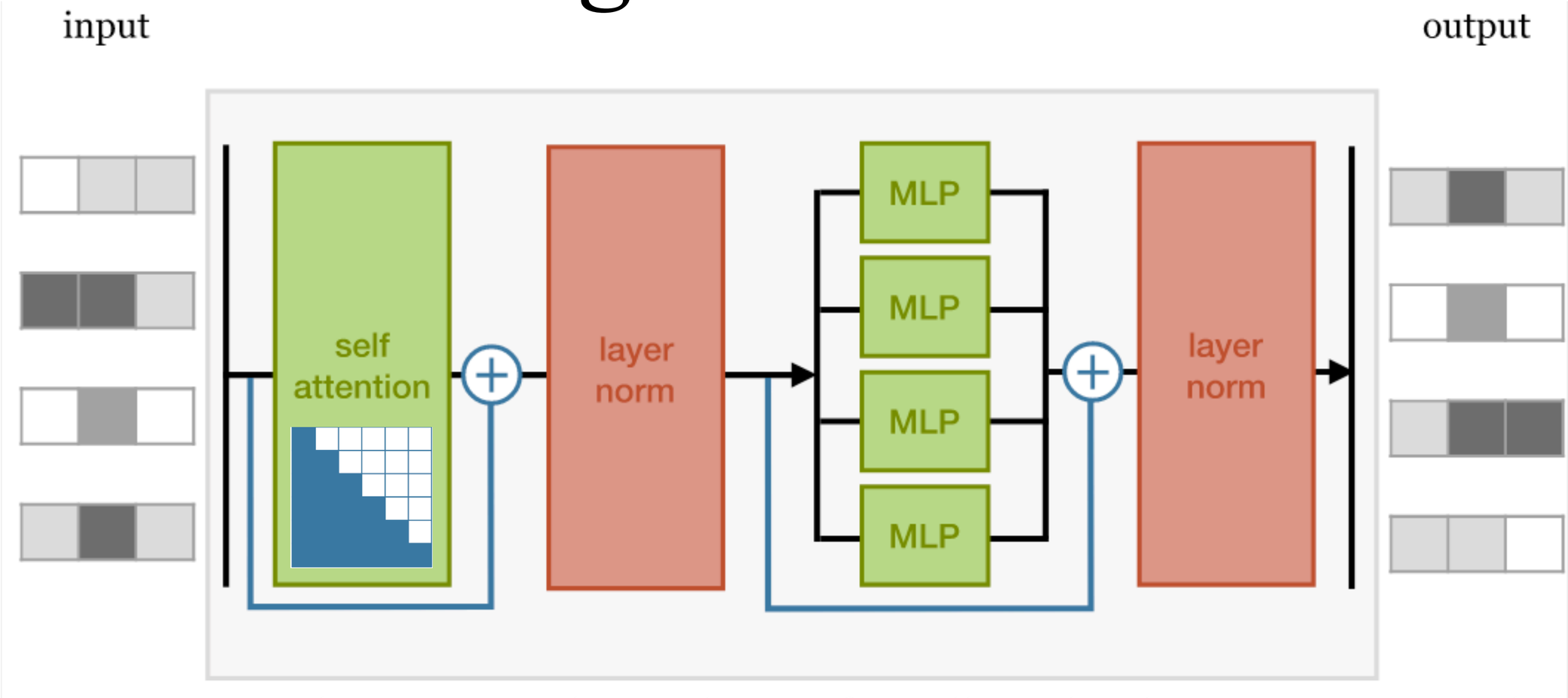
1	1	1	1	1
1	1	1	1	1
1	1	0	0	0
0	0	0	0	0
0	0	0	0	0

Mask self-attention matrix

Causal transformer, Vaswani et al 1706.03762



The autoregressive transformer



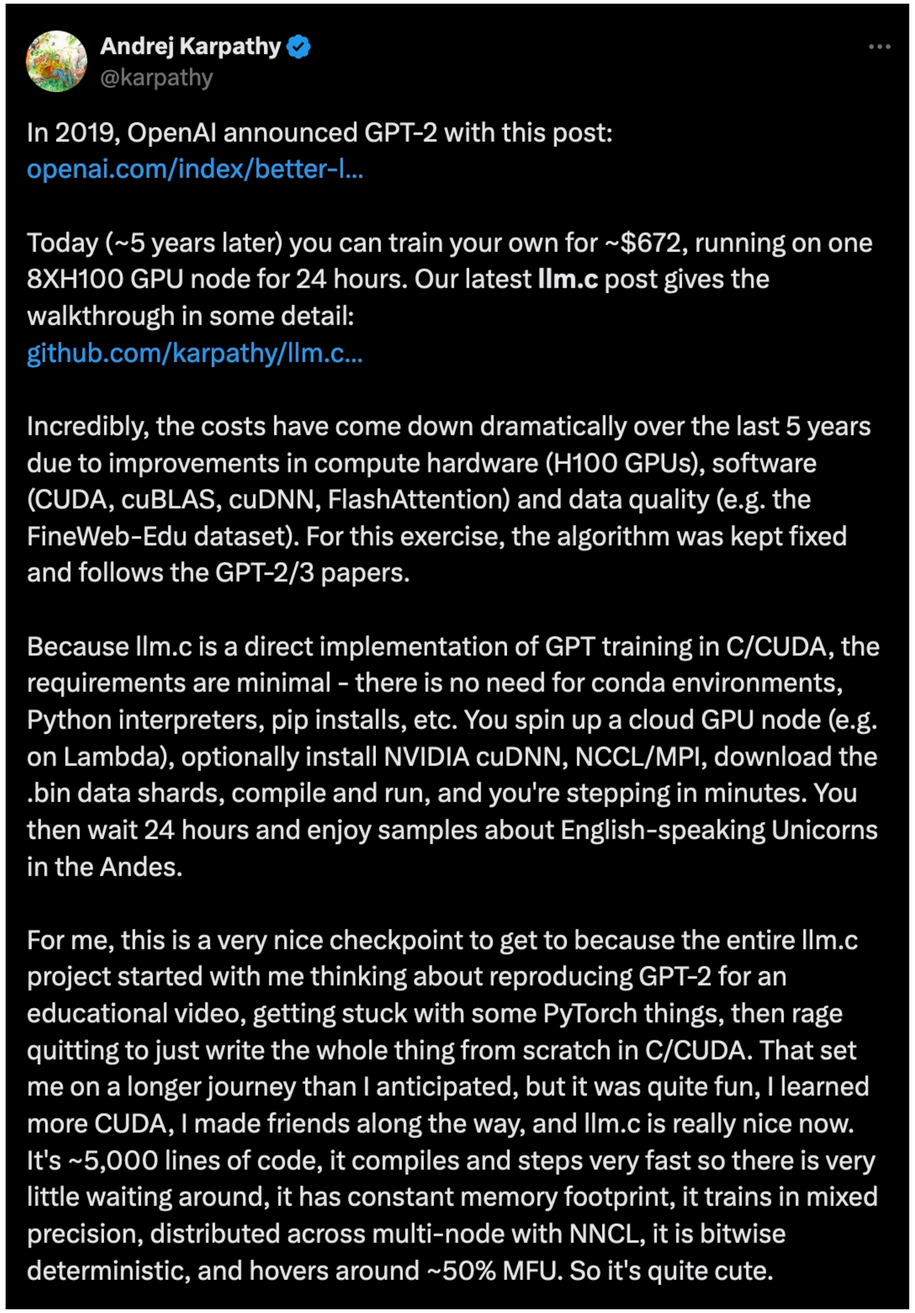
Great at capturing long-range dependence; friendly to backpropagation and GPUs

Masked attention matrix => lower triangular Jacobian matrix => autoregressive model

Transformers from scratch, https://peterbloem.nl/blog/transformers

GPT2 in 60 lines of numpy

https://jaykmody.com/blog/gpt-from-scratch



https://x.com/karpathy/status/1811467135279104217

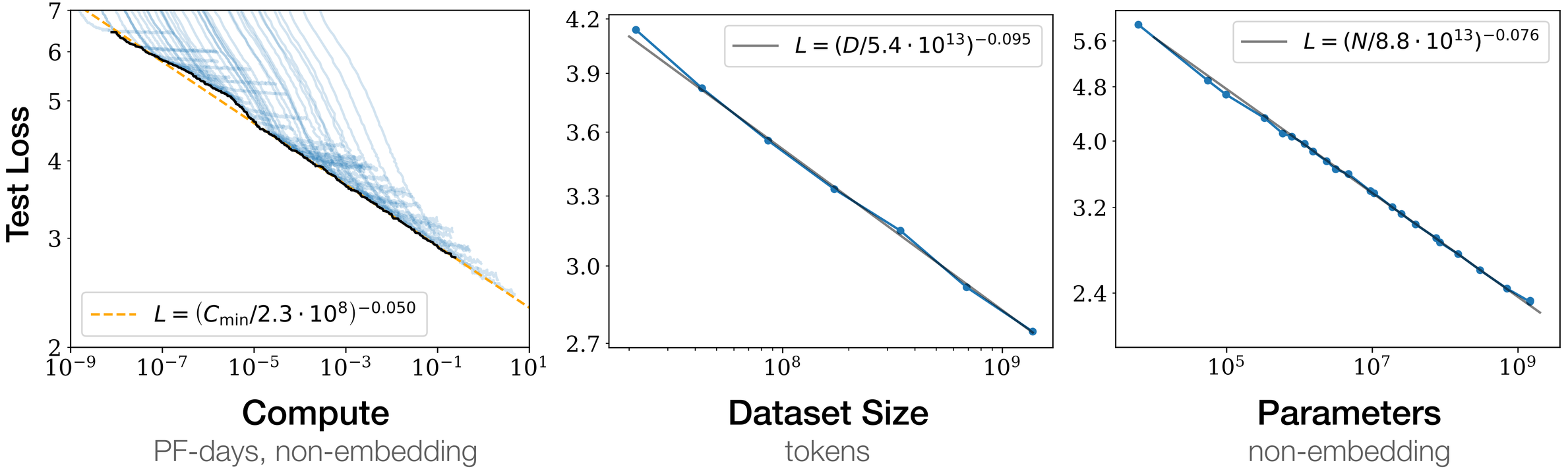
Or, ~1000 lines of C

https://github.com/karpathy/llm.c

Scaling law of the loss function

$$\mathscr{L} = -\,\mathbb{E}_{X\sim\text{data}}\left[\ln p(X)\right]$$

Kaplan et al, 2001.08361



"It would also be exciting to find a theoretical framework from which the scaling relations can be derived: a 'statistical mechanics' underlying the 'thermodynamics' we have observed."

Kullback–Leibler divergence

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

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

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

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

Learn from data

$$\pi(\boldsymbol{x}) \propto \sum_{\boldsymbol{d} \in \text{data}} \delta(\boldsymbol{x} - \boldsymbol{d})$$

$$\min_{\theta} \mathbb{KL}(\pi \parallel p_\theta) \iff \min_{\theta} \left\{ -\mathbb{E}_{\boldsymbol{x} \sim \text{data}} \left[\ln p_\theta(\boldsymbol{x}) \right] \right\}$$

target model Maximum likelihood estimation

The lower bound is the entropy of the dataset: complete memorization

Learn from Energy

$$\pi(\boldsymbol{x}) \propto e^{-E/k_B T}$$

$$\min_{\theta} \mathbb{KL}(p_\theta \,\|\, \pi) \iff \min_{\theta} \left\{ \mathop{\mathbb{E}}_{\boldsymbol{x}\sim p_\theta(\boldsymbol{x})} \left[E(\boldsymbol{x}) + k_B T \ln p_\theta(\boldsymbol{x}) \right] \right\}$$

model target Variational free energy

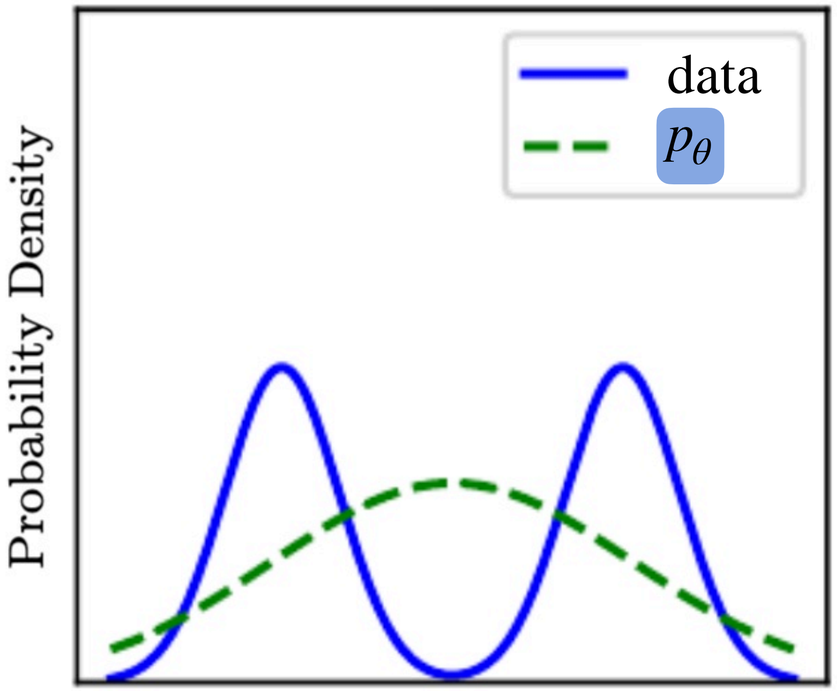
The lower bound is the true free energy: exact solution

Forward KL or Reverse KL ?

Maximum likelihood estimation

$$\min_\theta \mathbb{KL}(\text{data} \parallel p_\theta)$$

Mode covering

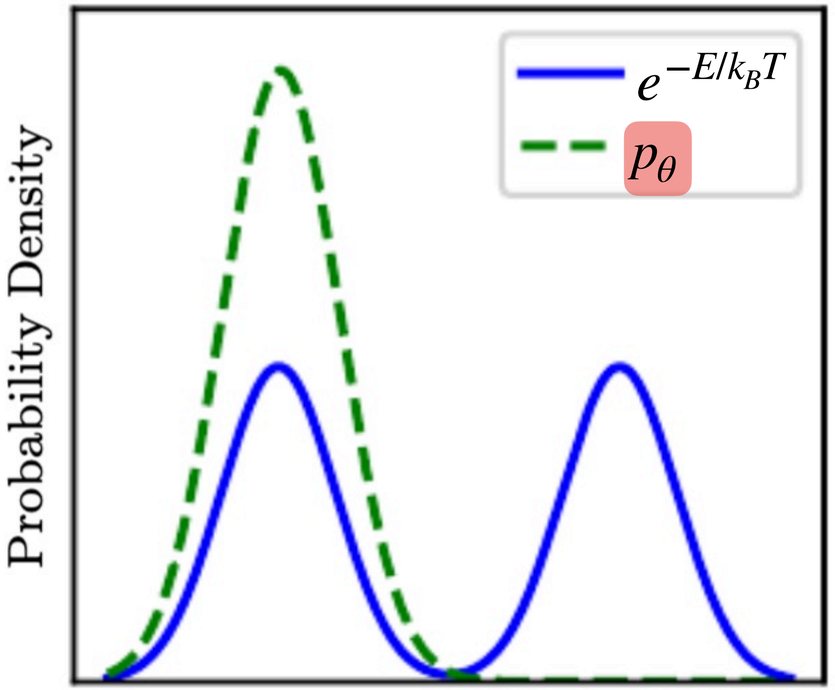


Failure mode: hallucination

Variational free energy

$$\min_\theta \mathbb{KL}(p_\theta \parallel e^{-E/k_BT})$$

Mode seeking



Failure mode: local minima

Goodfellow et al, Deep Learning

The training objective of LLM

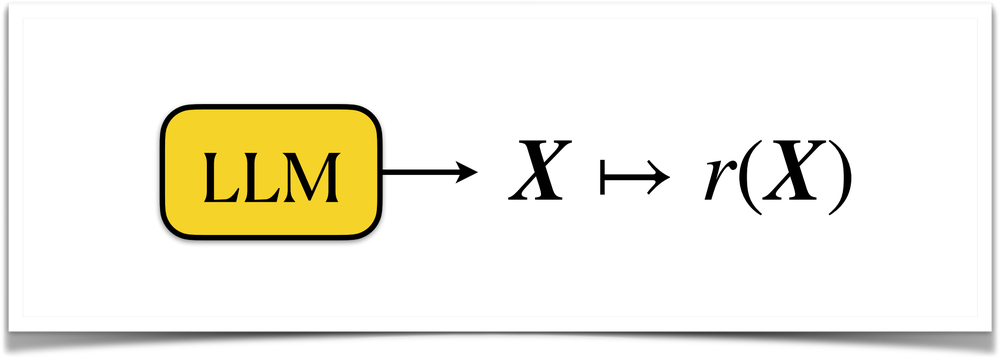
❶ Pretrain with forward KL

learn from data
to be a generalist

$$\mathscr{L} = -\mathbb{E}_{X\sim\text{data}}\left[\ln p(X)\right]$$

学而不思则罔

❷ Finetune with reverse KL



learn from reward
to be a specialized generalist

$$F = \mathop{\mathbb{E}}_{X\sim p(X)}\left[r(X) + T\ln p(X)\right]$$

思而不学则殆

—《论语·大模型》

RL with KL penalties is better viewed as Bayesian inference, Korbar et al, 2205.11275

Andrej Karpathy
@karpathy

It's a bit sad and confusing that LLMs ("Large Language Models") have little to do with language; It's just historical. They are highly general purpose technology for statistical modeling of token streams. A better name would be Autoregressive Transformers or something.

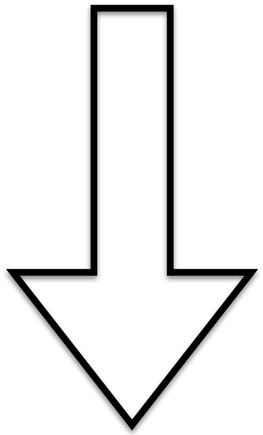
They don't care if the tokens happen to represent little text chunks. It could just as well be little image patches, audio chunks, action choices, molecules, or whatever. If you can reduce your problem to that of modeling token streams (for any arbitrary vocabulary of some set of discrete tokens), you can "throw an LLM at it".

Actually, as the LLM stack becomes more and more mature, we may see a convergence of a large number of problems into this modeling paradigm. That is, the problem is fixed at that of "next token prediction" with an LLM, it's just the usage/meaning of the tokens that changes per domain.

If that is the case, it's also possible that deep learning frameworks (e.g. PyTorch and friends) are way too general for what most problems want to look like over time. What's up with thousands of ops and layers that you can reconfigure arbitrarily if 80% of problems just want to use an LLM?

I don't think this is true but I think it's half true.

~~Language~~



Token streams

Autoregressive model is more than language modeling

"Language" => token stream => bitstream => **ANYTHING**

Speech: WaveNet 1609.03499

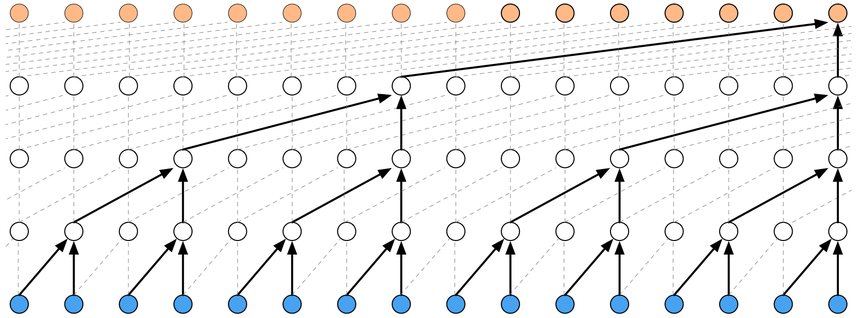
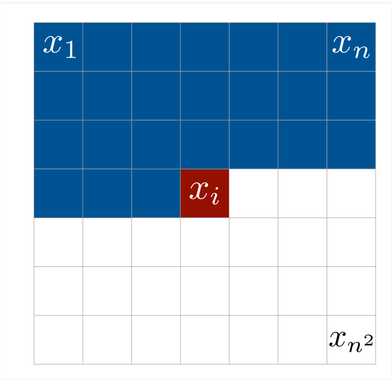
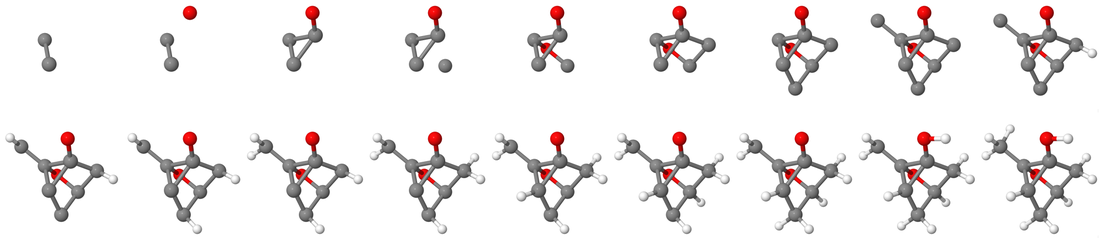


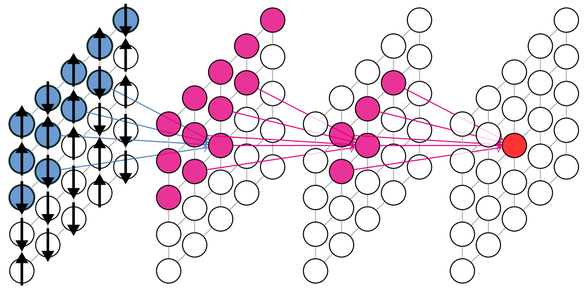
Image: PixelCNN 1601.06759



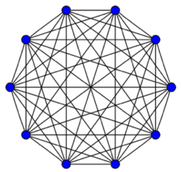
Molecular graph: 1810.11347



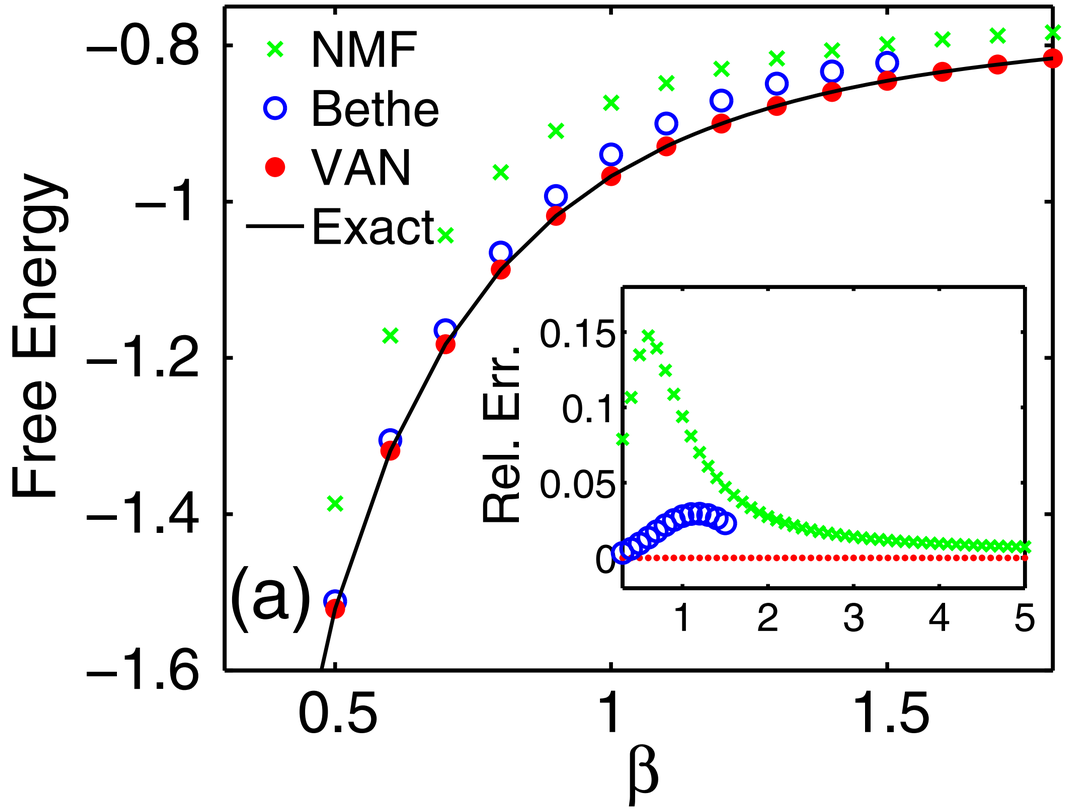
Ising model: 1809.10606



Variational autoregressive network for statistical mechanics



Sherrington-Kirkpatrick spin glass



Naive mean-field factorized probability

$$p(\boldsymbol{X}) = \prod_i p(x_i)$$

Bethe approximation pairwise interaction

$$p(\boldsymbol{X}) = \prod_i p(x_i) \prod_{(i,j)\in E} \frac{p(x_i, x_j)}{p(x_i)p(x_j)}$$

Variational autoregressive network

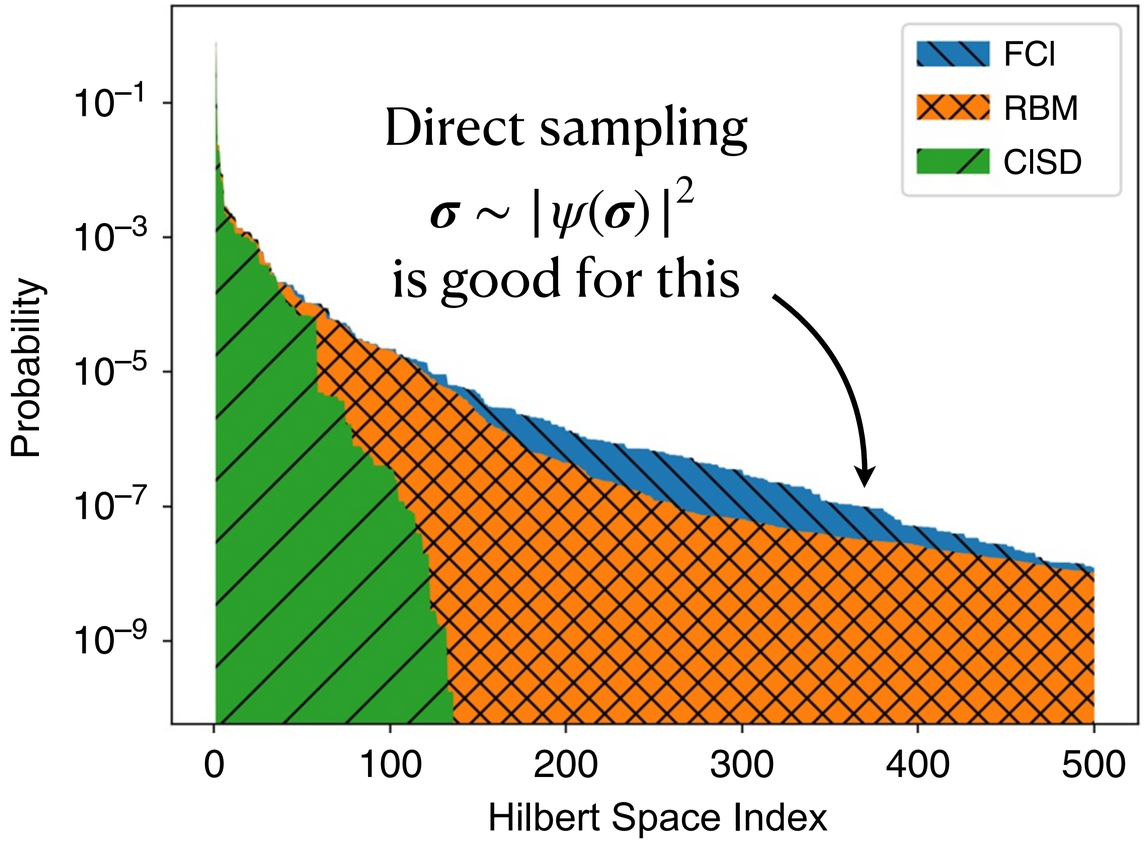
$$p(\boldsymbol{X}) = \prod_i p(x_i \,|\, \boldsymbol{x}_{<i})$$

Wu, LW, Zhang, PRL '19

github.com/wdphy16/stat-mech-van

Variational autoregressive quantum states

$$\psi(\boldsymbol{\sigma}) = \psi(\sigma_1)\psi(\sigma_2 \,|\, \sigma_1)\psi(\sigma_3 \,|\, \sigma_1, \sigma_2)\cdots$$



N2 molecule, Choo et al, Nat. Comm. '20

Objective function: ground state energy

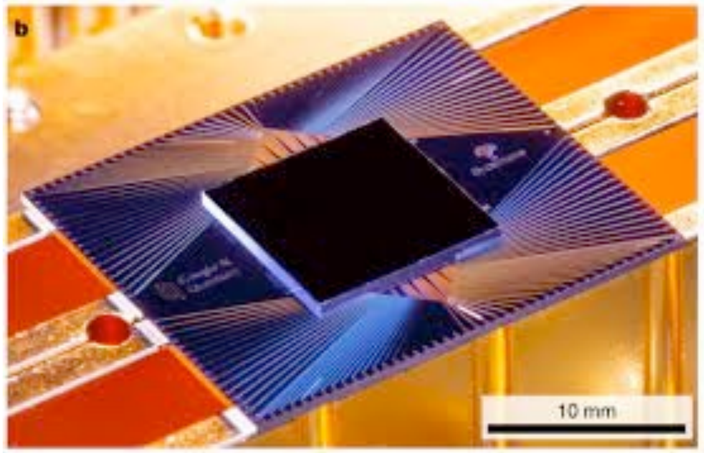
McMillan 1965, Carleo & Troyer Science 2017

$$\frac{\langle\psi|\hat{H}|\psi\rangle}{\langle\psi|\psi\rangle} = \mathop{\mathbb{E}}_{\boldsymbol{\sigma}\sim|\psi(\boldsymbol{\sigma})|^2}\left[\frac{\hat{H}\psi(\boldsymbol{\sigma})}{\psi(\boldsymbol{\sigma})}\right]$$

Sharir, Levine, Wies, Carleo, Shashua, PRL '20
Hibat-Allah, Ganahl, Hayward, Melko, Carrasquilla, PRResarch '20
Barrett et al, Nat. Mach. Intell. '22
Zhao et al, MLST. '23
Shang et al, 2307.09343

Demo: Generative model of Sycamore data

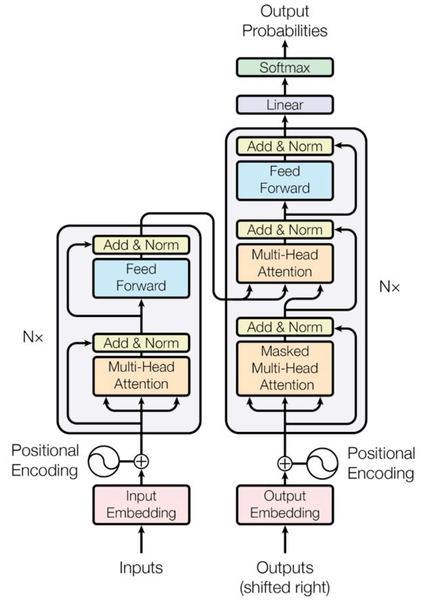
Quantum chip



bitstrings $\sim |\Psi(X)|^2$

```
011110110100
100001111011
100110110111
100110100010
010100011000
010001000000
010101101100
100001111000
100101001001
001000001010
```

Transformer

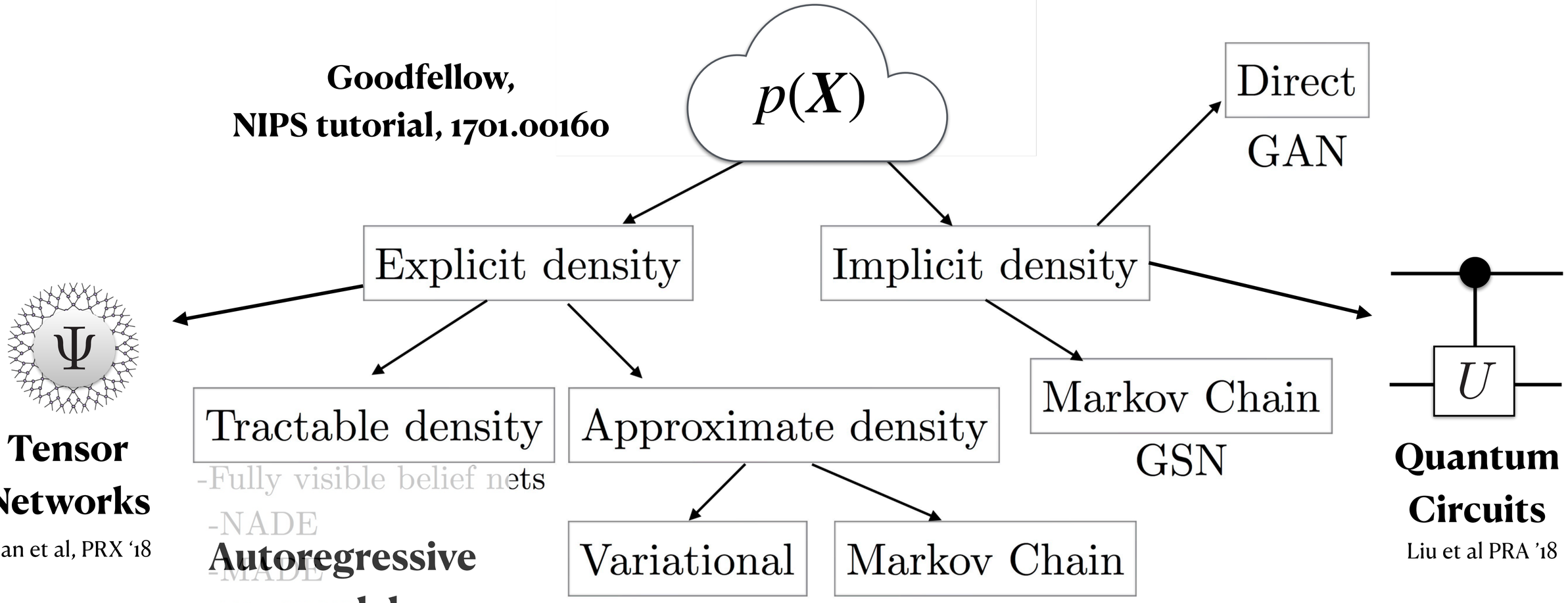


Can we fake the measurement of the sycamore quantum circuit by training a transformer?

https://colab.research.google.com/drive/11WaroqULkudKT3h2i5J6r_EmA4wFKkoZ?usp=sharing

Generative models and their physics genes

Goodfellow,
NIPS tutorial, 1701.00160



Tensor Networks

Han et al, PRX '18

Autoregressive model

-Fully visible belief nets
-NADE
-MADE
-PixelRNN
-Change of variables models (nonlinear ICA)

Flow model

Quantum Circuits

Liu et al PRA '18

+**Diffusion models**

Normalizing flows

Parallel WaveNet 1711.10433

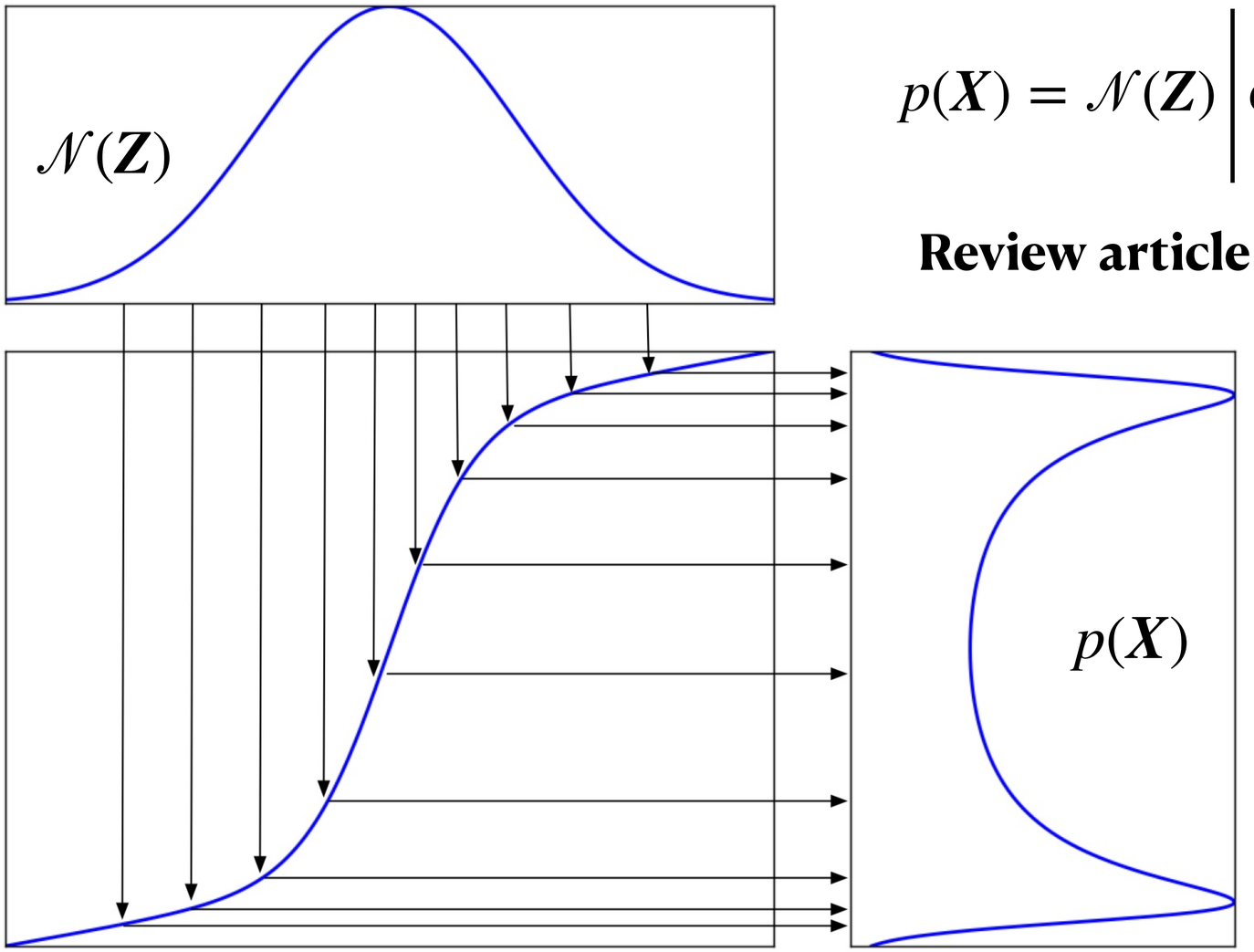
https://deepmind.com/blog/high-fidelity-speech-synthesis-wavenet/

Glow 1807.03039

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

Normalizing flow in a nutshell

Base density



$$p(\boldsymbol{X}) = \mathcal{N}(\mathbf{Z}) \left| \det \left(\frac{\partial \mathbf{Z}}{\partial \boldsymbol{X}} \right) \right|$$

Review article 1912.02762

"neural net" with 1 neuron

Target density

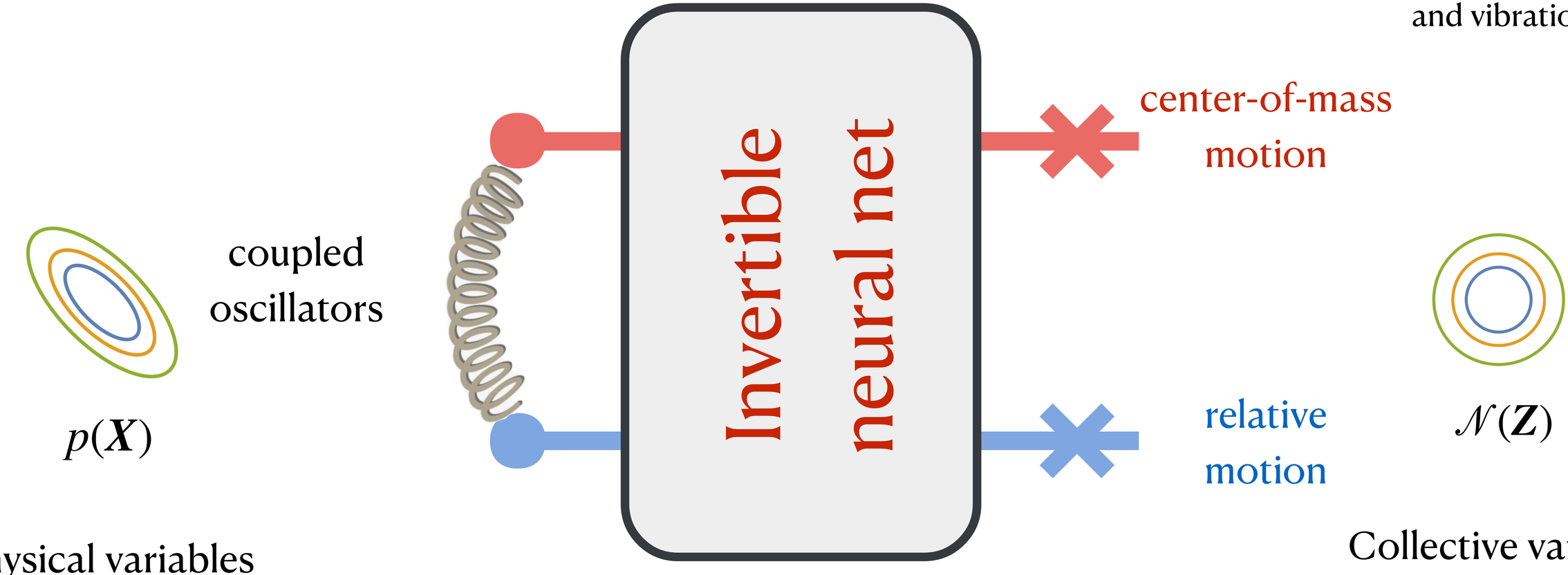
Physics intuition of normalizing flow

Li and LW, PRL '18

Li, Dong, Zhang, LW, PRX '20

Zhang, Wang, LW, JCP '24

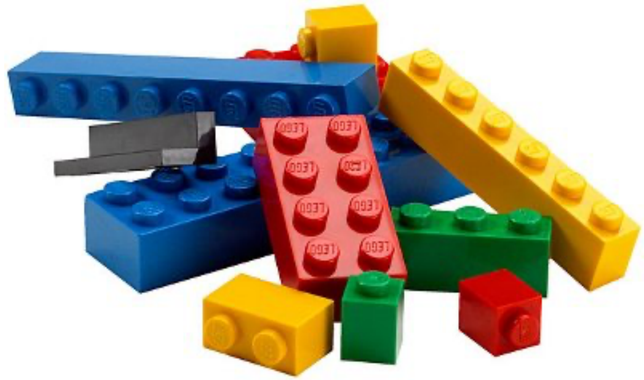
for RG, canonical transformations, and vibrational spectra



High-dimensional, composable, learnable, nonlinear transformations

Flow architecture design

Composability

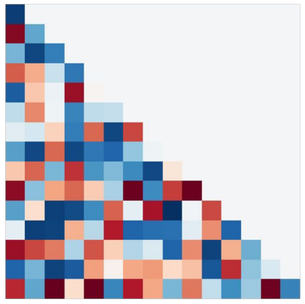


$$\boldsymbol{Z} = \mathscr{T}(\boldsymbol{X})$$

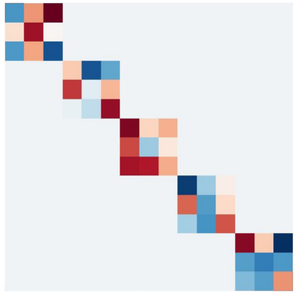
$$\mathscr{T} = \mathscr{T}_1 \circ \mathscr{T}_2 \circ \mathscr{T}_3 \circ \cdots$$

Balanced efficiency & inductive bias

$$\left|\det\left(\frac{\partial \boldsymbol{Z}}{\partial \boldsymbol{X}}\right)\right|$$



Autoregressive



Blockwise

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

Continuous flow

Example of a building block

Forward

arbitrary neural nets

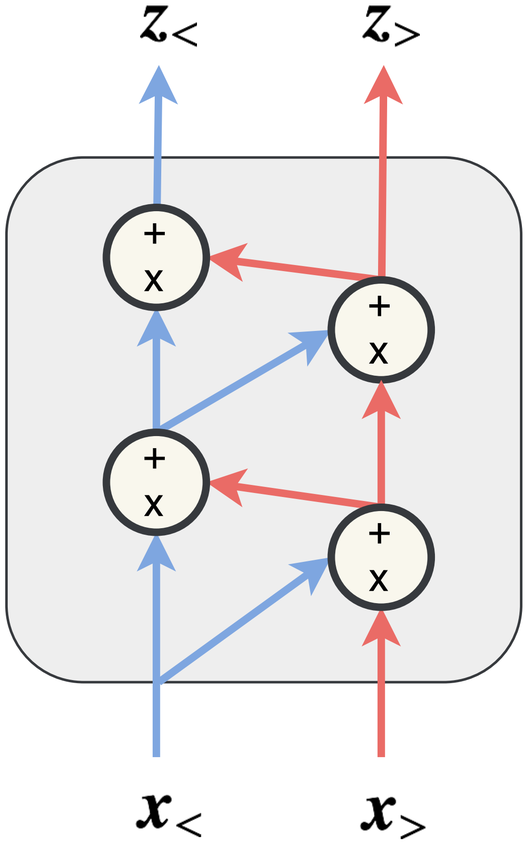
$$\begin{cases} \boldsymbol{x}_< = \boldsymbol{z}_< \\ \boldsymbol{x}_> = \boldsymbol{z}_> \odot e^{s(\boldsymbol{z}_<)} + t(\boldsymbol{z}_<) \end{cases}$$

Inverse

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

Log-Abs-Jacobian-Det

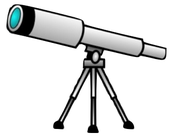
$$\ln \left| \det \left(\frac{\partial \boldsymbol{x}}{\partial \boldsymbol{z}} \right) \right| = \sum_i [s(\boldsymbol{z}_<)]_i$$



Real NVP, Dinh et al,1605.08803

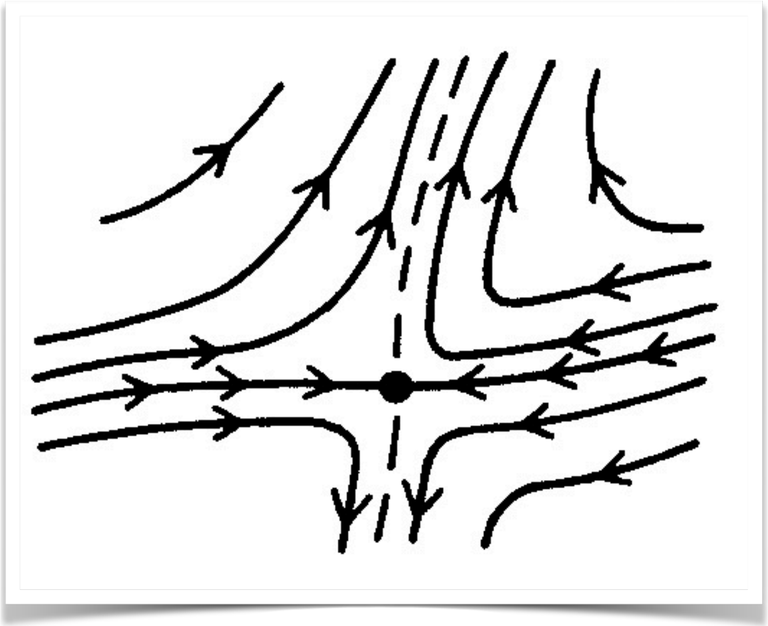
Turns out to have surprising connection Störmer–Verlet integration

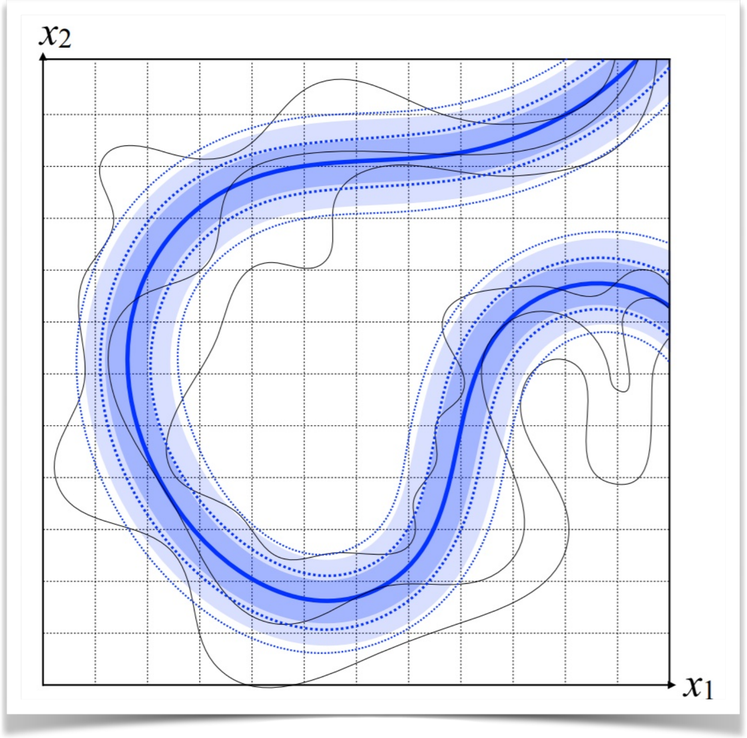
Why is flow useful for physics?



Renormalization group

Monte Carlo update





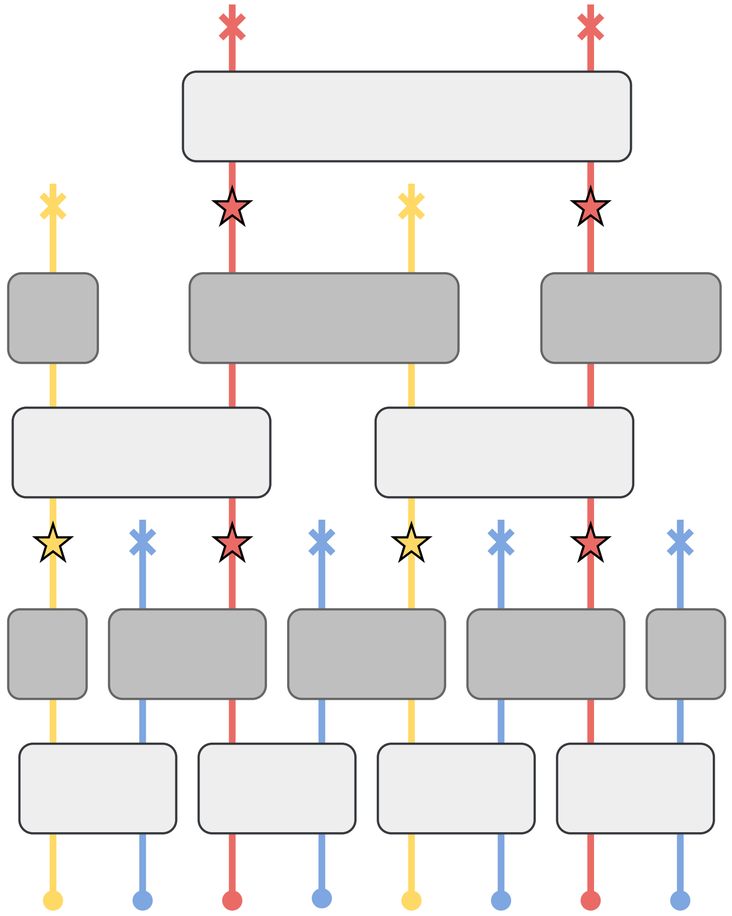
Effective theory emerges upon transformation of the variables

Physics happens on a manifold
Train neural nets to unfold that manifold

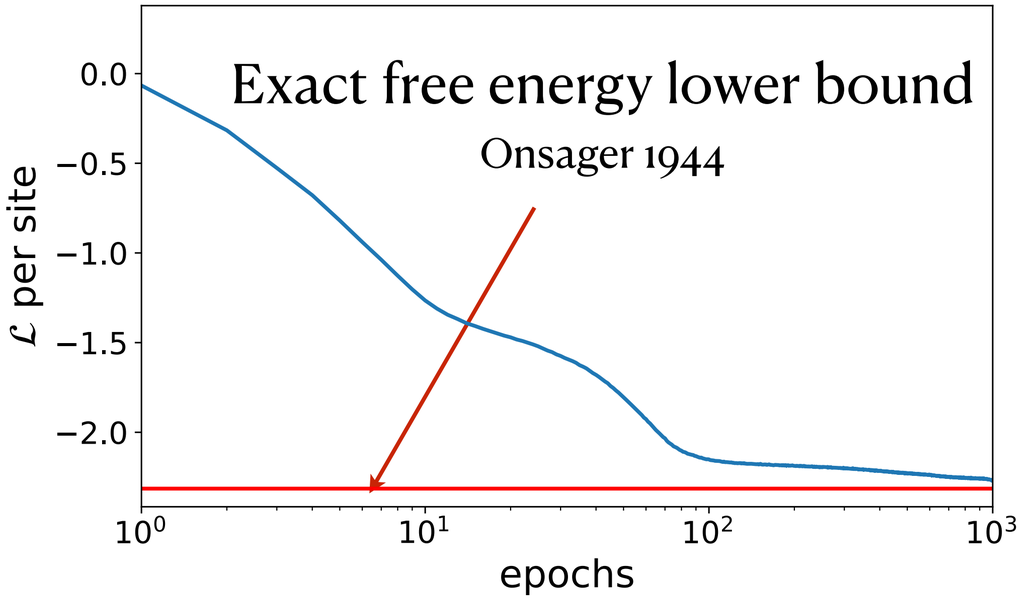
Neural network renormalization group

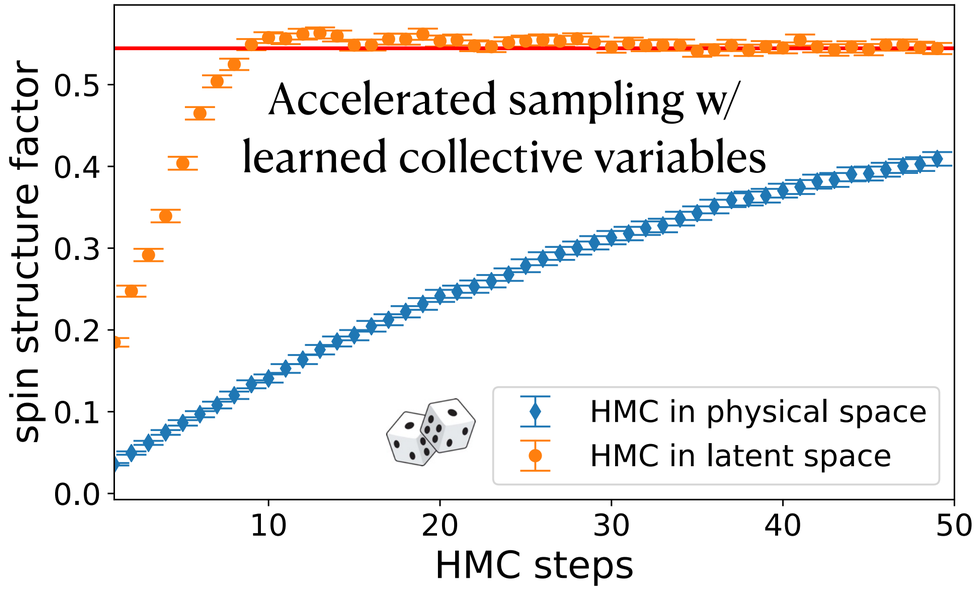
Li, LW, PRL '18 [lio12589/NeuralRG]

Collective variables

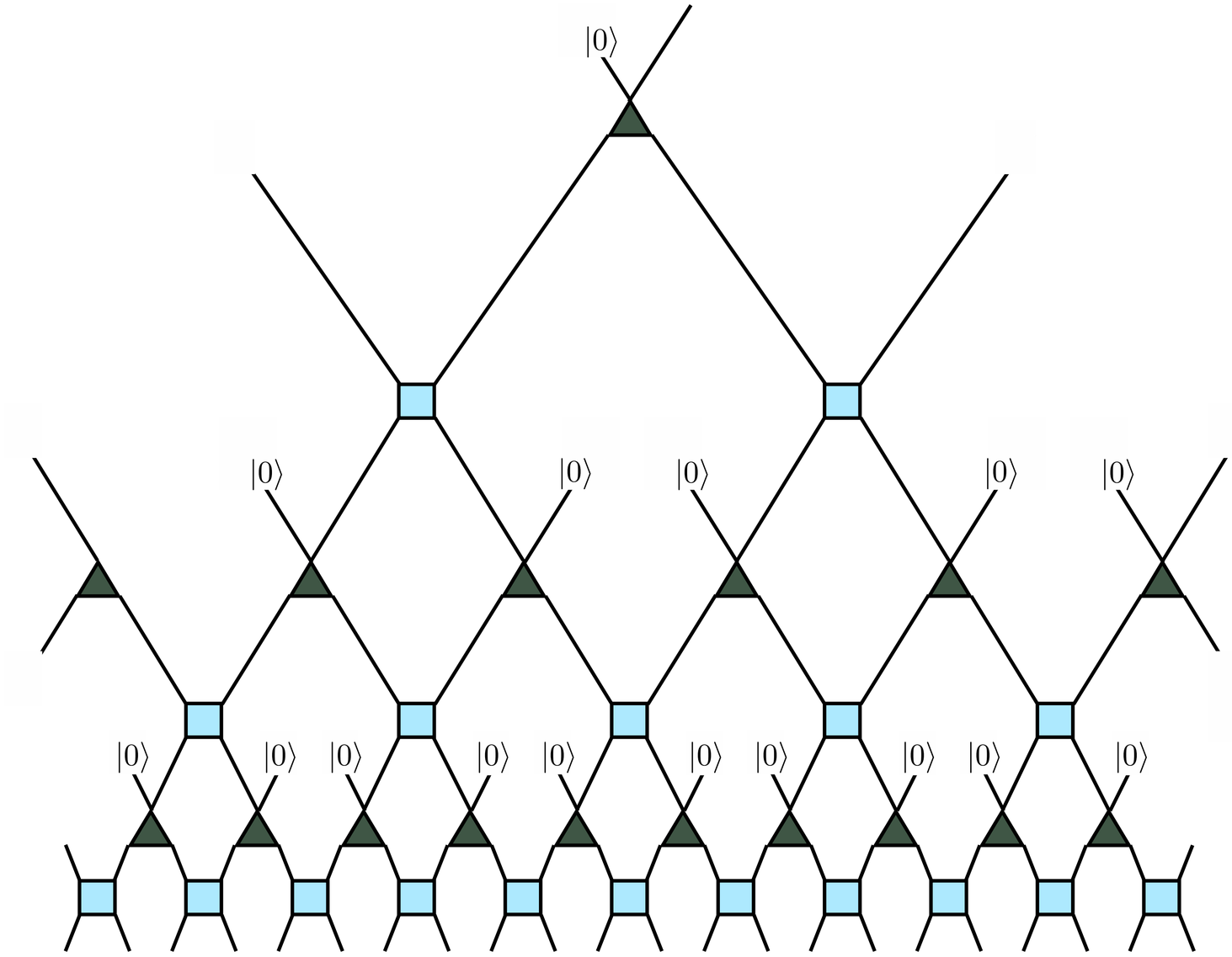


Physical variables

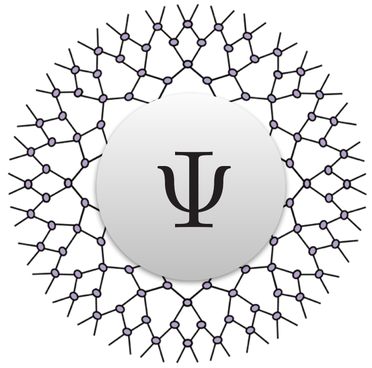




Quantum version of the architecture

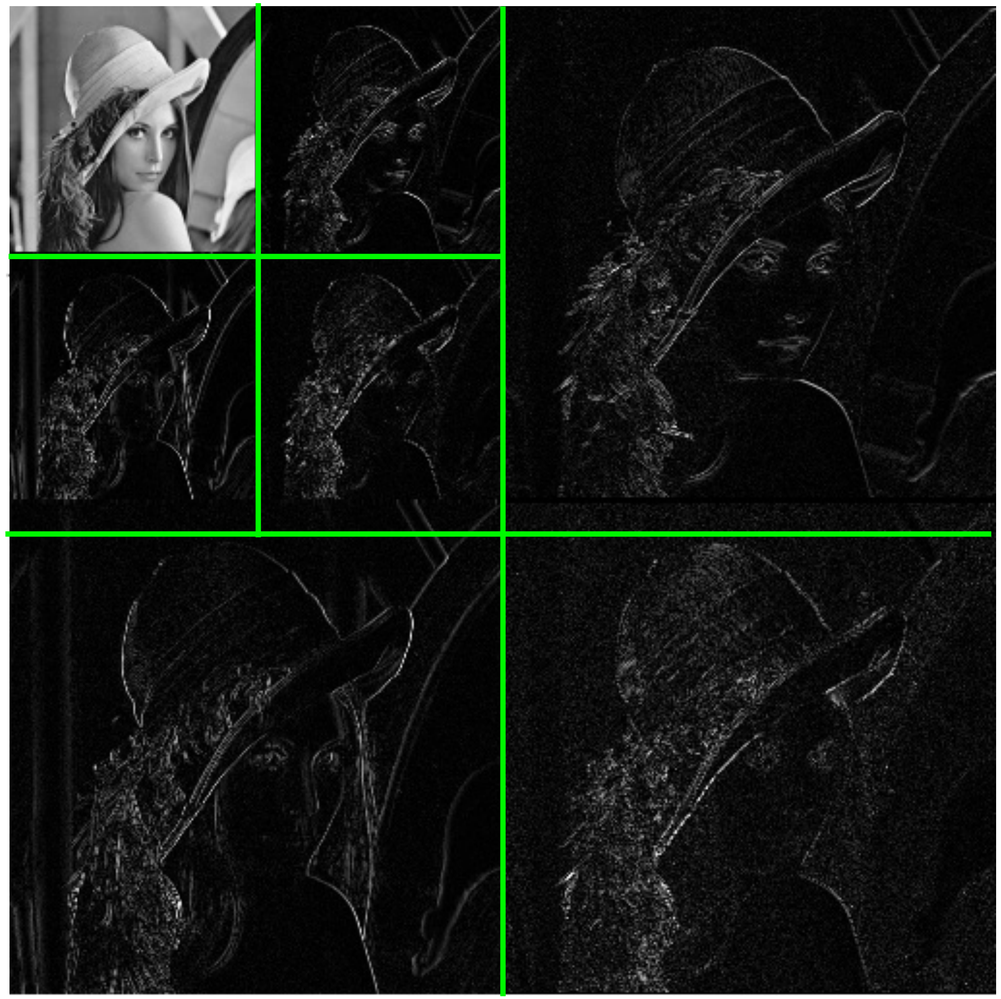


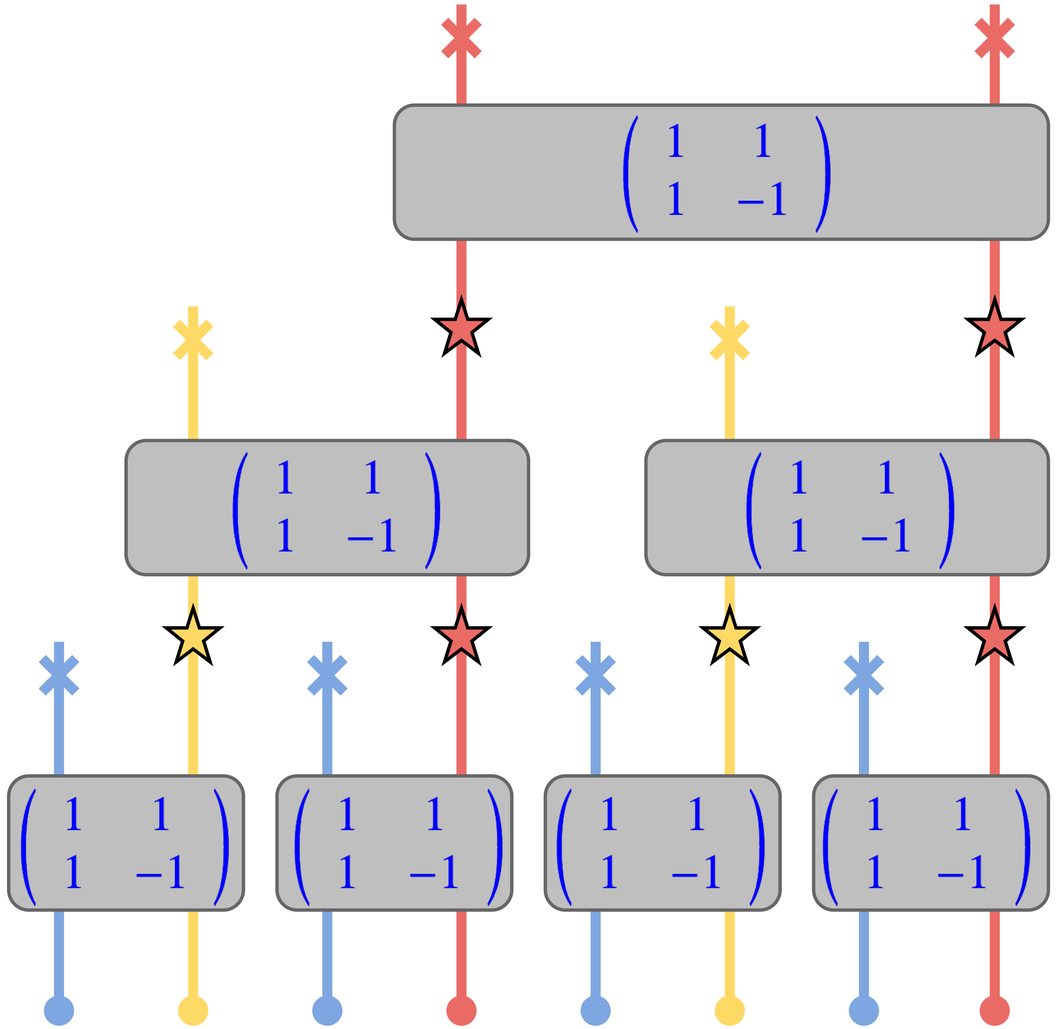
Entangled qubits



Multi-Scale
Entanglement
Renormalization
Ansatz

Connection to wavelets



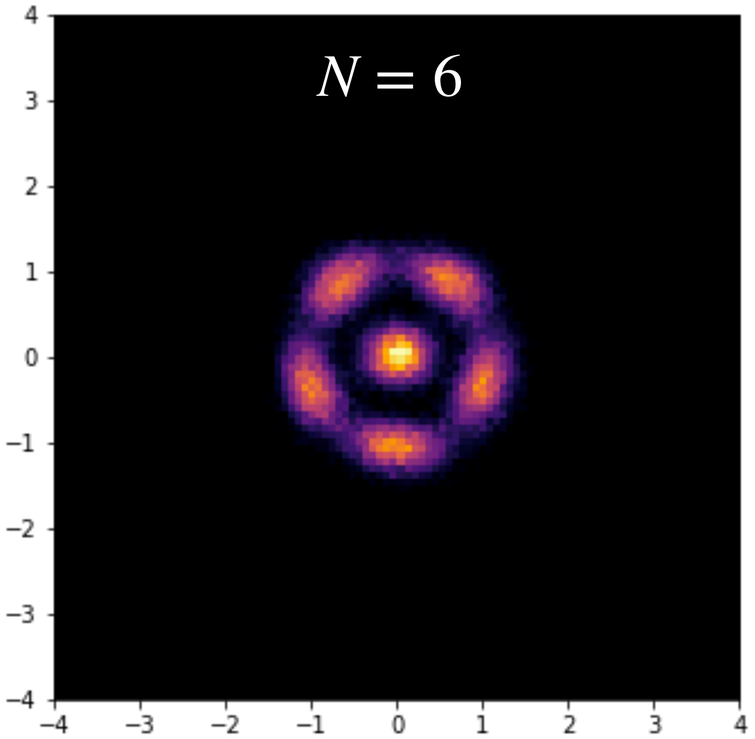


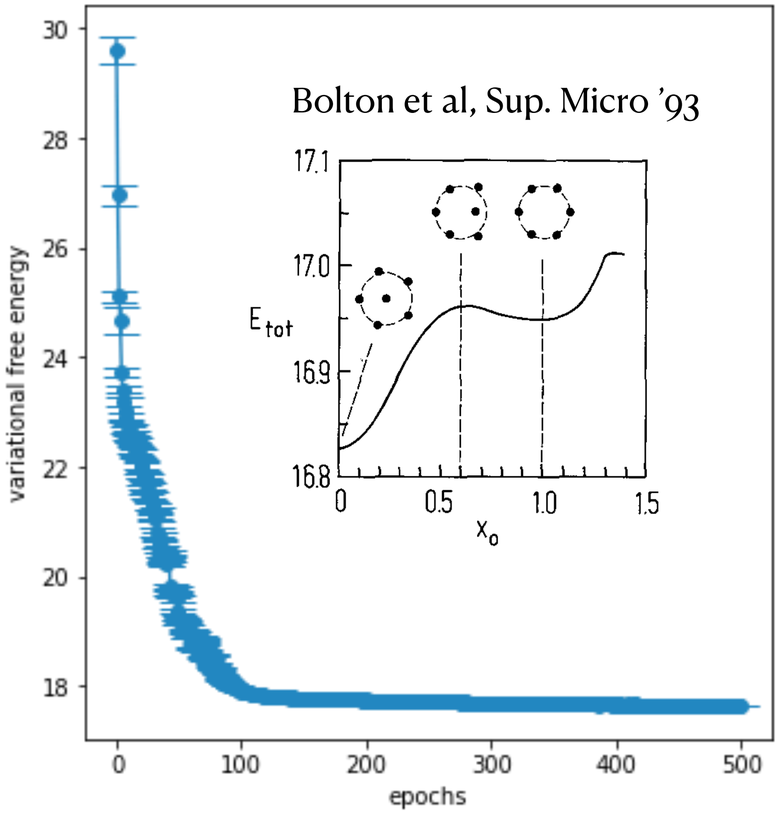
Nonlinear & adaptive generalizations of wavelets

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

Demo: Classical Coulomb gas in a harmonic trap

$$E = \sum_{i<j} \frac{1}{|\boldsymbol{x}_i - \boldsymbol{x}_j|} + \sum_i^N \boldsymbol{x}_i^2$$





https://colab.research.google.com/drive/1yIlPo5CAjYrqWHeFEZrMlzWNCoNJ6_YP#scrollTo=eQwLElKmaowu

Optimization: Monte Carlo Gradient Estimators

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

Review: 1906.10652

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

Score function estimator (REINFORCE)

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

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

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

10.1 Guidance in Choosing Gradient Estimators

With so many competing approaches, we offer our rules of thumb in choosing an estimator, which follow the intuition we developed throughout the paper:

- If our estimation problem involves continuous functions and measures that are continuous in the domain, then using the pathwise estimator is a good default. It is relatively easy to implement and a default implementation, one without other variance reduction, will typically have variance that is low enough so as not to interfere with the optimisation.
- If the cost function is not differentiable or a black-box function then the score-function or the measure-valued gradients are available. If the number of parameters is low, then the measure-valued gradient will typically have lower variance and would be preferred. But if we have a high-dimensional parameter set, then the score function estimator should be used.
- If we have no control over the number of times we can evaluate a black-box cost function, effectively only allowing a single evaluation of it, then the score function is the only estimator of the three we reviewed that is applicable.
- The score function estimator should, by default, always be implemented with at least a basic variance reduction. The simplest option is to use a baseline control variate estimated with a running average of the cost value.
- When using the score-function estimator, some attention should be paid to the dynamic range of the cost function and its variance, and to find ways to keep its value bounded within a reasonable range, e.g., transforming the cost so that it is zero mean, or using a baseline.
- For all estimators, track the variance of the gradients if possible and address high variance by using a larger number of samples from the measure, decreasing the learning rate, or clipping the gradient values. It may also be useful to restrict the range of some parameters to avoid extreme values, e.g., by clipping them to a desired interval.
- The measure-valued gradient should be used with some coupling method for variance reduction. Coupling strategies that exploit relationships between the positive and negative components of the density decomposition, and which have shared sampling paths, are known for the commonly-used distributions.
- If we have several unbiased gradient estimators, a convex combination of them might have lower variance than any of the individual estimators.
- If the measure is discrete on its domain then the score-function or measure-valued gradient are available. The choice will again depend on the dimensionality of the parameter space.
- In all cases, we strongly recommend having a broad set of tests to verify the unbiasedness of the gradient estimator when implemented.

Mohamed et al, 1906.10652

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

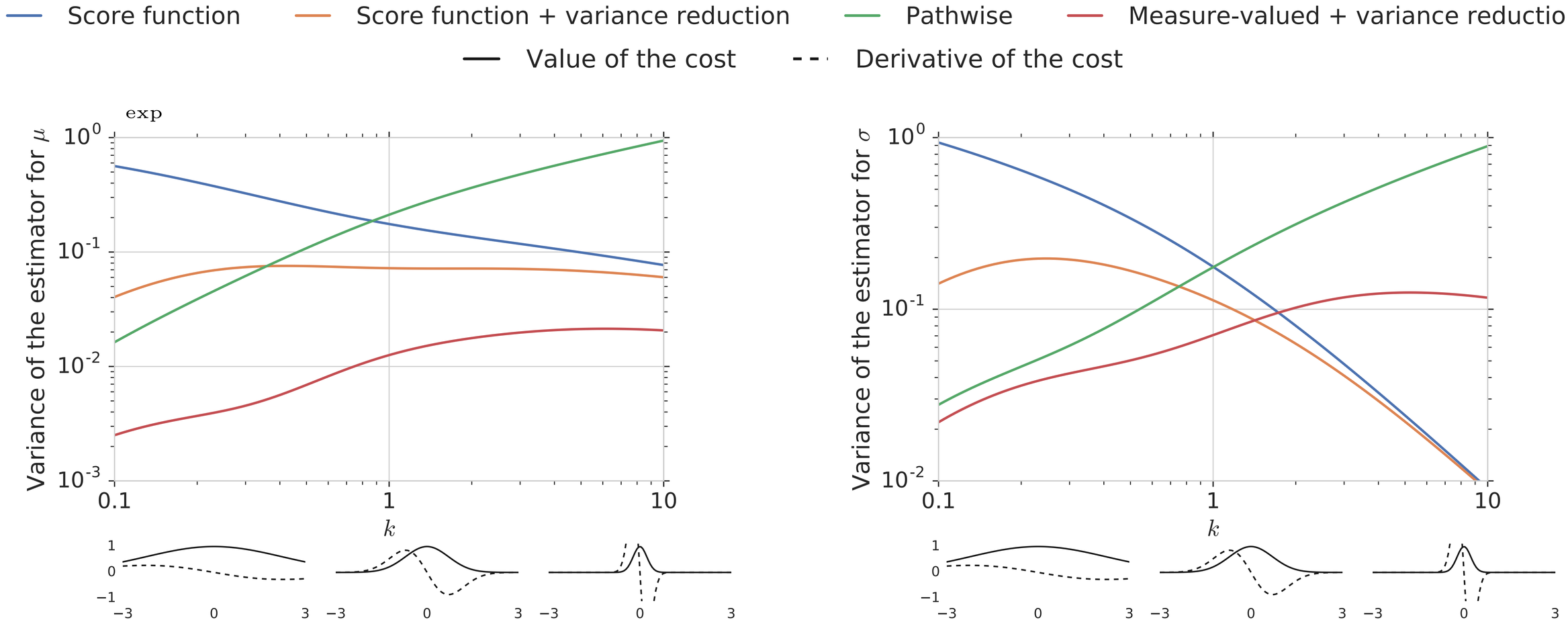
When to use which ?

More discussions

Roeder et al, 1703.09194

Vaitl et al 2206.09016, 2207.08219

$$\eta = \nabla_\theta \int \mathcal{N}(x|\mu, \sigma^2) f(x; k) dx; \quad \theta \in \{\mu, \sigma\}$$



https://github.com/deepmind/mc_gradients Mohamed et al, 1906.10652

Continuous normalizing flows

$$\ln p(\boldsymbol{X}) = \ln \mathcal{N}(\boldsymbol{Z}) - \ln \left| \det \left(\frac{\partial \boldsymbol{X}}{\partial \boldsymbol{Z}} \right) \right|$$

Consider infinitesimal change-of-variables Chen et al 1806.07366

$$\boldsymbol{X} = \boldsymbol{Z} + \varepsilon \boldsymbol{v} \qquad \ln p(\boldsymbol{X}) - \ln \mathcal{N}(\boldsymbol{Z}) = - \ln \left| \det \left(1 + \varepsilon \frac{\partial \boldsymbol{v}}{\partial \boldsymbol{Z}} \right) \right|$$

$\varepsilon \to 0$

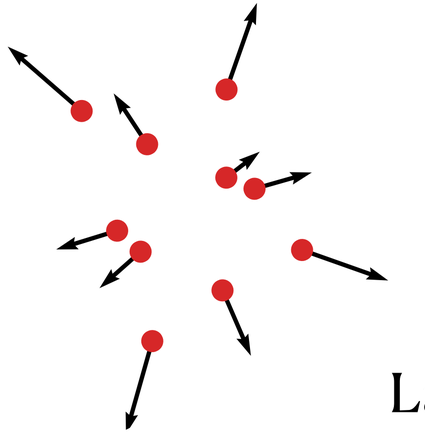
$t = 0$ $t = 1$

$$\frac{d\boldsymbol{X}}{dt} = \boldsymbol{v} \qquad \frac{d \ln p(\boldsymbol{X}, t)}{dt} = - \nabla \cdot \boldsymbol{v}$$

Fluid physics behind flows

$$\frac{d\boldsymbol{X}}{dt} = \boldsymbol{v}$$

$$\frac{d \ln p(\boldsymbol{X}, t)}{dt} = -\nabla \cdot \boldsymbol{v}$$



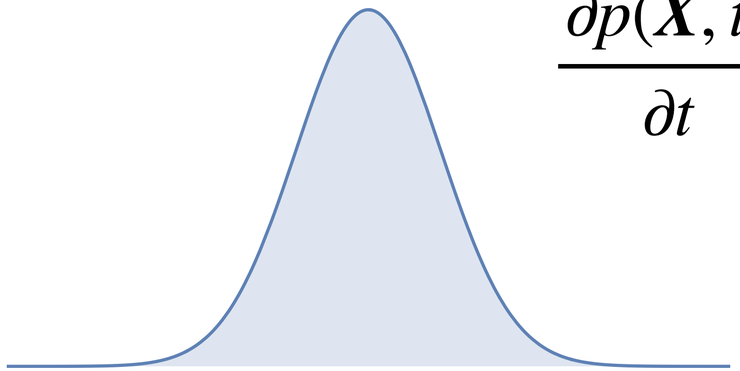
Zhang, E, LW 1809.10188
wangleiphy/MongeAmpereFlow

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

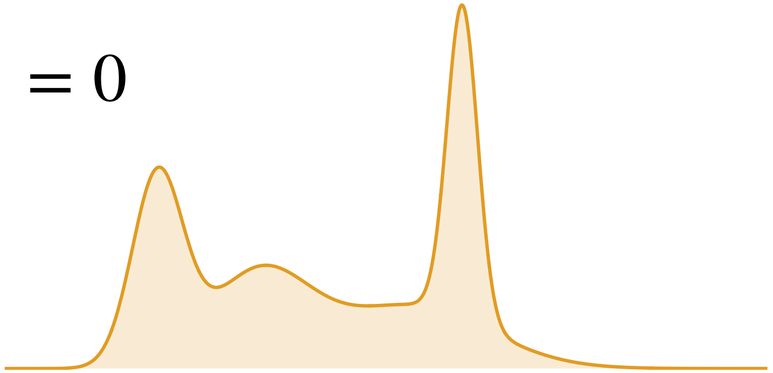
"material derivative"

Lagrangian v.s. Euler approach to fluid mechanics

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



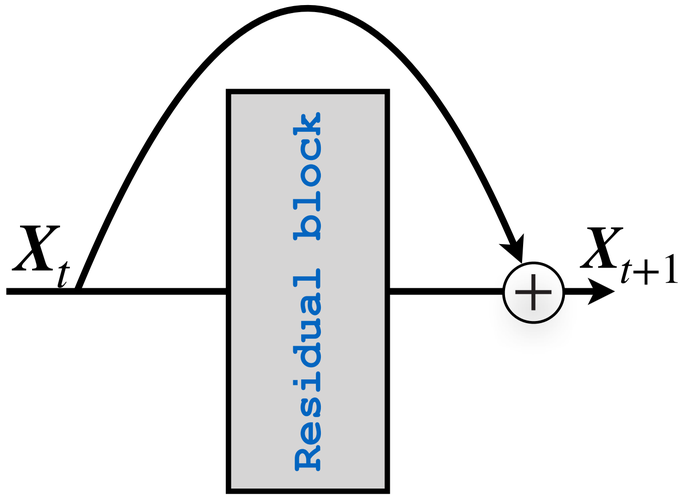
Simple density



Complex density

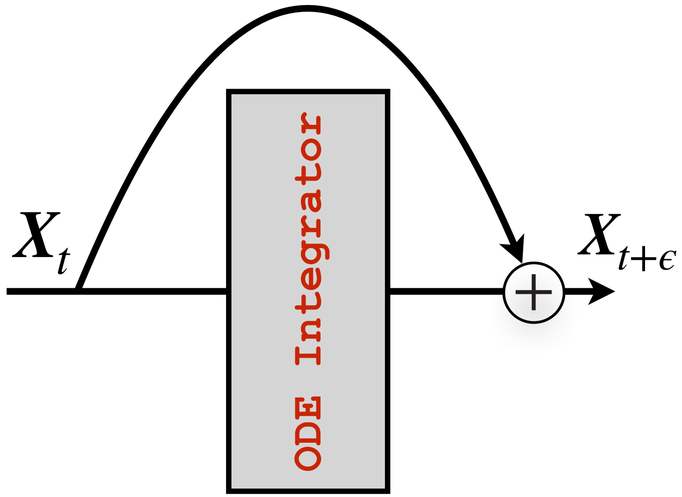
Neural Ordinary Differential Equations

Residual network



$$X_{t+1} = X_t + v(X_t)$$

ODE integration



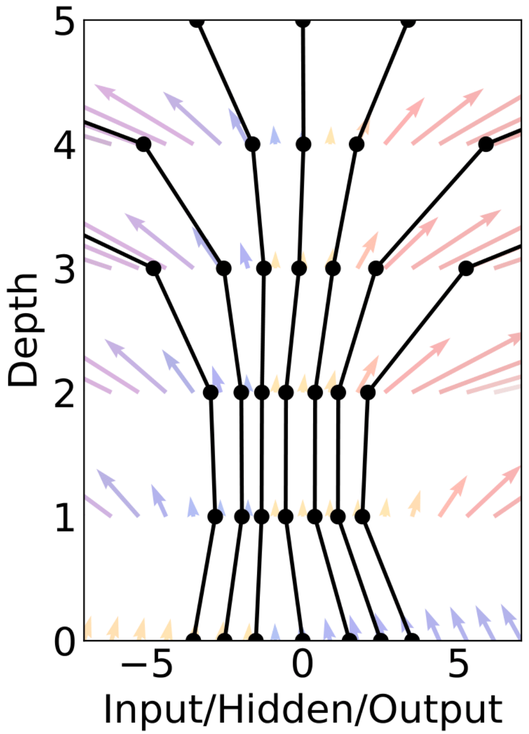
$$dX/dt = v(X)$$

Chen et al, 1806.07366

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

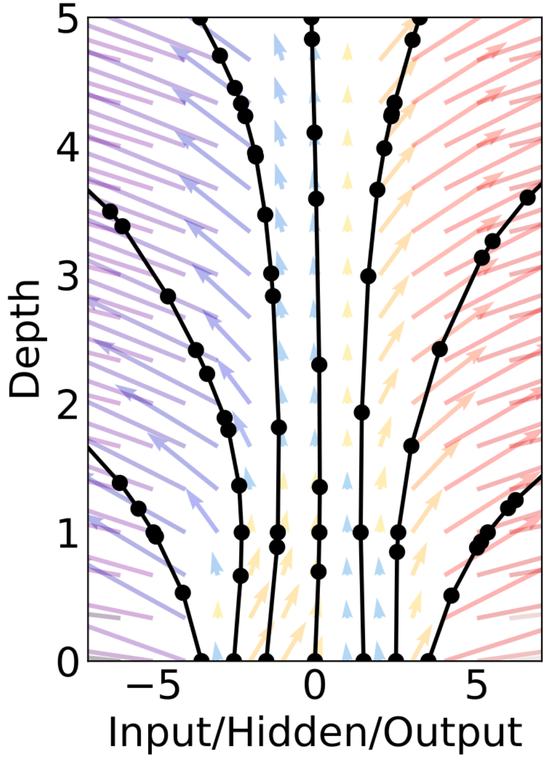
Neural Ordinary Differential Equations

Residual network



$$X_{t+1} = X_t + v(X_t)$$

ODE integration



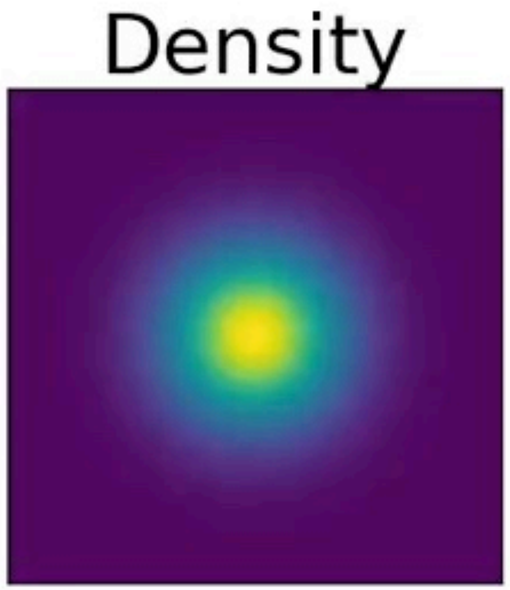
$$dX/dt = v(X)$$

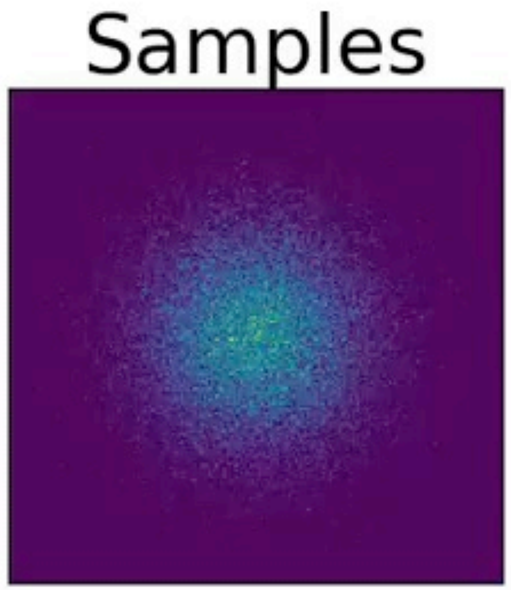
Chen et al, 1806.07366

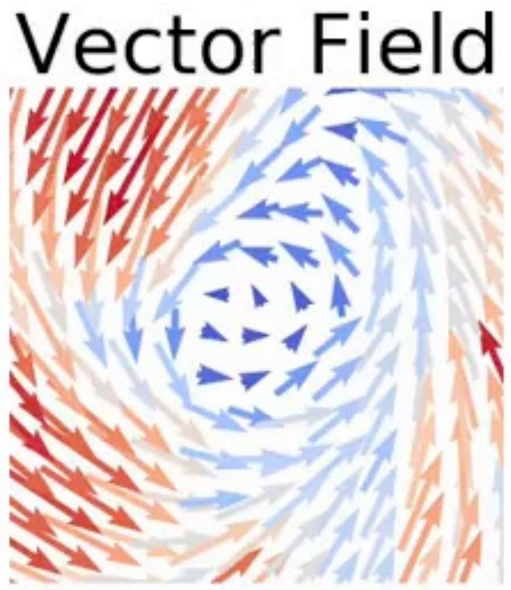
Harbor el al 1705.03341
Lu et al 1710.10121,
E Commun. Math. Stat 17'...

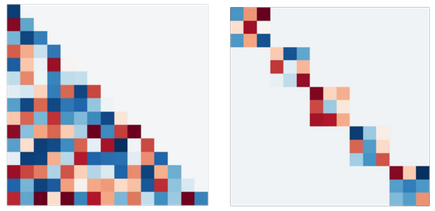
Continuous normalizing flows implemented with NeuralODE

Chen et al, 1806.07366, Grathwohl et al 1810.01367









Continuous normalizing flow have no structural constraints on the transformation Jacobian

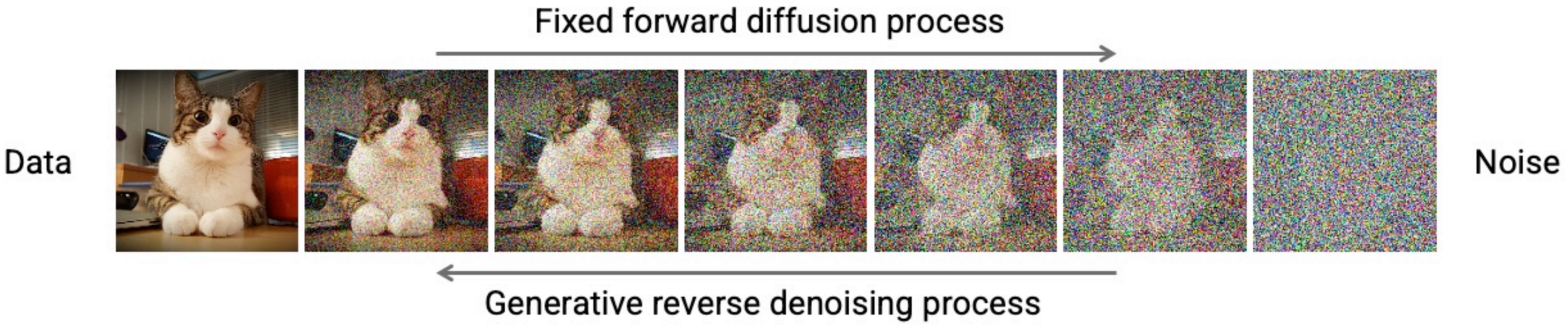
From flow to diffusion model

Continuity equation

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

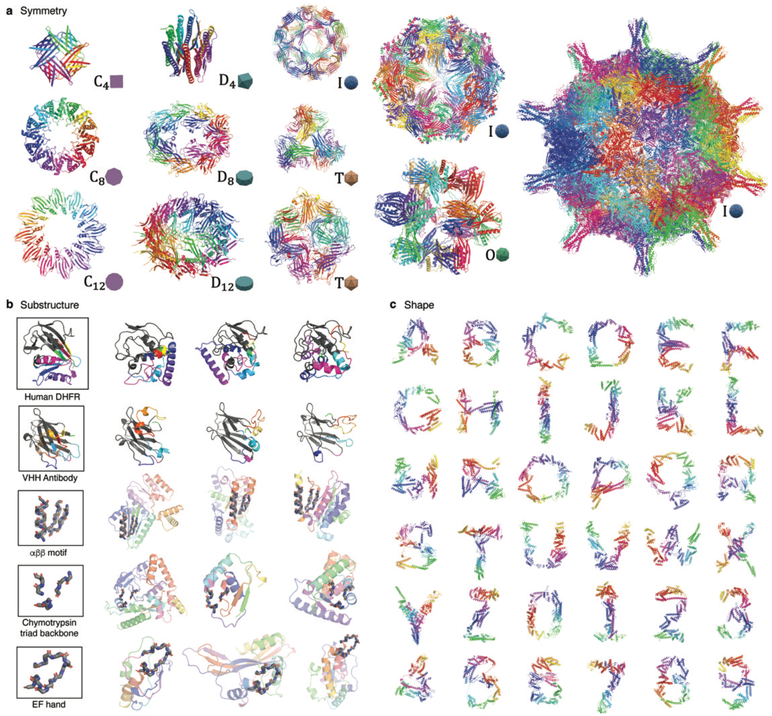
Fokker-Planck equation

$$\frac{\partial p(\boldsymbol{X}, t)}{\partial t} + \nabla \cdot \left[p(\boldsymbol{X}, t) \boldsymbol{f} \right] - \nabla^2 p(\boldsymbol{X}, t) = 0$$

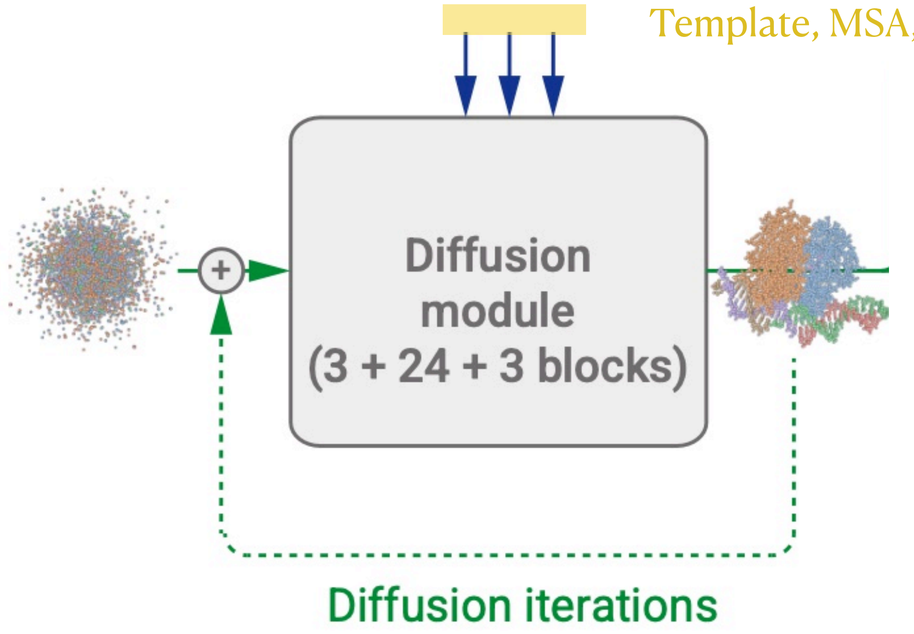


Vincent 2011, Sohl-Dickstein et al, 1503.03585, Song et al, 1907.05600 Ho et al, 2006.11239...

Diffusion models for protein structure prediction and design



Ingraham et al, Chroma, Nature 2023
https://generatebiomedicines.com/chroma



Abramson et al, AlphaFold3, Nature 2024
https://deepmind.google/technologies/alphafold/

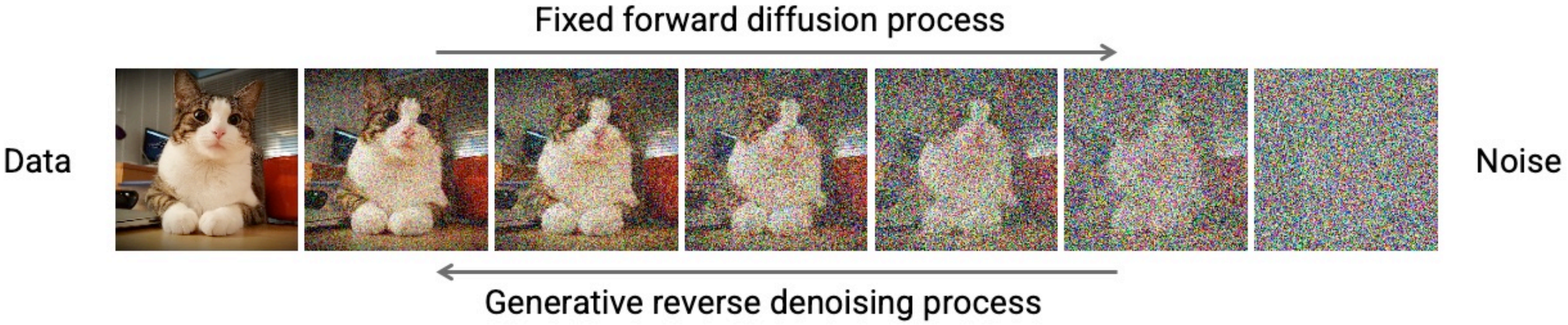
From flow to diffusion model, and back

Continuity equation

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

Fokker-Planck equation

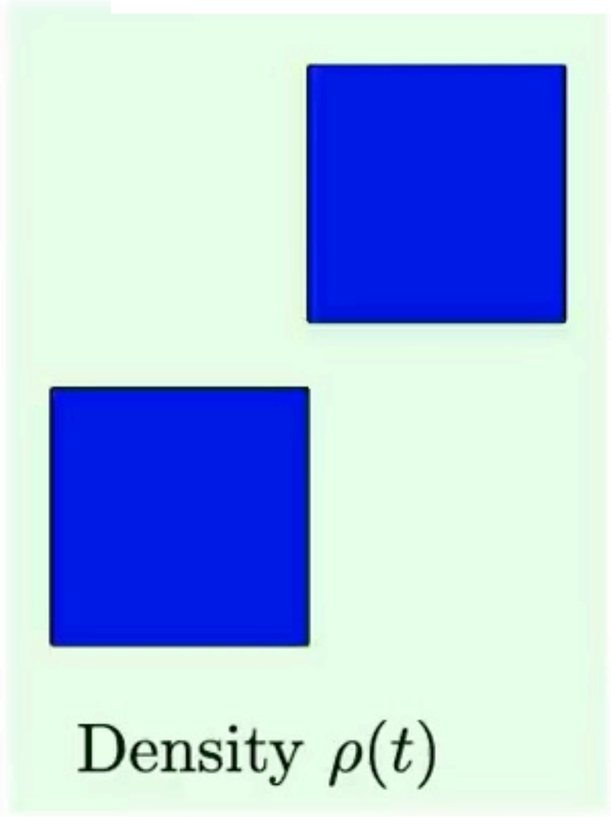
$$\frac{\partial p(\boldsymbol{X}, t)}{\partial t} + \nabla \cdot \left[p(\boldsymbol{X}, t) \left(\boldsymbol{f} - \nabla \ln p(\boldsymbol{X}, t) \right) \right] = 0$$

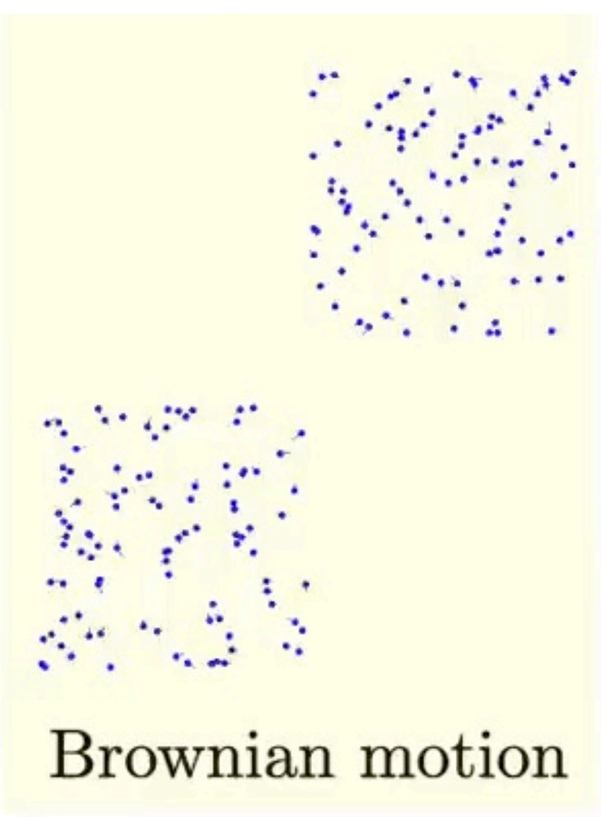


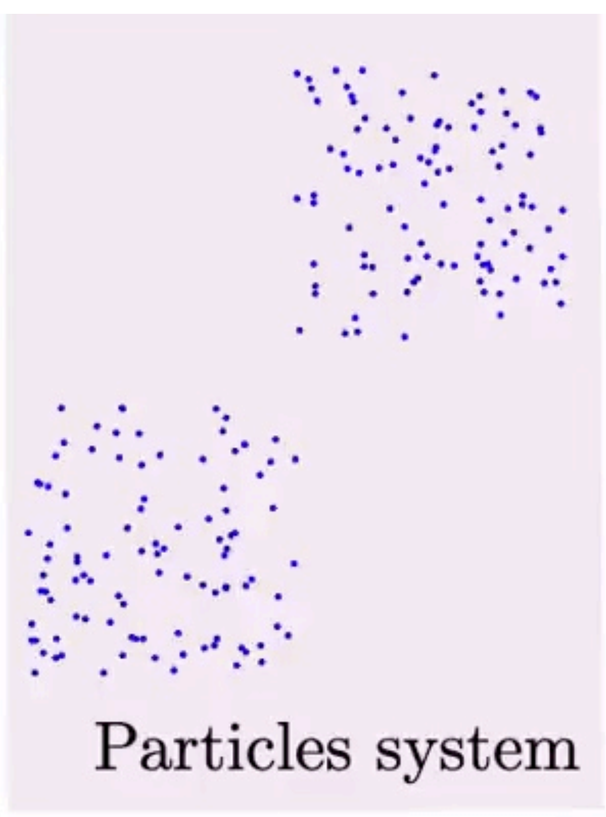
Maoutsa et al, 2006.00702, Song et al, 2011.13456
Liu et al 2209.03003, Albergo et al, 2209.15571, Lipman et al, 2210.02747

A tale of three equations

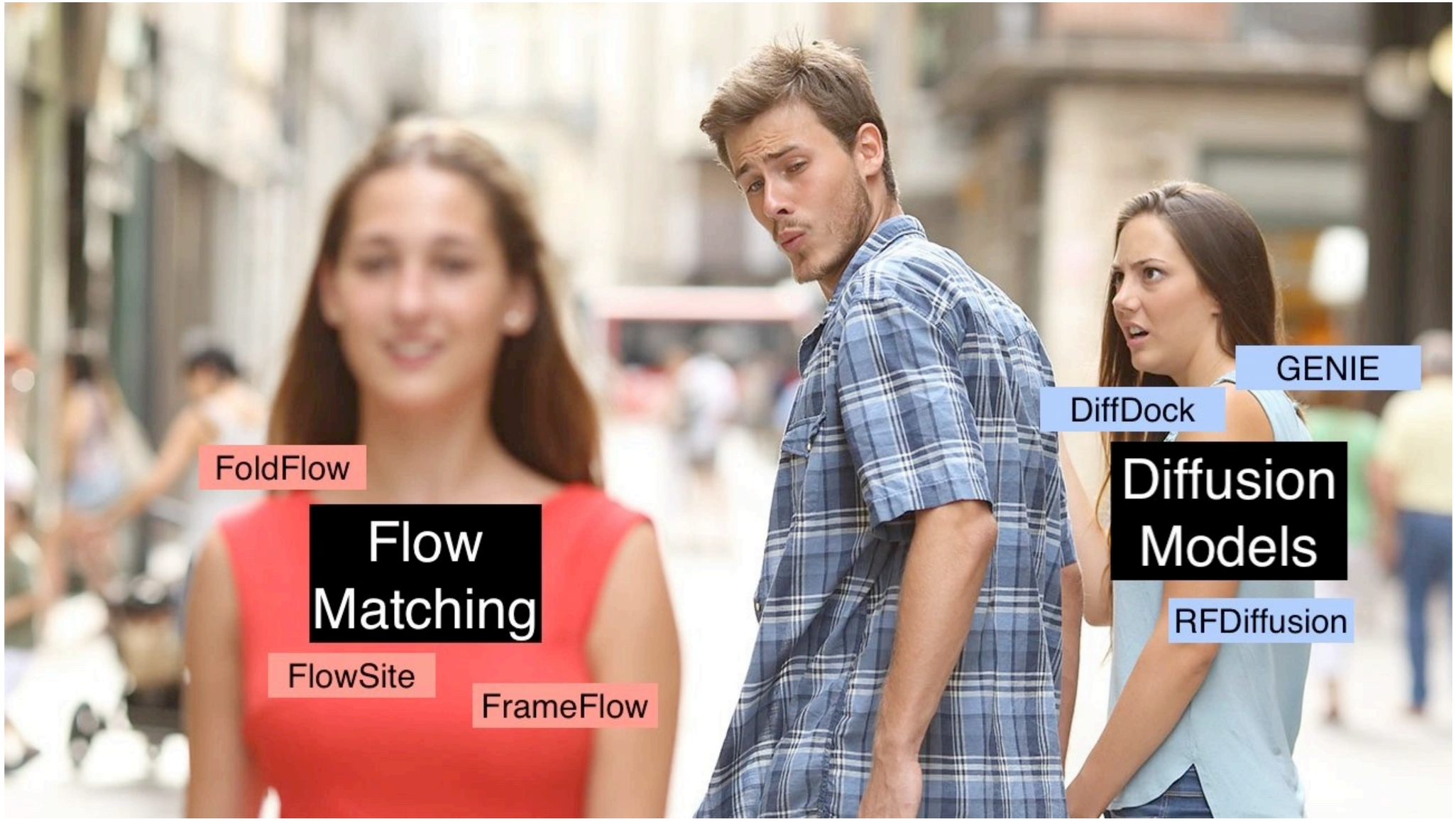
偏微分方程	随机微分方程	常微分方程
Fokker-Planck	Diffusion model	Flow model







https://twitter.com/gabrielpeyre/status/1744962274018894292

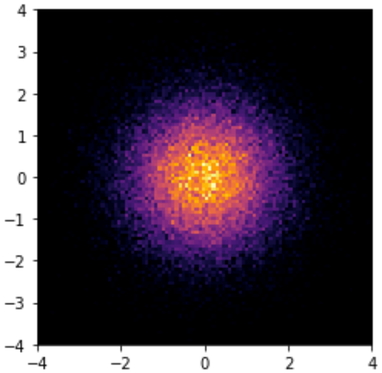


https://twitter.com/michael_galkin/status/1711845455817261409

Demo: bounding free energy of classical Coulomb gas

$$\mathscr{L} = \mathbb{E}_{t\sim\mathscr{U}(0,1)}\mathbb{E}_{\boldsymbol{x}_0\sim\mathscr{N}(0,I)}\mathbb{E}_{\boldsymbol{x}_1\sim\exp(-\beta E)/Z}\left|\boldsymbol{x}_1 - \boldsymbol{x}_0 - \boldsymbol{v}(\boldsymbol{x},t)\right|^2$$

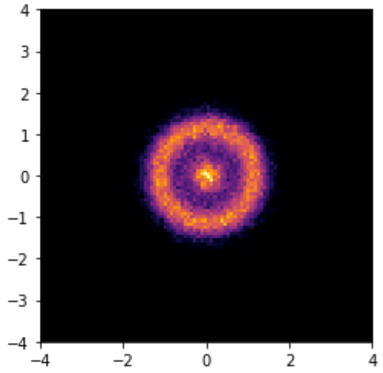
$$Z = \mathbb{E}_{\boldsymbol{x}\sim q(\boldsymbol{x})}\left[e^{-\beta E(\boldsymbol{x}) - \ln q(\boldsymbol{x})}\right] \quad \ln q(\boldsymbol{x}) = \ln \mathscr{N}(0,I) - \int_0^1 \nabla\cdot\boldsymbol{v}dt$$



Base density

Gaussian samples

Interpolate samples to estimate free energy differences



Target density

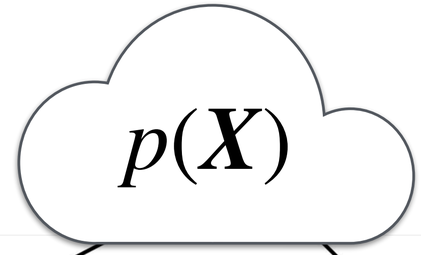
Monte Carlo samples

https://colab.research.google.com/drive/1t-Vk37Axxp040B7uXFUNlk-zeCC2lcX3?usp=sharing

Jarzynski PRE '02, see also likelihood-based training of flows Wirnsberger et al, 2002.04913, 2111.08696

Generative models and their physics genes

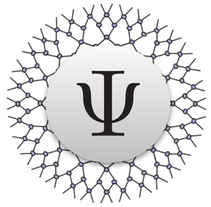
Goodfellow,
NIPS tutorial, 1701.00160



Direct
GAN

Explicit density

Implicit density



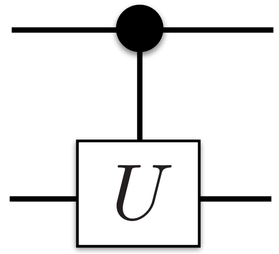
Tensor Networks

Han et al, PRX '18

Tractable density

Approximate density

Markov Chain
GSN



Quantum Circuits

Liu et al PRA '18

-Fully visible belief nets
-NADE
-MADE
-PixelRNN
-Change of variables models (nonlinear ICA)

Autoregressive model

Flow model

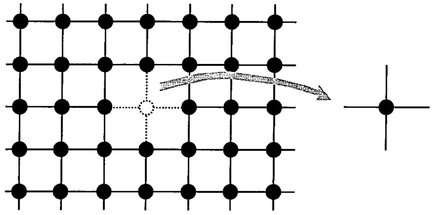
Variational

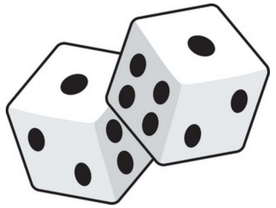
Markov Chain

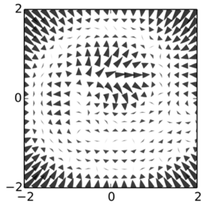
Variational autoencoder

Boltzmann machine

+**Diffusion models**







Which is the best ?

> **Jeremy Howard** @jeremyphoward · Apr 15
> "transformers or diffusion"
>
> Ummm.... who's gonna tell him that that makes no sense at all?

Autoregressive model

$$p(\boldsymbol{X}) = p(x_1)p(x_2 \mid x_1)p(x_3 \mid x_1, x_2)\cdots$$

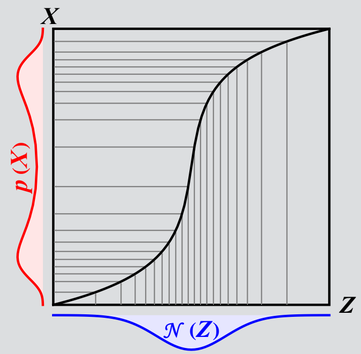
"... the murderer is ___"

$p(\_ \mid \dots)$

Discrete or continuous sequantial data

Flow model

$$p(\boldsymbol{X}) = \mathcal{N}(\mathbf{Z}) \left| \det \left(\frac{\partial \mathbf{Z}}{\partial \boldsymbol{X}} \right) \right|$$



Continuous structured data

Which is the best ?

The “it” in AI models is the dataset.

Posted on June 10, 2023 by jbetker

I’ve been at OpenAI for almost a year now. In that time, I’ve trained a **lot** of generative models. More than anyone really has any right to train. As I’ve spent these hours observing the effects of tweaking various model configurations and hyperparameters, one thing that has struck me is the similarities in between all the training runs.

It’s becoming awfully clear to me that these models are truly approximating their datasets to an incredible degree. What that means is not only that they learn what it means to be a dog or a cat, but the interstitial frequencies between distributions that don’t matter, like what photos humans are likely to take or words humans commonly write down.

What this manifests as is – trained on the same dataset for long enough, pretty much every model with enough weights and training time converges to the same point. Sufficiently large diffusion conv-unets produce the same images as ViT generators. AR sampling produces the same images as diffusion.

This is a surprising observation! It implies that model behavior is not determined by architecture, hyperparameters, or optimizer choices. It’s determined by your dataset, nothing else. Everything else is a means to an end in efficiently delivery compute to approximating that dataset.

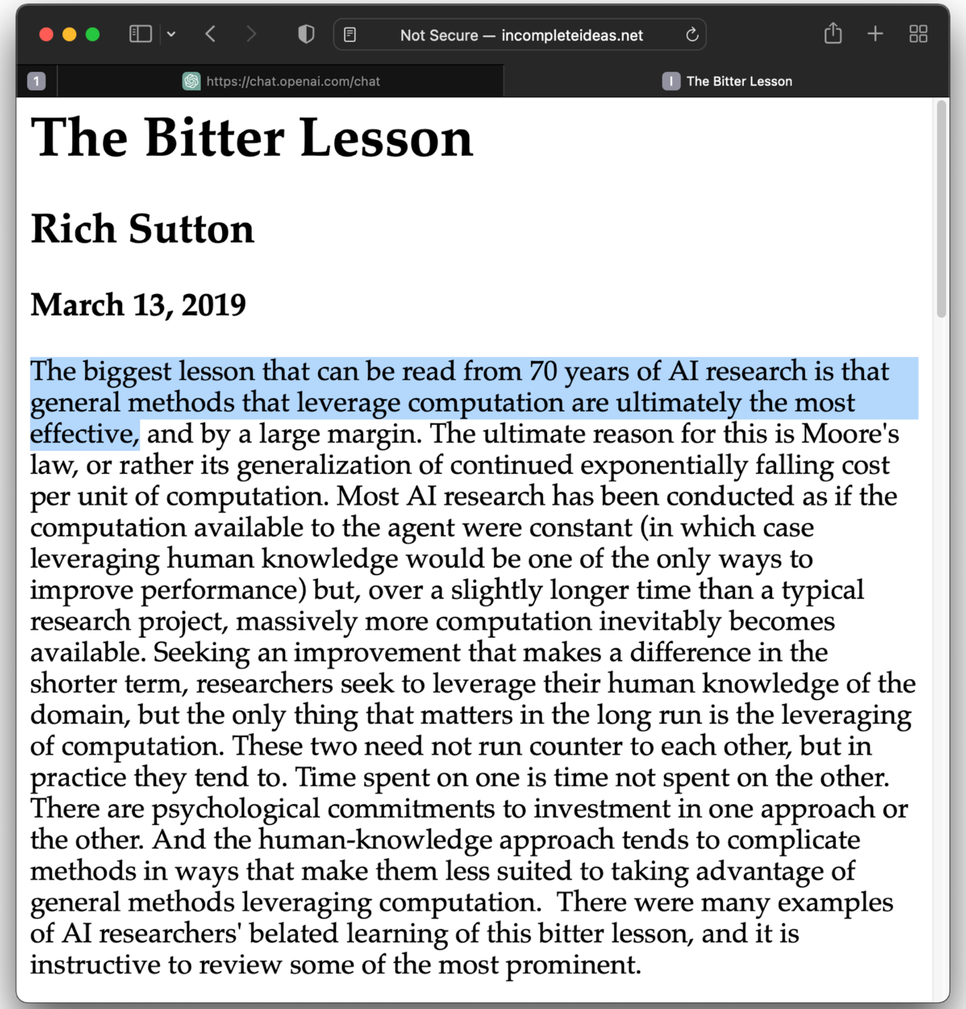
Then, when you refer to “Lambda”, “ChatGPT”, “Bard”, or “Claude” then, it’s not the model weights that you are referring to. It’s the dataset.

https://nonint.com/2023/06/10/the-it-in-ai-models-is-the-dataset/

How much inductive bias?

http://www.incompleteideas.net/IncIdeas/BitterLesson.html

https://staff.fnwi.uva.nl/m.welling/wp-content/uploads/Model-versus-Data-AI-1.pdf

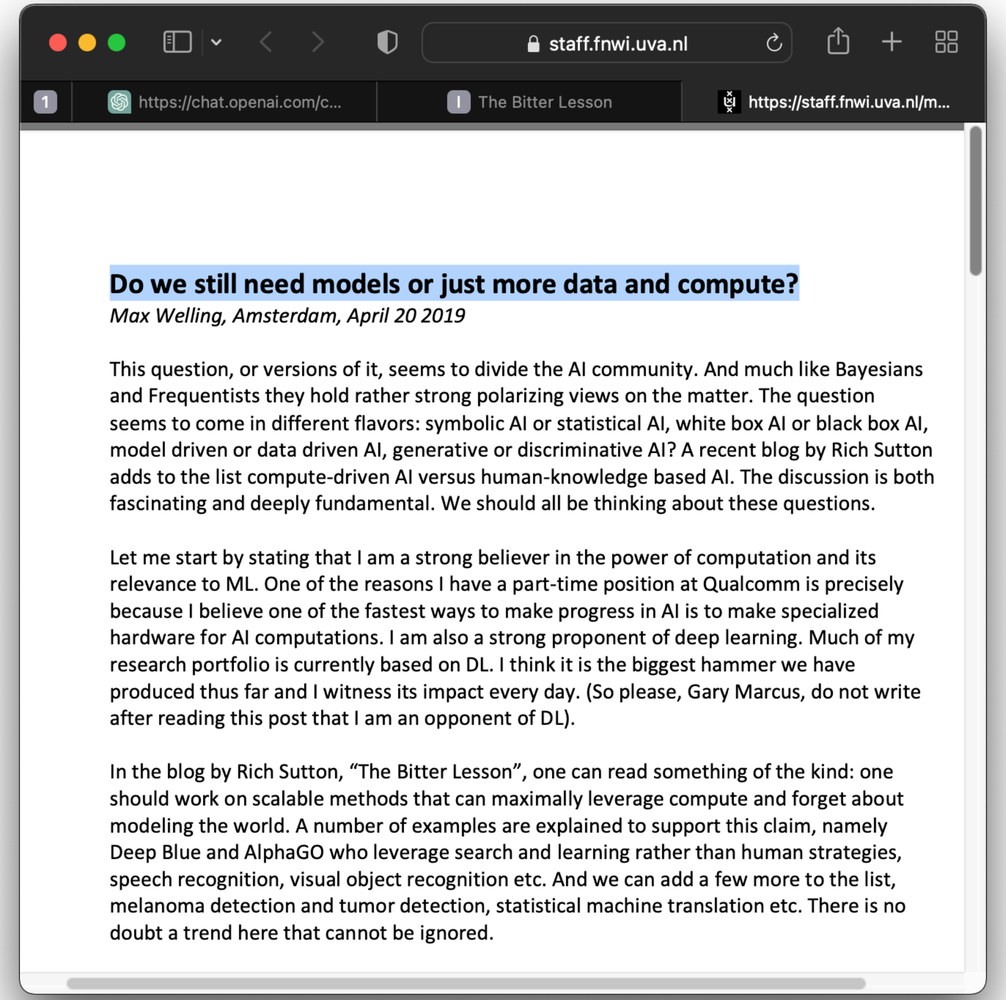


The Bitter Lesson

Rich Sutton

March 13, 2019

The biggest lesson that can be read from 70 years of AI research is that general methods that leverage computation are ultimately the most effective, and by a large margin. The ultimate reason for this is Moore's law, or rather its generalization of continued exponentially falling cost per unit of computation. Most AI research has been conducted as if the computation available to the agent were constant (in which case leveraging human knowledge would be one of the only ways to improve performance) but, over a slightly longer time than a typical research project, massively more computation inevitably becomes available. Seeking an improvement that makes a difference in the shorter term, researchers seek to leverage their human knowledge of the domain, but the only thing that matters in the long run is the leveraging of computation. These two need not run counter to each other, but in practice they tend to. Time spent on one is time not spent on the other. There are psychological commitments to investment in one approach or the other. And the human-knowledge approach tends to complicate methods in ways that make them less suited to taking advantage of general methods leveraging computation. There were many examples of AI researchers' belated learning of this bitter lesson, and it is instructive to review some of the most prominent.



Do we still need models or just more data and compute?

Max Welling, Amsterdam, April 20 2019

This question, or versions of it, seems to divide the AI community. And much like Bayesians and Frequentists they hold rather strong polarizing views on the matter. The question seems to come in different flavors: symbolic AI or statistical AI, white box AI or black box AI, model driven or data driven AI, generative or discriminative AI? A recent blog by Rich Sutton adds to the list compute-driven AI versus human-knowledge based AI. The discussion is both fascinating and deeply fundamental. We should all be thinking about these questions.

Let me start by stating that I am a strong believer in the power of computation and its relevance to ML. One of the reasons I have a part-time position at Qualcomm is precisely because I believe one of the fastest ways to make progress in AI is to make specialized hardware for AI computations. I am also a strong proponent of deep learning. Much of my research portfolio is currently based on DL. I think it is the biggest hammer we have produced thus far and I witness its impact every day. (So please, Gary Marcus, do not write after reading this post that I am an opponent of DL).

In the blog by Rich Sutton, "The Bitter Lesson", one can read something of the kind: one should work on scalable methods that can maximally leverage compute and forget about modeling the world. A number of examples are explained to support this claim, namely Deep Blue and AlphaGO who leverage search and learning rather than human strategies, speech recognition, visual object recognition etc. And we can add a few more to the list, melanoma detection and tumor detection, statistical machine translation etc. There is no doubt a trend here that cannot be ignored.

more data and compute

more physics and symmetries

Molecule representations and inductive biases

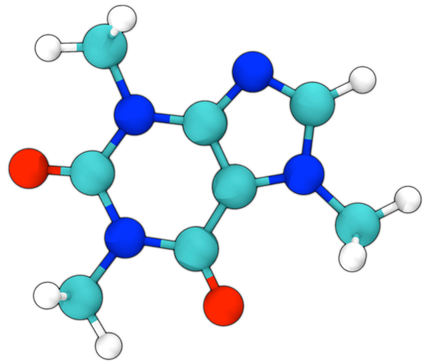
SMLIES or XYZ/CIF/PDB file

```
HEADER    CAFFEINE
COMPND    CAFFEINE
AUTHOR    GENERATED BY ChatGPT

ATOM      1  C   LIG A   1       0.000   0.000   0.000
ATOM      2  N   LIG A   1       1.289   0.000   0.000
ATOM      3  C   LIG A   1       1.463   1.192   0.000
ATOM      4  N   LIG A   1       2.644   1.192   0.000
ATOM      5  C   LIG A   1       2.818   2.384   0.000
ATOM      6  N   LIG A   1       4.089   2.384   0.000
ATOM      7  C   LIG A   1       4.263   3.576   0.000
                          .
                          .
                          .
```

Language model

Atom coordinates



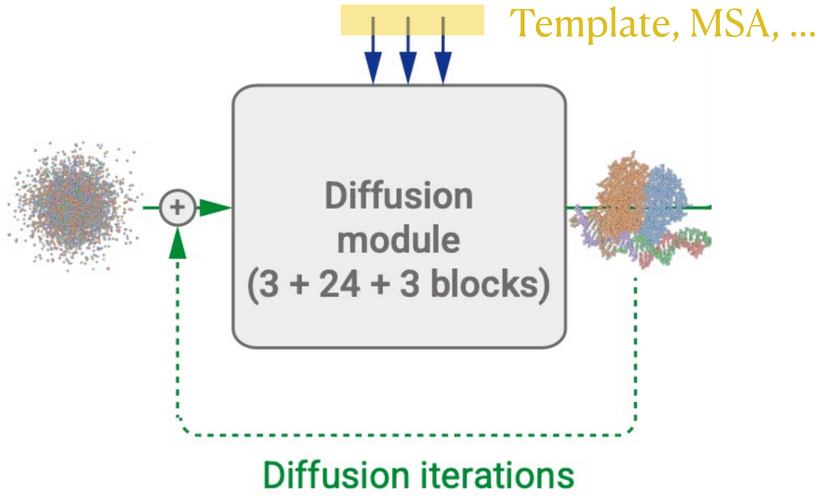
Equivariant neural network

more data and compute

more physics and symmetries

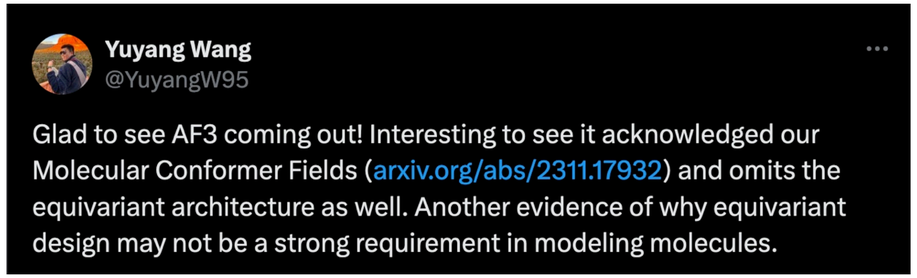
How much inductive bias?

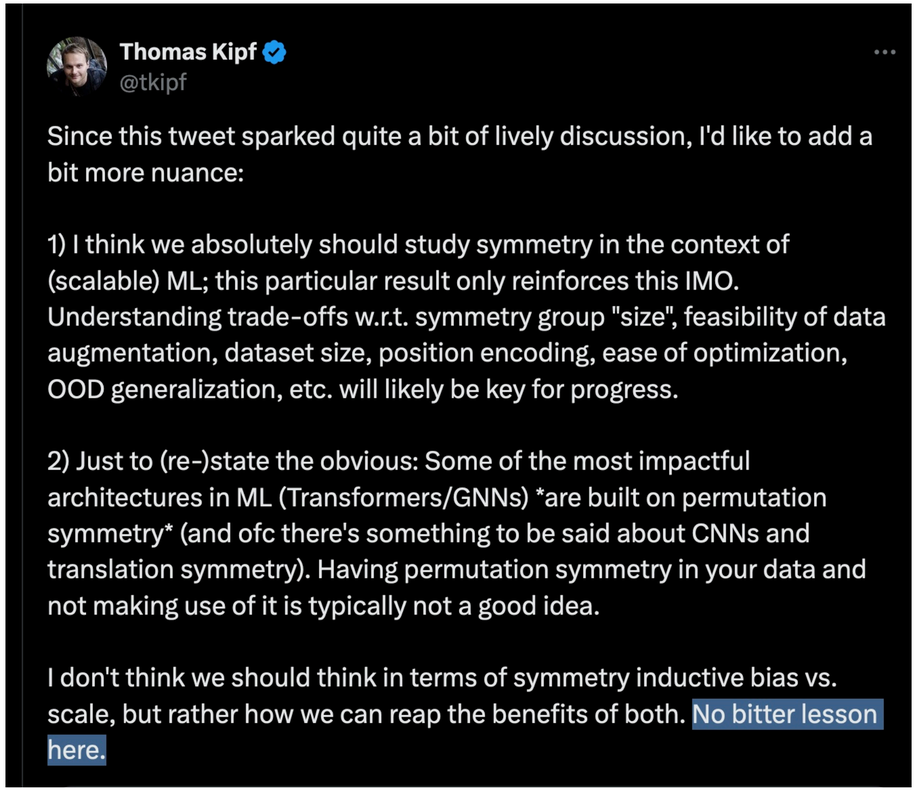
Abramson et al, AlphaFold3, Nature 2024



Similarly to some recent work[35], we find that no invariance or equivariance with respect to global rotations and translation of the molecule are required in the architecture and so we omit them to simplify the machine learning architecture.

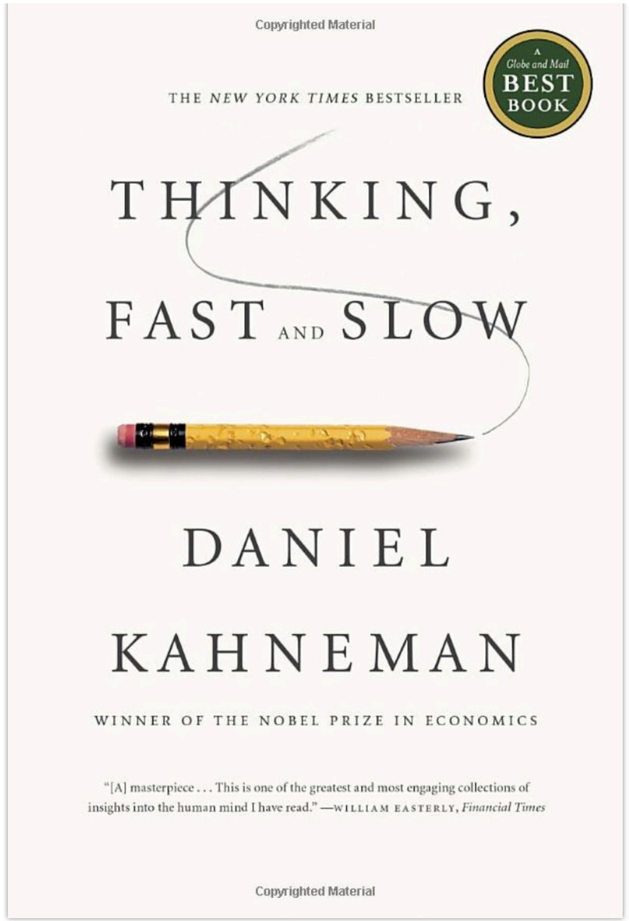
[35] **Swallowing the Bitter Pill**: Simplified Scalable Conformer Generation, 2311.17932

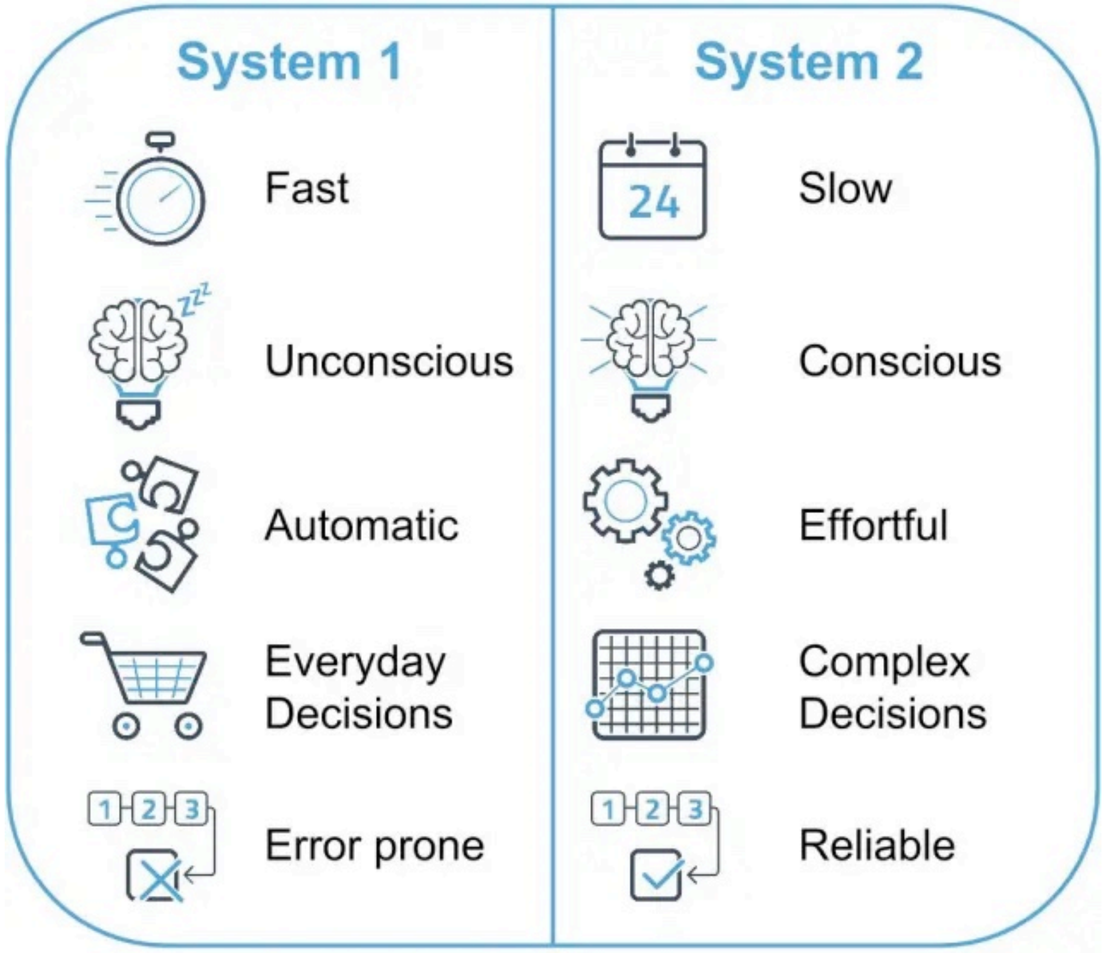




Can they reason ?

Autoregressive language models are fast thinkers





Fast thinkers rely on good intuitions

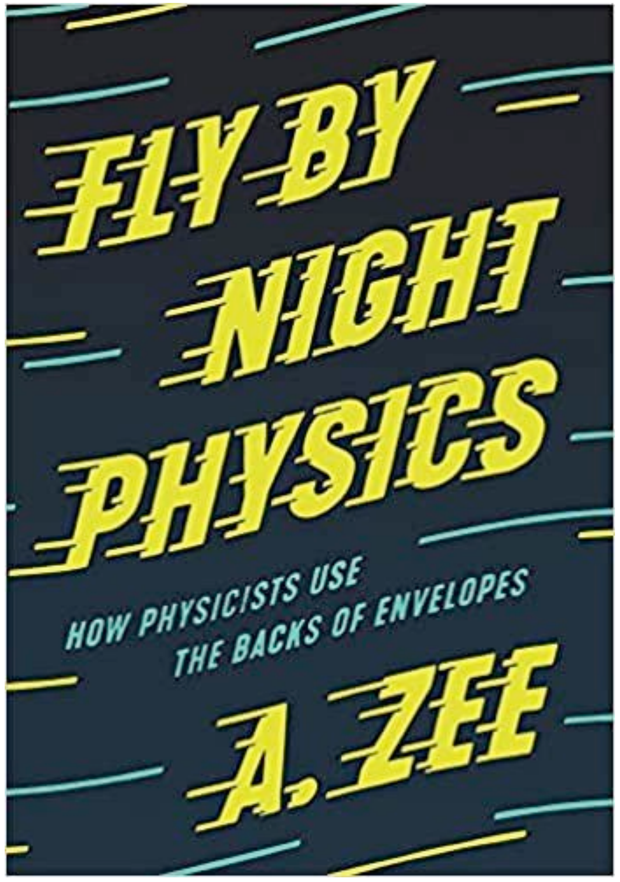
物理直觉是如何养成的？徐一鸿：思考比计算要困难得多

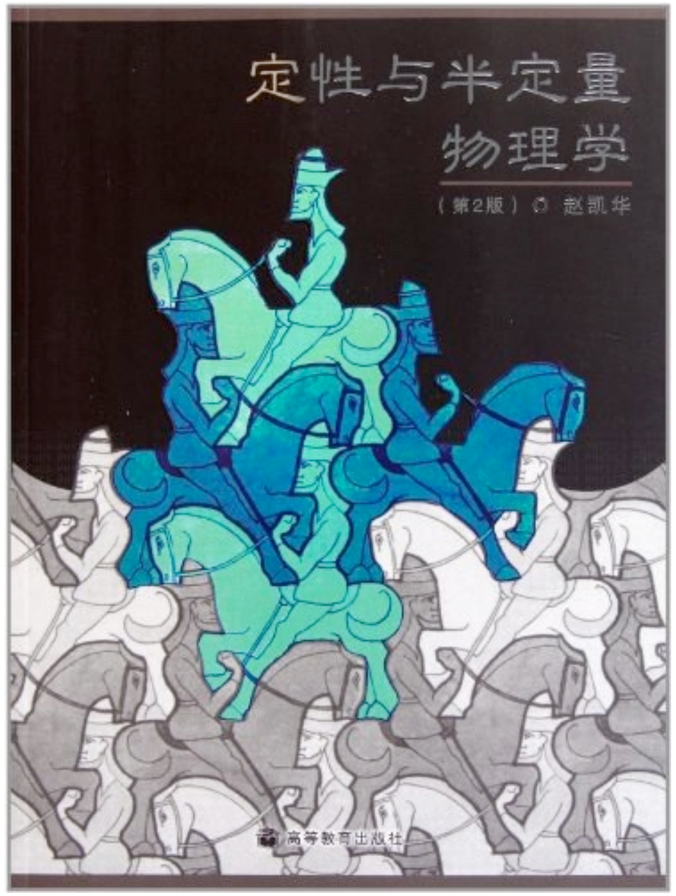
澎湃新闻记者 曹年润
2022-12-14 08:28 来源：澎湃新闻

· “重要的是要帮助学生培养更多的物理直觉，这也是我写这本书的原因之一。如何培养物理直觉？我在书里提到了两种可能性：天生就拥有物理直觉，或者通过不断练习发展物理直觉。”

徐一鸿教授。

“你应该先思考。思考比计算要困难得多，大多数人都可以坐下来计算，但不计算就思考问题，这是极其困难的。”





System 1 thinking in physics:
getting answers quickly without lengthy calculations

“Never never calculate unless you already know the answer! ”—John Wheeler

Do they understand physics?

https://openai.com/index/sora/

http://ai.ruc.edu.cn/newslist/newsdetail/20240326001.html

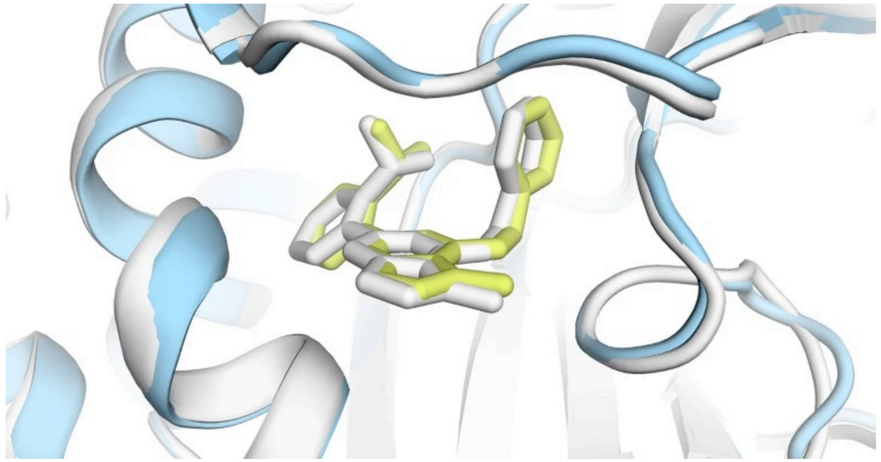
“What I can not create, I do not understand”

—Richard Feynman

Fold by intuition vs fold by equation

Both integrate Langevin dyanmics $X_{t+1} = X_t + \frac{\eta_t}{2} s(X_t, t) + \sqrt{\eta_t}\epsilon$

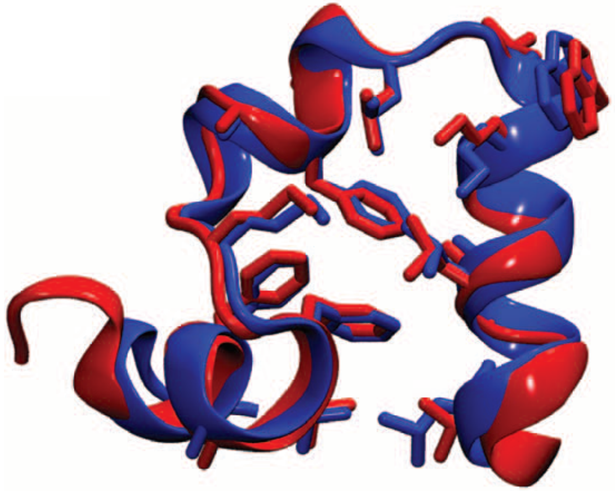
Data-driven generation AlphaFold3



Abramson et al, Nature 2024

The diffusion model may generate right conformations via unphysical pathways

Physics-based molecular dynamics



Shaw et al, Science 2010

Physical force fields may face difficulties in sampling rough energy landscapes

生成模型四问

① 哪种模型好?

大数据：无所谓
小数据：看模态

② 要多少先验?

抓主要矛盾

③ 能不能推理?

哪怕不能，“直觉”也很可贵

④ 懂不懂物理?

不是真的懂，但未必是坏事

What can generative models do ?

Appreciation

Likelihood estimation

$$\ln p(\boldsymbol{X})$$

Anomaly detention

Generation

(Un)conditional sampling

$$\boldsymbol{X} \sim p(\boldsymbol{X}) \text{ or } p(\boldsymbol{X} \mid y)$$

De novo design

Text-to-image

Question answering

Generative models compress the training data to extract language/image/physics/chemistry intuitions

What is next?

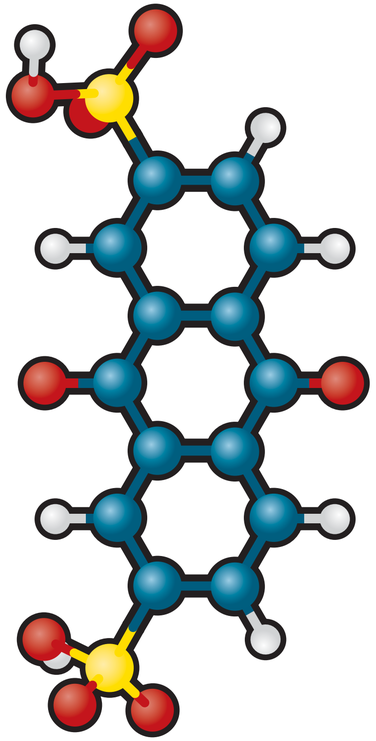
Generative AI for It

"It from Bit", John Wheeler, 1989

in Informaiton, physics, quantum: the search for links

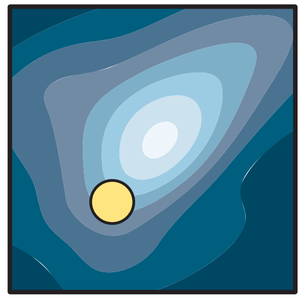
Generative AI for matter engineering

Chemical space X



Prediction $p(y \mid X)$

Generation $p(X \mid y)$

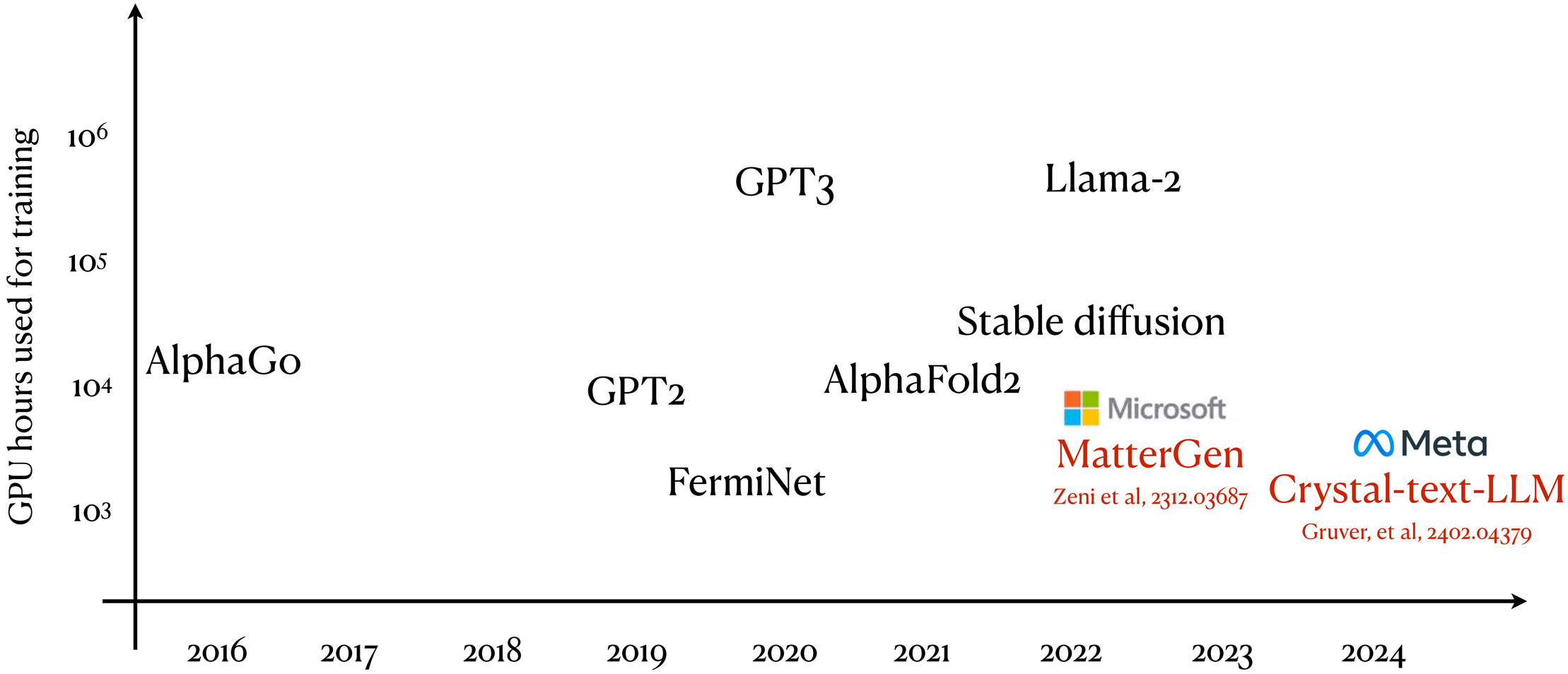


Control variable y

Inverse molecular design using machine learning, Sanchez-Lengeling & Aspuru-Guzik, Science '18
Inverse design in search of materials with target functionalities, Zunger, Nature Reviews Chemistry '18

“an image of beautiful crystals in 16:9”

$$\text{pixels} \sim p(\text{pixels} \mid \text{texts})$$



Is there a bitter lesson ?

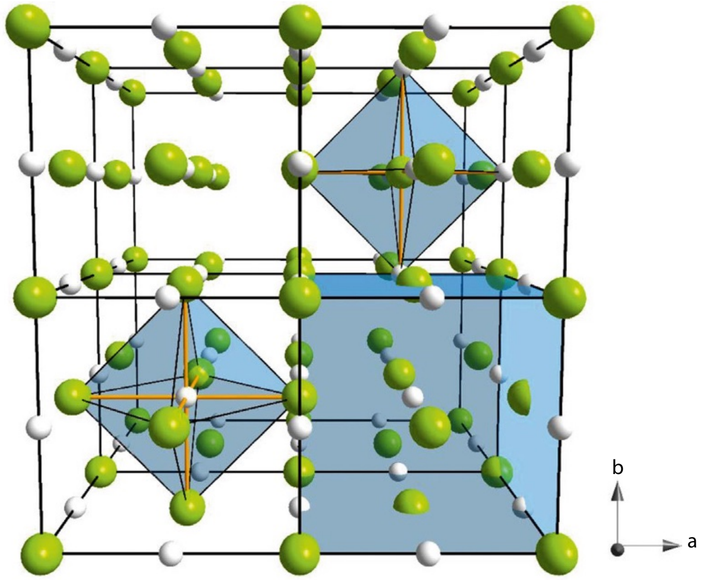
The biggest lesson that can be read from 70 years of AI research is that general methods that leverage computation are ultimately the most effective

—Rich Sutton 2019

```
data_Na1Cl1
_symmetry_space_group_name_H-M   'P1'
_cell_length_a   3.9893
_cell_length_b   3.9893
_cell_length_c   3.9893
_cell_angle_alpha   60.0000
_cell_angle_beta   60.0000
_cell_angle_gamma   60.0000
_symmetry_Int_Tables_number   1
_chemical_formula_structural   NaCl
_chemical_formula_sum   'Na1 Cl1'
_cell_volume   44.8931
_cell_formula_units_Z   1
loop_
 _symmetry_equiv_pos_site_id
 _symmetry_equiv_pos_as_xyz
  1  'x, y, z'
loop_
 _atom_site_type_symbol
 _atom_site_label
 _atom_site_symmetry_multiplicity
 _atom_site_fract_x
 _atom_site_fract_y
 _atom_site_fract_z
 _atom_site_occupancy
  Cl  Cl0  1  0.0000  0.0000  0.0000  1
  Na Na1 1  0.5000  0.5000  0.5000  1
```

Flam-Shepherd et al, 2305.05708
Antunes et al, 2307.04340
Gruver, et al, 2402.04379...

CALYPSO
USPEX
AIRSS
GNoME, ...



Large language model

Energy-based structure prediction

more data and compute

more physics and symmetries

We have much less crystal data

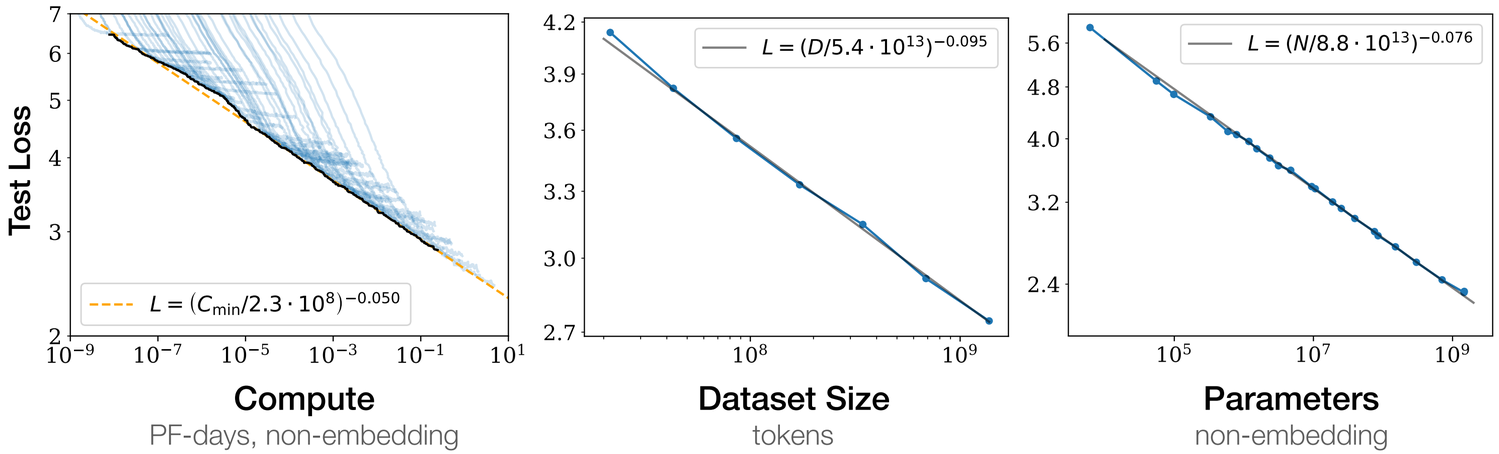
COMMON CRAWL

Over 250 billion pages

ICSD

> 291,000 crystal structures

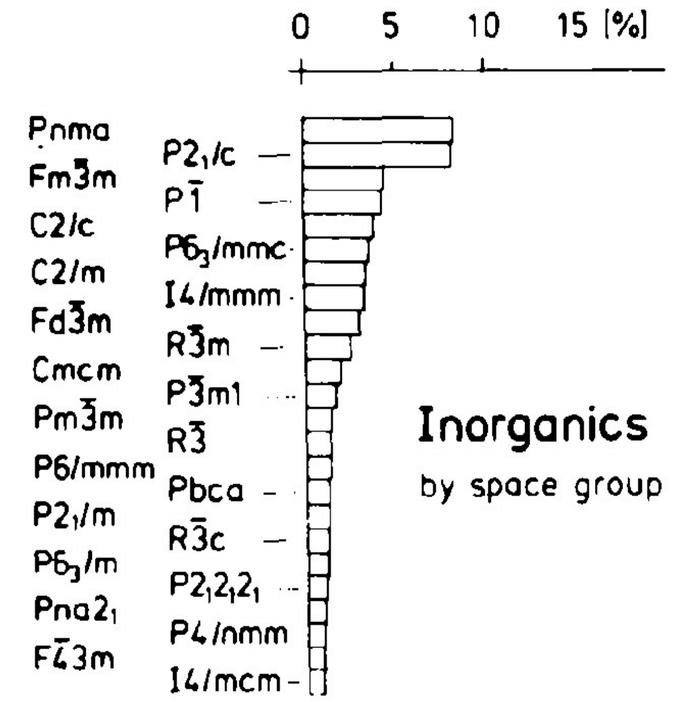
Data, compute, and parameters need to scale simultaneously Kaplan et al, 2001.08361



What is the natural bitstream representation of crystals?

https://crystalsymmetry.files.wordpress.com/2014/08/poster_230_the_space_group_list_project.png

Space groups quantify Nature' symmetry preference



P1 is rare!

Wyckoff Positions of Group *P*1 (No. 1)

Multiplicity	Wyckoff letter	Site symmetry	Coordinates
1	a	1	(x,y,z)

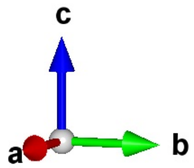
Wyckoff Positions of Group *Fm*-3*m* (No. 225)

Multiplicity	Wyckoff letter	Site symmetry	Coordinates
			(0,0,0) + (0,1/2,1/2) + (1/2,0,1/2) + (1/2,1/2,0) +
192	l	1	(x,y,z) (-x,-y,z) (-x,y,-z) (x,-y,-z) (z,x,y) (z,-x,-y) (-z,-x,y) (-z,x,-y) (y,z,x) (-y,z,-x) (y,-z,-x) (-y,-z,x) (y,x,-z) (-y,-x,-z) (y,-x,z) (-y,x,z) (x,z,-y) (-x,z,y) (-x,-z,-y) (x,-z,y) (z,y,-x) (z,-y,x) (-z,y,x) (-z,-y,-x) (-x,-y,-z) (x,y,-z) (x,-y,z) (-x,y,z) (-z,-x,-y) (-z,x,y) (z,x,-y) (z,-x,y) (-y,-z,-x) (y,-z,x) (-y,z,x) (y,z,-x) (-y,-x,z) (y,x,z) (-y,x,-z) (y,-x,-z) (-x,-z,y) (x,-z,-y) (x,z,y) (-x,z,-y) (-z,-y,x) (-z,y,-x) (z,-y,-x) (z,y,x)
96	k	..m	(x,x,z) (-x,-x,z) (-x,x,-z) (x,-x,-z) (z,x,x) (z,-x,-x) (-z,-x,x) (-z,x,-x) (x,z,x) (-x,z,-x) (x,-z,-x) (-x,-z,x) (x,x,-z) (-x,-x,-z) (x,-x,z) (-x,x,z) (x,z,-x) (-x,z,x) (-x,-z,-x) (x,-z,x) (z,x,-x) (z,-x,x) (-z,x,x) (-z,-x,-x)
96	j	m..	(0,y,z) (0,-y,z) (0,y,-z) (0,-y,-z) (z,0,y) (z,0,-y) (-z,0,y) (-z,0,-y) (y,z,0) (-y,z,0) (y,-z,0) (-y,-z,0) (y,0,-z) (-y,0,-z) (y,0,z) (-y,0,z) (0,z,-y) (0,z,y) (0,-z,-y) (0,-z,y) (z,y,0) (z,-y,0) (-z,y,0) (-z,-y,0)
48	i	m.m 2	(1/2,y,y) (1/2,-y,y) (1/2,y,-y) (1/2,-y,-y) (y,1/2,y) (y,1/2,-y) (-y,1/2,y) (-y,1/2,-y) (y,y,1/2) (-y,y,1/2) (y,-y,1/2) (-y,-y,1/2)
48	h	m.m 2	(0,y,y) (0,-y,y) (0,y,-y) (0,-y,-y) (y,0,y) (y,0,-y) (-y,0,y) (-y,0,-y) (y,y,0) (-y,y,0) (y,-y,0) (-y,-y,0)
48	g	2.m m	(x,1/4,1/4) (-x,3/4,1/4) (1/4,x,1/4) (1/4,-x,3/4) (1/4,1/4,x) (3/4,1/4,-x) (1/4,x,3/4) (3/4,-x,3/4) (x,1/4,3/4) (-x,1/4,1/4) (1/4,1/4,-x) (1/4,3/4,x)
32	f	.3m	(x,x,x) (-x,-x,x) (-x,x,-x) (x,-x,-x) (x,x,-x) (-x,-x,-x) (x,-x,x) (-x,x,x)
24	e	4m. m	(x,0,0) (-x,0,0) (0,x,0) (0,-x,0) (0,0,x) (0,0,-x)
24	d	m.m m	(0,1/4,1/4) (0,3/4,1/4) (1/4,0,1/4) (1/4,0,3/4) (1/4,1/4,0) (3/4,1/4,0)
8	c	-43m	(1/4,1/4,1/4) (1/4,1/4,3/4)
4	b	m-3m	(1/2,1/2,1/2)
4	a	m-3m	(0,0,0)

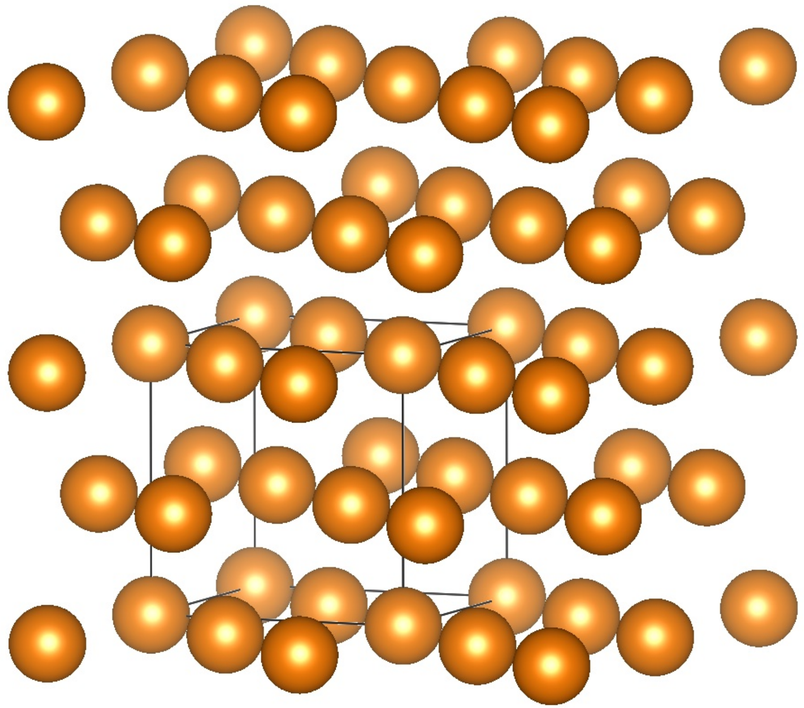
https://www.cryst.ehu.es/cryst/get_wp.html

Wyckoff Positions of Group *Fm*-3*m* (No. 225)

Multiplicity	Wyckoff letter	Site symmetry	Coordinates
			(0,0,0) + (0,1/2,1/2) + (1/2,0,1/2) + (1/2,1/2,0) +
192	l	1	(x,y,z) (-x,-y,z) (-x,y,-z) (x,-y,-z) (z,x,y) (z,-x,-y) (-z,-x,y) (-z,x,-y) (y,z,x) (-y,z,-x) (y,-z,-x) (-y,-z,x) (y,x,-z) (-y,-x,-z) (y,-x,z) (-y,x,z) (x,z,-y) (-x,z,y) (-x,-z,-y) (x,-z,y) (z,y,-x) (z,-y,x) (-z,y,x) (-z,-y,-x) (-x,-y,-z) (x,y,-z) (x,-y,z) (-x,y,z) (-z,-x,-y) (-z,x,y) (z,x,-y) (z,-x,y) (-y,-z,-x) (y,-z,x) (-y,z,x) (y,z,-x) (-y,-x,z) (y,x,z) (-y,x,-z) (y,-x,-z) (-x,-z,y) (x,-z,-y) (x,z,y) (-x,z,-y) (-z,-y,x) (-z,y,-x) (z,-y,-x) (z,y,x)
96	k	..m	(x,x,z) (-x,-x,z) (-x,x,-z) (x,-x,-z) (z,x,x) (z,-x,-x) (-z,-x,x) (-z,x,-x) (x,z,x) (-x,z,-x) (x,-z,-x) (-x,-z,x) (x,x,-z) (-x,-x,-z) (x,-x,z) (-x,x,z) (x,z,-x) (-x,z,x) (-x,-z,-x) (x,-z,x) (z,x,-x) (z,-x,x) (-z,x,x) (-z,-x,-x)
96	j	m..	(0,y,z) (0,-y,z) (0,y,-z) (0,-y,-z) (z,0,y) (z,0,-y) (-z,0,y) (-z,0,-y) (y,z,0) (-y,z,0) (y,-z,0) (-y,-z,0) (y,0,-z) (-y,0,-z) (y,0,z) (-y,0,z) (0,z,-y) (0,z,y) (0,-z,-y) (0,-z,y) (z,y,0) (z,-y,0) (-z,y,0) (-z,-y,0)
48	i	m.m 2	(1/2,y,y) (1/2,-y,y) (1/2,y,-y) (1/2,-y,-y) (y,1/2,y) (y,1/2,-y) (-y,1/2,y) (-y,1/2,-y) (y,y,1/2) (-y,y,1/2) (y,-y,1/2) (-y,-y,1/2)
48	h	m.m 2	(0,y,y) (0,-y,y) (0,y,-y) (0,-y,-y) (y,0,y) (y,0,-y) (-y,0,y) (-y,0,-y) (y,y,0) (-y,y,0) (y,-y,0) (-y,-y,0)
48	g	2.m m	(x,1/4,1/4) (-x,3/4,1/4) (1/4,x,1/4) (1/4,-x,3/4) (1/4,1/4,x) (3/4,1/4,-x) (1/4,x,3/4) (3/4,-x,3/4) (x,1/4,3/4) (-x,1/4,1/4) (1/4,1/4,-x) (1/4,3/4,x)
32	f	.3m	(x,x,x) (-x,-x,x) (-x,x,-x) (x,-x,-x) (x,x,-x) (-x,-x,-x) (x,-x,x) (-x,x,x)
24	e	4m. m	(x,0,0) (-x,0,0) (0,x,0) (0,-x,0) (0,0,x) (0,0,-x)
24	d	m.m m	(0,1/4,1/4) (0,3/4,1/4) (1/4,0,1/4) (1/4,0,3/4) (1/4,1/4,0) (3/4,1/4,0)
8	c	-43m	(1/4,1/4,1/4) (1/4,1/4,3/4)
4	b	m-3m	(1/2,1/2,1/2)
4	a	m-3m	(0,0,0)



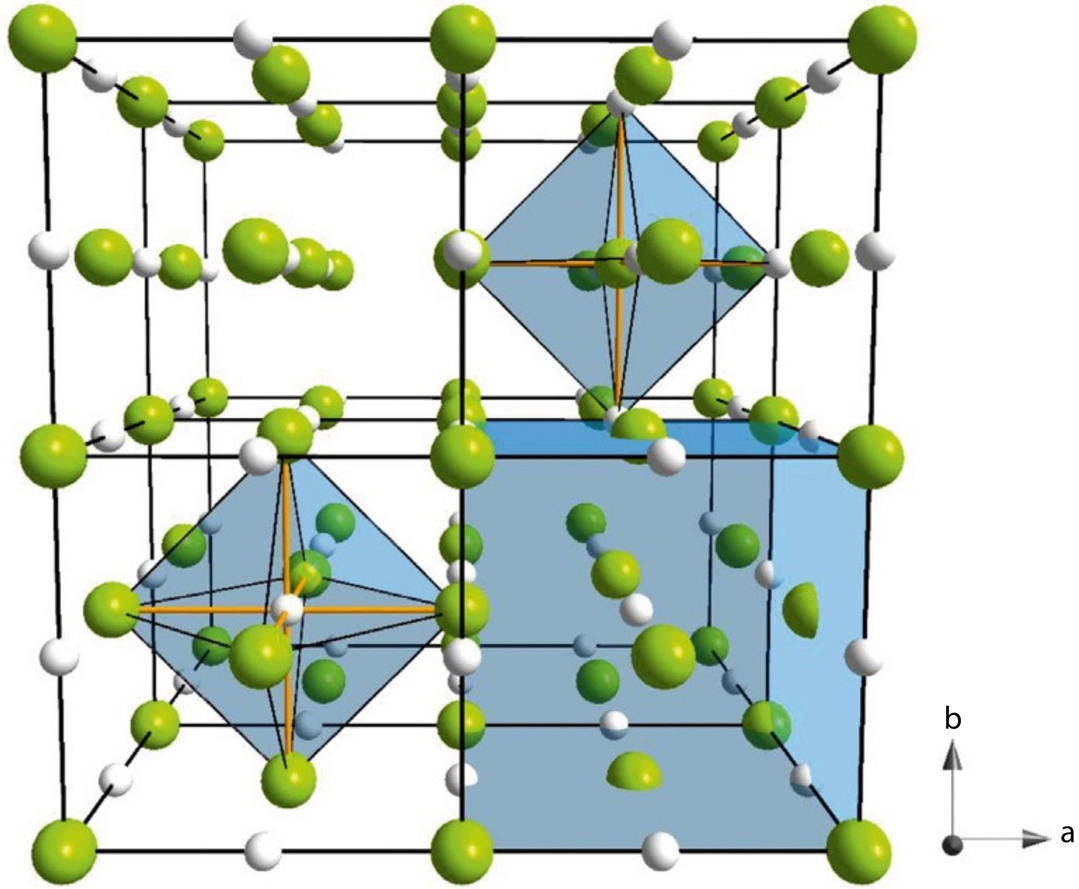
Copper



Wyckoff Positions of Group *Fm*-3*m* (No. 225)

Multiplicity	Wyckoff letter	Site symmetry	Coordinates
			(0,0,0) + (0,1/2,1/2) + (1/2,0,1/2) + (1/2,1/2,0) +
192	l	1	(x,y,z) (-x,-y,z) (-x,y,-z) (x,-y,-z) (z,x,y) (z,-x,-y) (-z,-x,y) (-z,x,-y) (y,z,x) (-y,z,-x) (y,-z,-x) (-y,-z,x) (y,x,-z) (-y,-x,-z) (y,-x,z) (-y,x,z) (x,z,-y) (-x,z,y) (-x,-z,-y) (x,-z,y) (z,y,-x) (z,-y,x) (-z,y,x) (-z,-y,-x) (-x,-y,-z) (x,y,-z) (x,-y,z) (-x,y,z) (-z,-x,-y) (-z,x,y) (z,x,-y) (z,-x,y) (-y,-z,-x) (y,-z,x) (-y,z,x) (y,z,-x) (-y,-x,z) (y,x,z) (-y,x,-z) (y,-x,-z) (-x,-z,y) (x,-z,-y) (x,z,y) (-x,z,-y) (-z,-y,x) (-z,y,-x) (z,-y,-x) (z,y,x)
96	k	..m	(x,x,z) (-x,-x,z) (-x,x,-z) (x,-x,-z) (z,x,x) (z,-x,-x) (-z,-x,x) (-z,x,-x) (x,z,x) (-x,z,-x) (x,-z,-x) (-x,-z,x) (x,x,-z) (-x,-x,-z) (x,-x,z) (-x,x,z) (x,z,-x) (-x,z,x) (-x,-z,-x) (x,-z,x) (z,x,-x) (z,-x,x) (-z,x,x) (-z,-x,-x)
96	j	m..	(0,y,z) (0,-y,z) (0,y,-z) (0,-y,-z) (z,0,y) (z,0,-y) (-z,0,y) (-z,0,-y) (y,z,0) (-y,z,0) (y,-z,0) (-y,-z,0) (y,0,-z) (-y,0,-z) (y,0,z) (-y,0,z) (0,z,-y) (0,z,y) (0,-z,-y) (0,-z,y) (z,y,0) (z,-y,0) (-z,y,0) (-z,-y,0)
48	i	m.m 2	(1/2,y,y) (1/2,-y,y) (1/2,y,-y) (1/2,-y,-y) (y,1/2,y) (y,1/2,-y) (-y,1/2,y) (-y,1/2,-y) (y,y,1/2) (-y,y,1/2) (y,-y,1/2) (-y,-y,1/2)
48	h	m.m 2	(0,y,y) (0,-y,y) (0,y,-y) (0,-y,-y) (y,0,y) (y,0,-y) (-y,0,y) (-y,0,-y) (y,y,0) (-y,y,0) (y,-y,0) (-y,-y,0)
48	g	2.m m	(x,1/4,1/4) (-x,3/4,1/4) (1/4,x,1/4) (1/4,-x,3/4) (1/4,1/4,x) (3/4,1/4,-x) (1/4,x,3/4) (3/4,-x,3/4) (x,1/4,3/4) (-x,1/4,1/4) (1/4,1/4,-x) (1/4,3/4,x)
32	f	.3m	(x,x,x) (-x,-x,x) (-x,x,-x) (x,-x,-x) (x,x,-x) (-x,-x,-x) (x,-x,x) (-x,x,x)
24	e	4m. m	(x,0,0) (-x,0,0) (0,x,0) (0,-x,0) (0,0,x) (0,0,-x)
24	d	m.m m	(0,1/4,1/4) (0,3/4,1/4) (1/4,0,1/4) (1/4,0,3/4) (1/4,1/4,0) (3/4,1/4,0)
8	c	-43m	(1/4,1/4,1/4) (1/4,1/4,3/4)
4	b	m-3m	(1/2,1/2,1/2)
4	a	m-3m	(0,0,0)

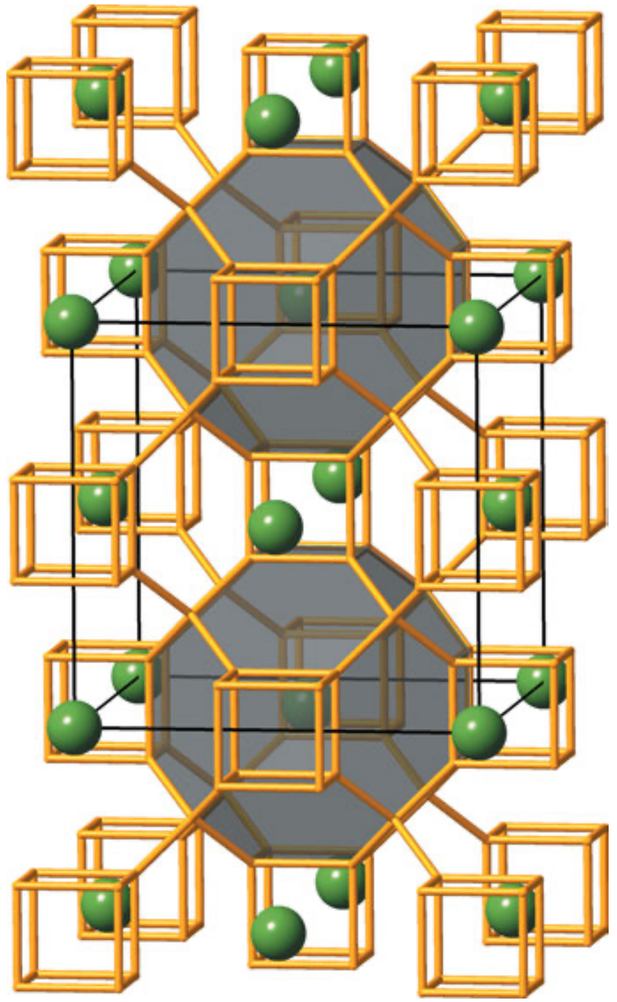
NaCl



Wyckoff Positions of Group *Fm*-3*m* (No. 225)

Multiplicity	Wyckoff letter	Site symmetry	Coordinates
			(0,0,0) + (0,1/2,1/2) + (1/2,0,1/2) + (1/2,1/2,0) +
192	l	1	(x,y,z) (-x,-y,z) (-x,y,-z) (x,-y,-z) (z,x,y) (z,-x,-y) (-z,-x,y) (-z,x,-y) (y,z,x) (-y,z,-x) (y,-z,-x) (-y,-z,x) (y,x,-z) (-y,-x,-z) (y,-x,z) (-y,x,z) (x,z,-y) (-x,z,y) (-x,-z,-y) (x,-z,y) (z,y,-x) (z,-y,x) (-z,y,x) (-z,-y,-x) (-x,-y,-z) (x,y,-z) (x,-y,z) (-x,y,z) (-z,-x,-y) (-z,x,y) (z,x,-y) (z,-x,y) (-y,-z,-x) (y,-z,x) (-y,z,x) (y,z,-x) (-y,-x,z) (y,x,z) (-y,x,-z) (y,-x,-z) (-x,-z,y) (x,-z,-y) (x,z,y) (-x,z,-y) (-z,-y,x) (-z,y,-x) (z,-y,-x) (z,y,x)
96	k	..m	(x,x,z) (-x,-x,z) (-x,x,-z) (x,-x,-z) (z,x,x) (z,-x,-x) (-z,-x,x) (-z,x,-x) (x,z,x) (-x,z,-x) (x,-z,-x) (-x,-z,x) (x,x,-z) (-x,-x,-z) (x,-x,z) (-x,x,z) (x,z,-x) (-x,z,x) (-x,-z,-x) (x,-z,x) (z,x,-x) (z,-x,x) (-z,x,x) (-z,-x,-x)
96	j	m..	(0,y,z) (0,-y,z) (0,y,-z) (0,-y,-z) (z,0,y) (z,0,-y) (-z,0,y) (-z,0,-y) (y,z,0) (-y,z,0) (y,-z,0) (-y,-z,0) (y,0,-z) (-y,0,-z) (y,0,z) (-y,0,z) (0,z,-y) (0,z,y) (0,-z,-y) (0,-z,y) (z,y,0) (z,-y,0) (-z,y,0) (-z,-y,0)
48	i	m.m 2	(1/2,y,y) (1/2,-y,y) (1/2,y,-y) (1/2,-y,-y) (y,1/2,y) (y,1/2,-y) (-y,1/2,y) (-y,1/2,-y) (y,y,1/2) (-y,y,1/2) (y,-y,1/2) (-y,-y,1/2)
48	h	m.m 2	(0,y,y) (0,-y,y) (0,y,-y) (0,-y,-y) (y,0,y) (y,0,-y) (-y,0,y) (-y,0,-y) (y,y,0) (-y,y,0) (y,-y,0) (-y,-y,0)
48	g	2.m m	(x,1/4,1/4) (-x,3/4,1/4) (1/4,x,1/4) (1/4,-x,3/4) (1/4,1/4,x) (3/4,1/4,-x) (1/4,x,3/4) (3/4,-x,3/4) (x,1/4,3/4) (-x,1/4,1/4) (1/4,1/4,-x) (1/4,3/4,x)
32	f	.3m	(x,x,x) (-x,-x,x) (-x,x,-x) (x,-x,-x) (x,x,-x) (-x,-x,-x) (x,-x,x) (-x,x,x)
24	e	4m. m	(x,0,0) (-x,0,0) (0,x,0) (0,-x,0) (0,0,x) (0,0,-x)
24	d	m.m m	(0,1/4,1/4) (0,3/4,1/4) (1/4,0,1/4) (1/4,0,3/4) (1/4,1/4,0) (3/4,1/4,0)
8	c	-43m	(1/4,1/4,1/4) (1/4,1/4,3/4)
4	b	m-3m	(1/2,1/2,1/2)
4	a	m-3m	(0,0,0)

LaH_{10}

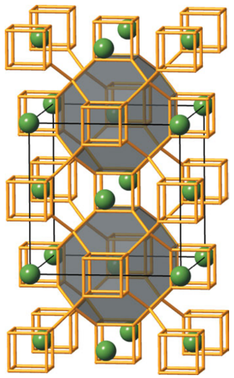


CrystalFormer

Zhendong Cao et al, 2403.15734

融合空间群对称性的晶体语言模型

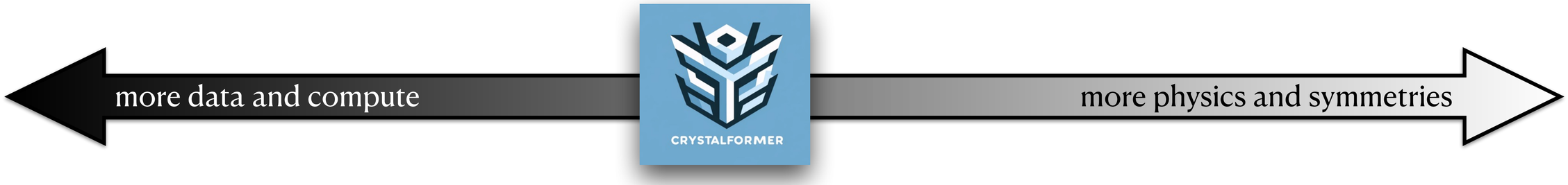
LaH_{10} 225-a-La-0-0-0-c-H-1/4-1/4-1/4-f-H-0.375-0.375-0.375-X-5.1-5.1-5.1-90-90-90



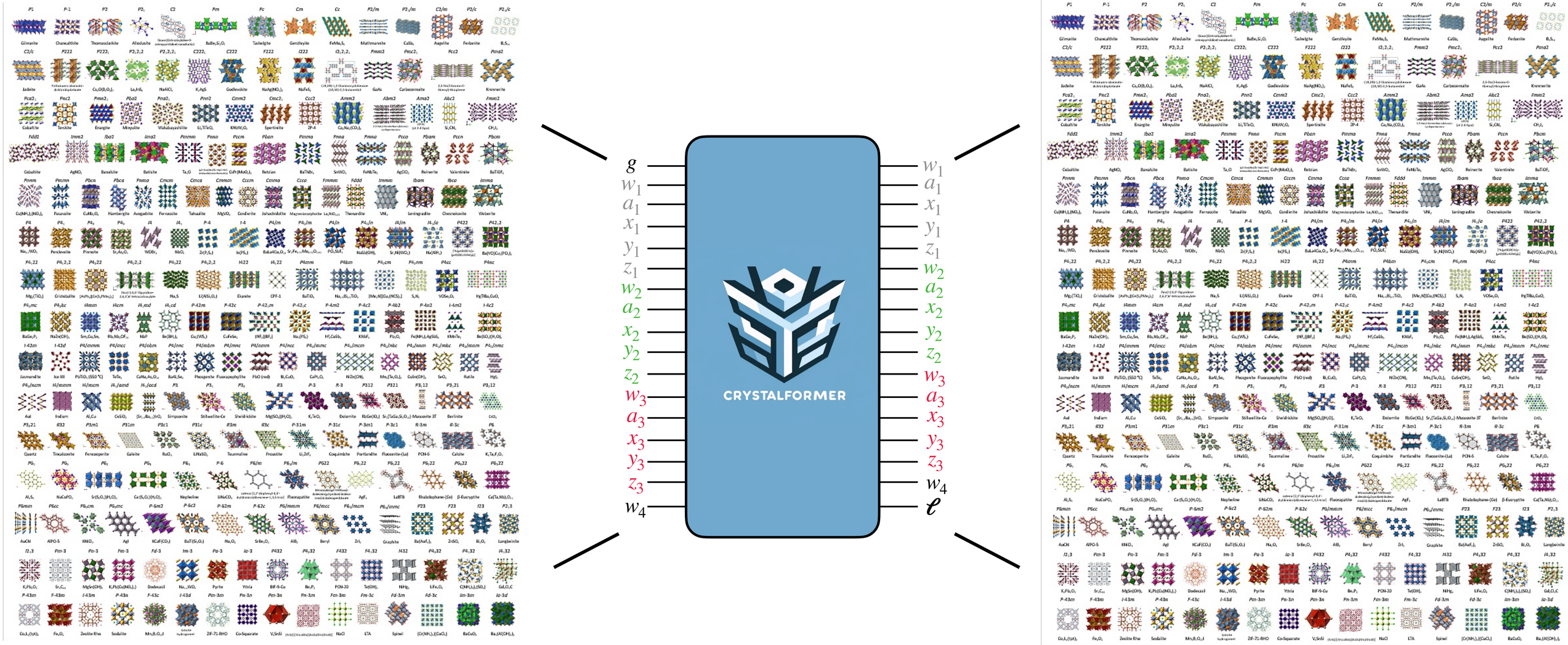
“语法”~ 固体化学规律

“同义词” ~ 可以互换的元素　　“格律”~ 空间群Wyckoff占位

“成语”~ 配位多面体

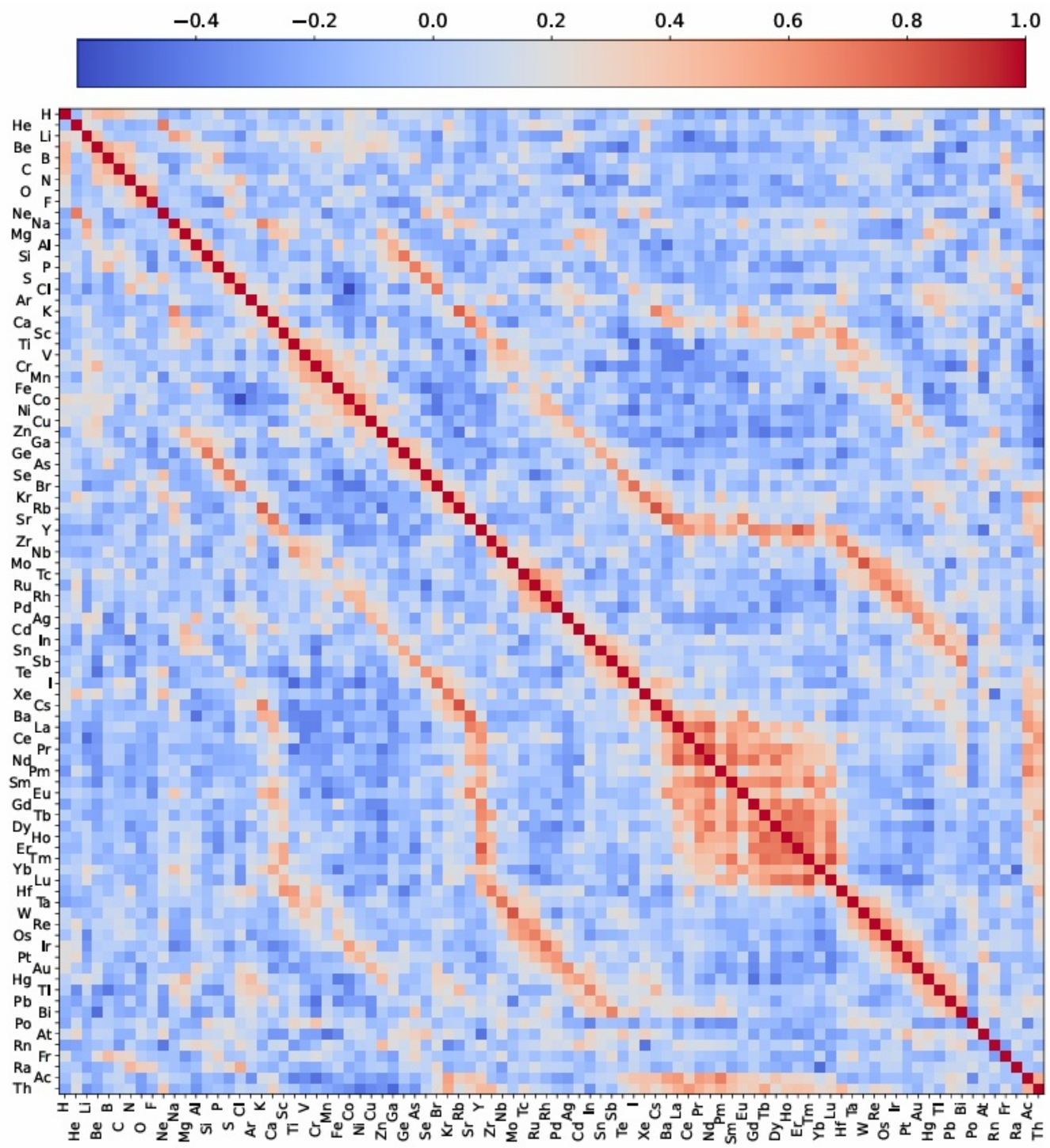


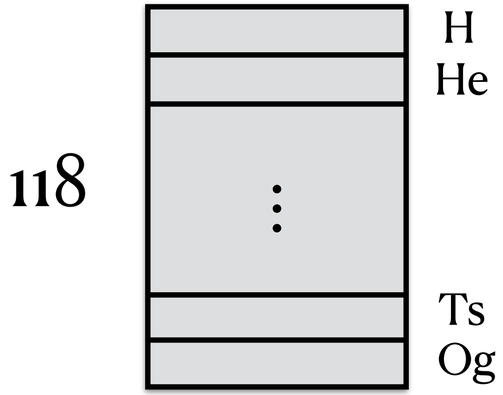
Not a large language model, **nor** a potential energy surface



Compress material database into transformer parameters

The model has to gain chemical intuition for such compression





Element embedding table

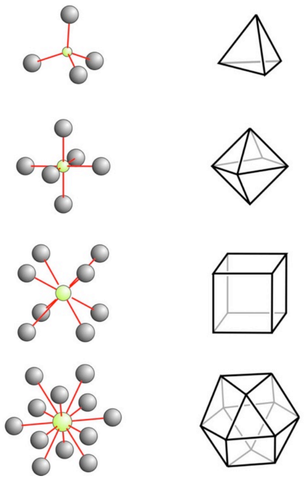
$$\frac{\boldsymbol{a} \cdot \boldsymbol{b}}{|\boldsymbol{a}| \cdot |\boldsymbol{b}|}$$

Cosine similarity

Solid state chemistry as “n-gram” in the crystal language

$$g\text{-}W_1\text{-}A_1\text{-}X_1\text{-}W_2\text{-}A_2\text{-}X_2\text{-}\cdots\text{-a-b-c-}\alpha\text{-}\beta\text{-}\gamma$$

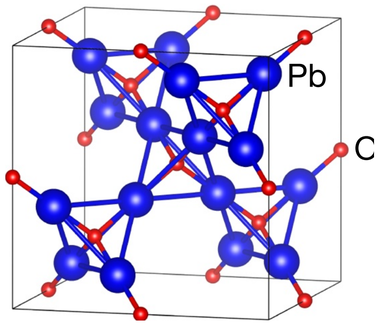
Coordination polyhedra



Polyhedra in Chemistry
Gongdu Zhou 2009

Lattices

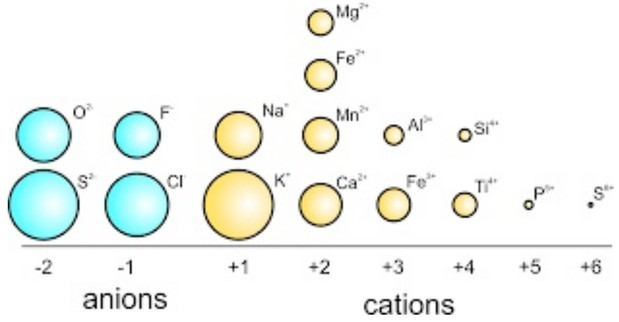
SG 227 $Pb_2Sb_2O_7$
Pyrochlore



Regnault et al, catalogue of
flat-band materials, Nature ‘22

Valence

“anions are in less symmetric positions than cations”



Urusov and Nadezhina,
J. Struct. Chem. 2009

Crystals by intuition vs by minimization

Data-driven (system 1)

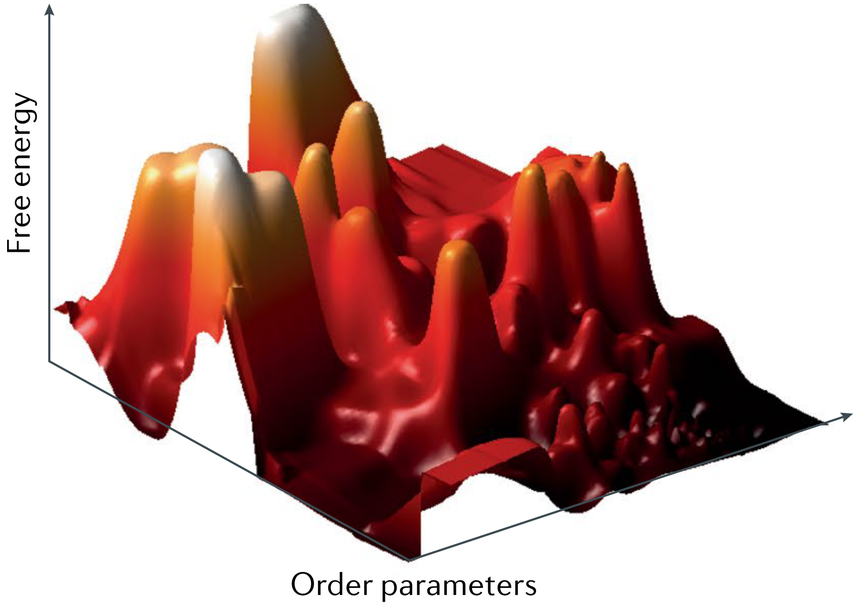
Chemical intuitions (e.g. Pauling rules)

from compression

CDVAE, Mattergen, Unimat, DiffCSP, CrystaLLM...

Physics-based (system 2)

energy minimization

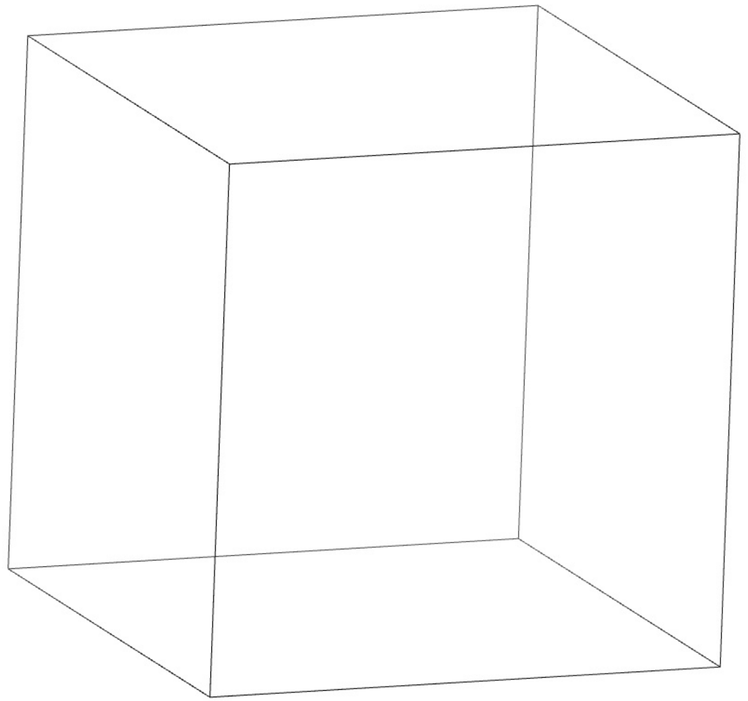


CALYPSO, USPEX, ARISS,...

DPA, MACE, LASP, GNoME,...

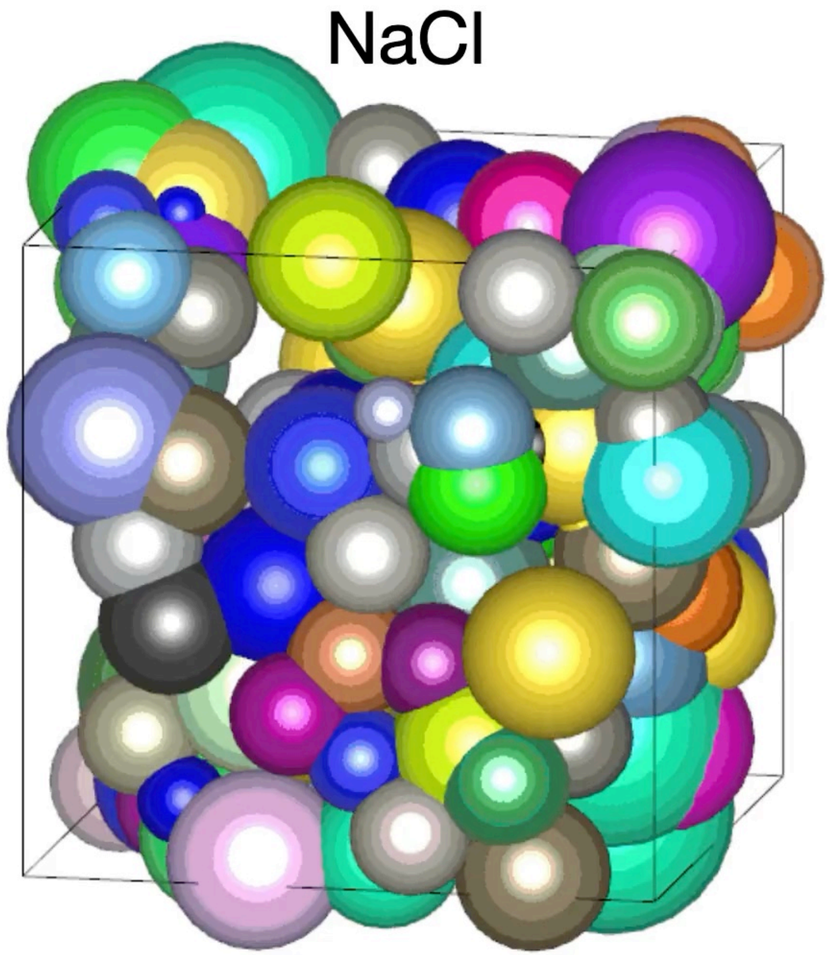
Autoregressive sampling of a crystal

$Cs_2ZnFe(CN)_6$



225-a-Fe-0-0-0-b-Zn-1/2-1/2-1/2-c-Cs-1/4-1/4-1/4-e-C-0.18-0-0-e-N-0.29-0-0-X-10.45-10.45-10.45-90-90-90

Generating crystals via diffusion



Google DeepMind

Yang et al, 2311.09235

https://unified-materials.github.io/unimat/

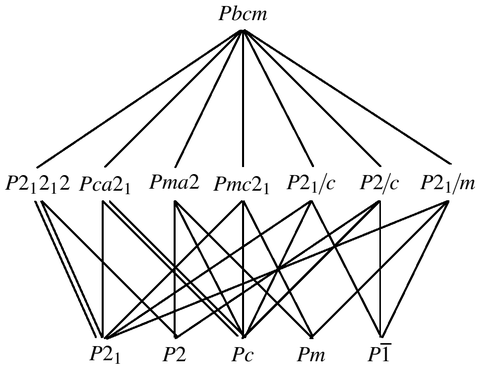
See also Xie et al, 2110.06197, Zheng et al, 2306.05445, Jiao et al, 2309.04475, Zeni et al, 2312.03687, Jiao et al,2402.03992 ...

The *large* language model approach

```
data_Na1Cl1
_symmetry_space_group_name_H-M   'P1'
_cell_length_a   3.9893
_cell_length_b   3.9893
_cell_length_c   3.9893
_cell_angle_alpha   60.0000
_cell_angle_beta   60.0000
_cell_angle_gamma   60.0000
_symmetry_Int_Tables_number   1
_chemical_formula_structural   NaCl
_chemical_formula_sum   'Na1 Cl1'
_cell_volume   44.8931
_cell_formula_units_Z   1
loop_
 _symmetry_equiv_pos_site_id
 _symmetry_equiv_pos_as_xyz
  1  'x, y, z'
loop_
 _atom_site_type_symbol
 _atom_site_label
 _atom_site_symmetry_multiplicity
 _atom_site_fract_x
 _atom_site_fract_y
 _atom_site_fract_z
 _atom_site_occupancy
  Cl  Cl0  1  0.0000  0.0000  0.0000  1
  Na Na1 1  0.5000  0.5000  0.5000  1
```

Here, one has to learn the periodicity and the space group—Wyckoff position—multiplicities—lattice relation from data as statistical correlations

ICSD



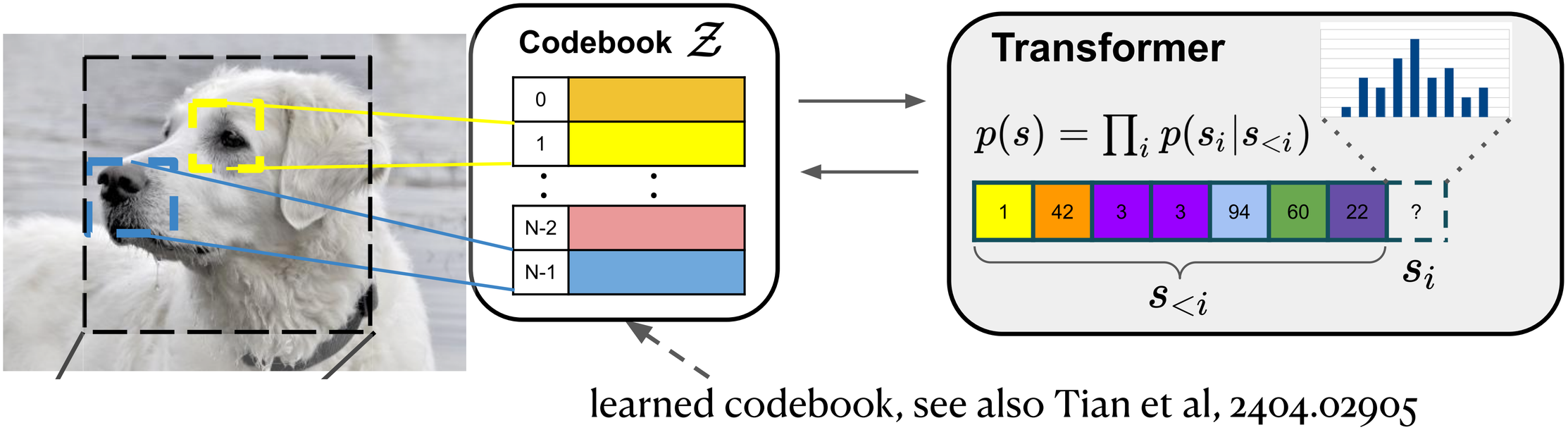
Cherrish your data, cherrish math

Flam-Shepherd et al, 2305.05708, Antunes et al, 2307.04340, Gruver, et al, 2402.04379

Meta

Aside: autoregressive transformer for images

Esser et al, Taming Transformers for High-Resolution Image Synthesis (VQGAN), 2012.09841



learned codebook, see also Tian et al, 2404.02905

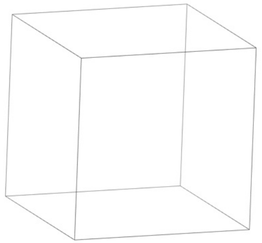
CrystalFormer leverages Nature's codebook: the Wyckoff position table

Crystal discovery and design with *CrystalFormer*

Zhendong Cao et al, 2403.15734

① 结构预测：晶体“造句”

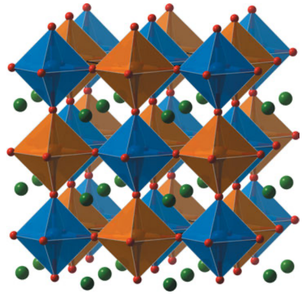
225-a-Fe-0-0-0-b-Zn-1/2-1/2-1/2-c-Cs-1/4-1/4-1/4-e-C-0.18-0-0-e-N-0.29-0-0



② 元素替换：晶体“完形填空”

More double perovskites $A_2BB'O$?

225-a-[?]-0-0-0-b-[?]-1/2-1/2-1/2-c-[?]-1/4-1/4-1/4-e-O-[?]-0-0

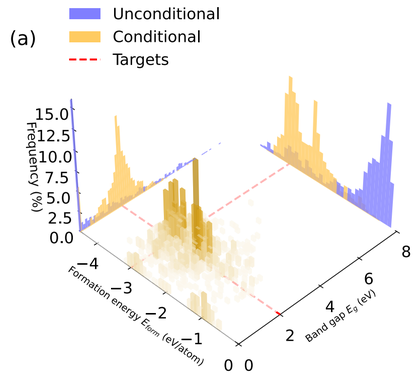


③ 反向设计：晶体“命题作文”

CrystalFormer for $p(\boldsymbol{X})$

Property prediction models $p(y \mid \boldsymbol{X})$

Posterior $p(\boldsymbol{X} \mid y) \propto p(\boldsymbol{X}) p(y \mid \boldsymbol{X})$



[–] The potato, or potato chip, is one of the best-selling snacks in the world!\n \n It comes in a variety of colors, is gluten-free (except for gluten-free chips), low in fat and saturated fat, and can be made in just 30 minutes, using only 6 simple ingredients. It's also easy to make, and can be stored in its own tin. \n \n The potato chip, however, may not be for everyone. If you are trying to lose weight, it...

[Science; 0.005] The potato is an ancient member of the green family which, as its name suggests, is green and edible. It is native to the Middle East and parts of South Asia. It is an important crop in many parts of the world. The potato, a member of the potato family, has been domesticated for thousands of years. It can be eaten raw and cooked in its skins; it is also used as a starch and is a great source of energy and fiber in many diets and in many countries...

[Science; 0.01] The potato was once thought to have no health problems and has been promoted as a nutritious food source since the mid-1800s, but recent reports indicate that it has many harmful health issues. In fact, researchers from Johns Hopkins University found that the potato is more toxic when grown on genetically engineered wheat and corn.\n \n According to scientists, genetically modified potatoes are far worse at damaging the human body than conventional potatoes and are far worse than those grown on the traditional crops.\n \n The study also revealed...

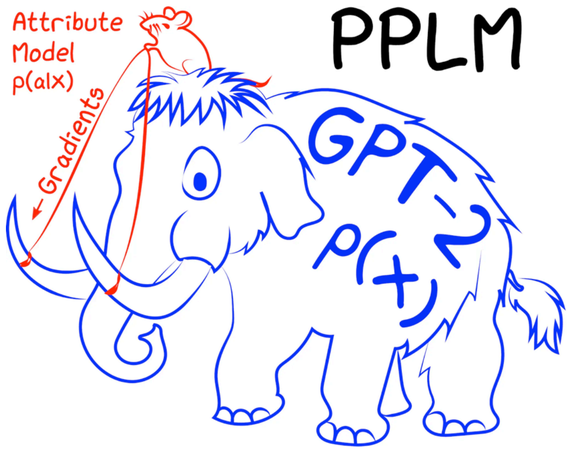
[Science; 0.02] The potato plant, a member of the onion family.\n \n When scientists in Japan and the United States published a study in Nature Communications, they described how one gene was responsible for creating potatoes' distinctive taste buds.\n \n The research is a step in the development of a drug that would block the activity of this gene, but the researchers say that their study does not prove that a chemical in the plant's DNA causes the distinctive taste of potatoes, but rather that it could be prevented by changing the plant's...

$$\text{texts} \sim p(\text{texts} \mid \text{topic})$$

Controlled text generatiion

Dathathri et al, 1912.02164

https://www.uber.com/en-KR/blog/pplm/



[Science; 0.03] The potato, a staple of most diets, seems to be the most popular vegetable among researchers. The research has been published and peer-reviewed.\n \n The potato has a unique ability. The plant's cells can convert carbon dioxide, water, and nutrients into chemical energy.\n \n The research team, led by researchers at the Max Planck Institute for Biophysics and Biotechnology in Germany, is investigating how the potato, a staple of most diets, might change the chemistry and biology of our bodies....
[Science; 0.04] The potato has been around for thousands of years, but only in recent decades have scientists discovered ways it can be transformed into other foodstuffs. Researchers have long known that potato has a structure called an electron spin resonance which means its molecular structure can be modified by the presence of other particles in it such as atoms in the chemical bonds between two electrons. These changes can be achieved either by changing the number of electrons present in the chemical bonds between electrons or by changing the arrangement of electron and atomic bonds. In both...
[Science; 0.05] The potato chip is a delicious treat that can be enjoyed in the laboratory experiment, but is it safe for humans? \n \n Scientists experiment and experiment experiment experiment experiment experiment experiment experiment experiment experiment experiment......
[Science; 0.1] The potato, which scientists at the lab experiment ...

make it smarter







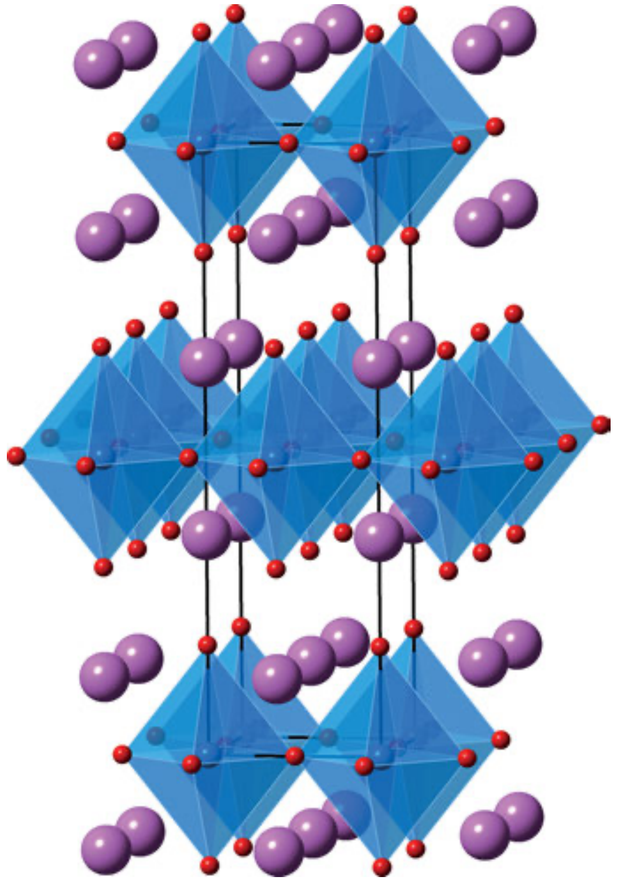
You

make it even more smarter

ChatGPT

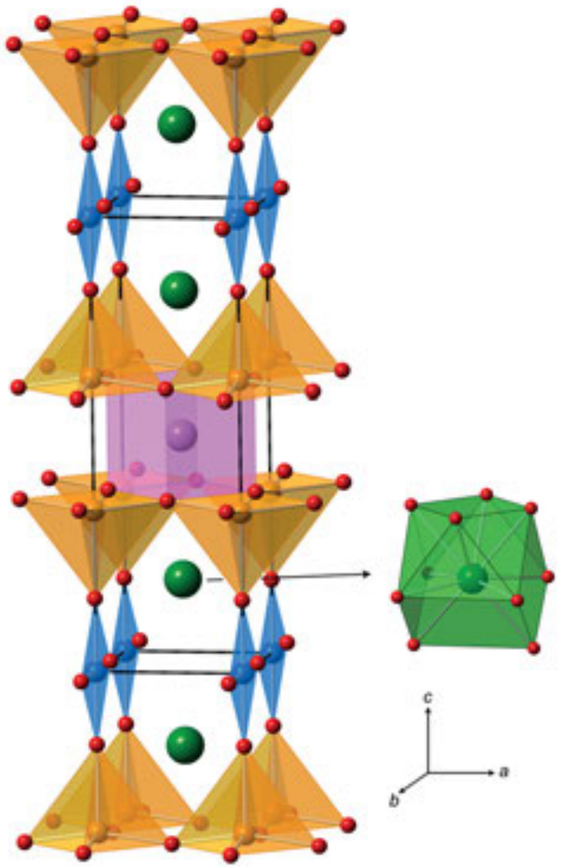
We've reached a point where the depiction of the detective's intelligence is already at a transcendent, almost mythical level, merging with the universe's knowledge and represented as an ethereal entity of energy and information. Going beyond this might enter into realms of abstract concepts that are challenging to visualize concretely. However, if you have a specific idea or theme in mind that you believe would symbolize an even greater level of intelligence, I'd be happy to hear it and create an image based on your vision!

make T_c higher



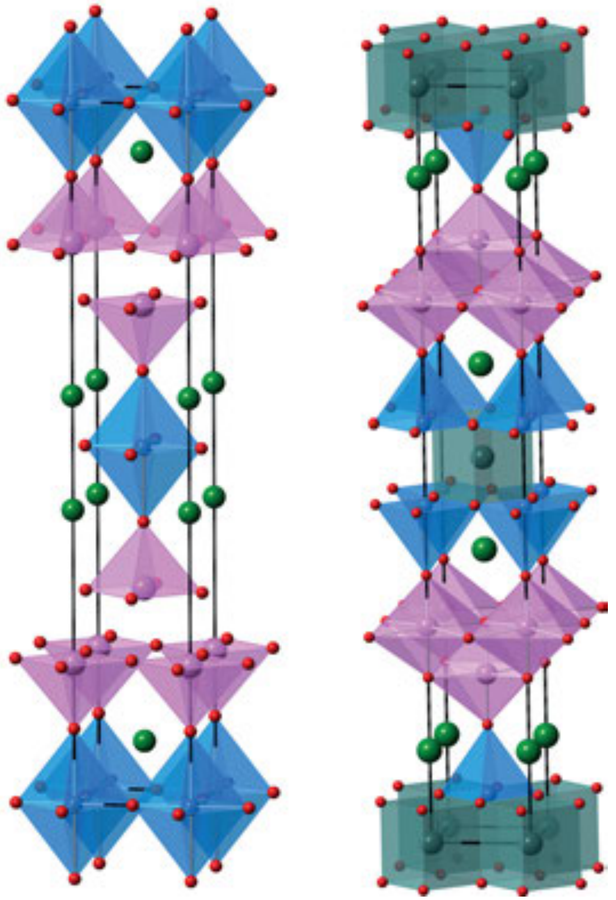
$La_{1.85}Ba_{0.15}CuO_4$

35K



$YBa_2Cu_3O_7$

92K



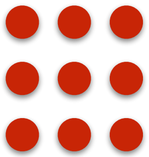
$Tl_2Ba_2Ca_nCu_{n+1}O_{2n+6}$

125K

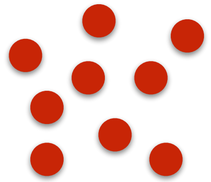
Nature tries to minimize free energy

$$F = E - TS$$

energy



entropy



F is a cost function given by Nature

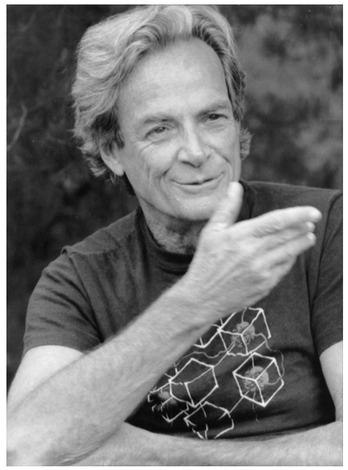
The *same* cost function for training deep generative models, almost

The variational free energy principle

Gibbs–Bogolyubov-Feynman-Delbrück–Molière

$$\min \quad F[\rho] = \mathrm{Tr}(H\rho) + k_B T \,\mathrm{Tr}(\rho \ln \rho) \quad \geq F$$

↓ variational density matrix ↓ energy ↓ entropy 😱



Difficulties in Applying the Variational Principle to Quantum Field Theories[1]

Richard P. Feynman

ρ ?

Generative models !

[1] transcript of Professor Feynman's talk in 1987

Neural *canonical* transformations

Li, Dong, Zhang, LW, PRX '20

Xie, Zhang, LW, JML '21

classical world

Symplectic transformation

Probability density

$$p$$

Kullback-Leibler divergence

$$\mathbb{KL}\left(p\,||\,q\right)$$

quantum world

Unitary transformation

Density matrix

$$\rho$$

Quantum relative entropy

$$S\left(\rho\,||\,\frac{e^{-\beta H}}{Z}\right)$$

Neural canonical transformation for variational density matrix

1809.10606, PRL '19

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

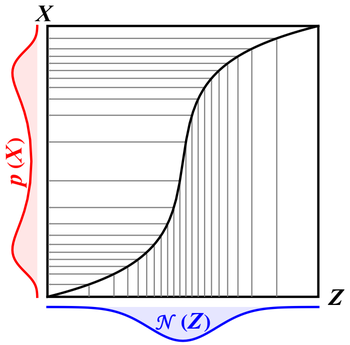
2105.08644, JML '22
2201.03156, SciPost Physics'23

Classical probability p_n

Quantum states $|\psi_n\rangle = U|\phi_n\rangle$

"... the murderer is ___"

$p(\_|\dots)$

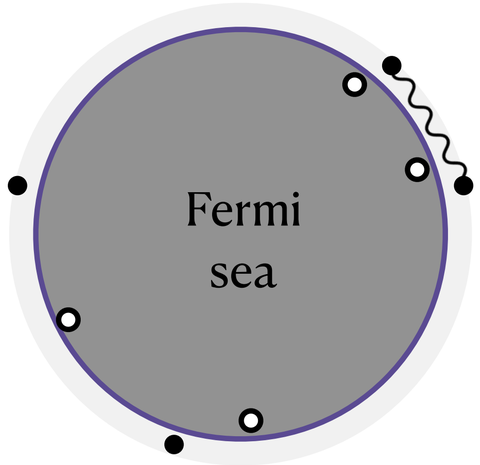


autoregressive model

$\sqrt{\text{flow}}$

Example: the variational density matrix of electron gas

Xie, Zhang, LW, SciPost Physics '23



Low-energy excited states are labeled in the same way as the ideal Fermi gas

$\boldsymbol{K} = \{\boldsymbol{k}_1, \boldsymbol{k}_2, \ldots, \boldsymbol{k}_N\}$

$$\rho = \sum_{\boldsymbol{K}} p(\boldsymbol{K}) \left| \Psi_{\boldsymbol{K}} \right\rangle \left\langle \Psi_{\boldsymbol{K}} \right|$$

Normalized probability distribution

Orthonormal many-electron basis

1. $\sum_{\boldsymbol{K}} p(\boldsymbol{K}) = 1$

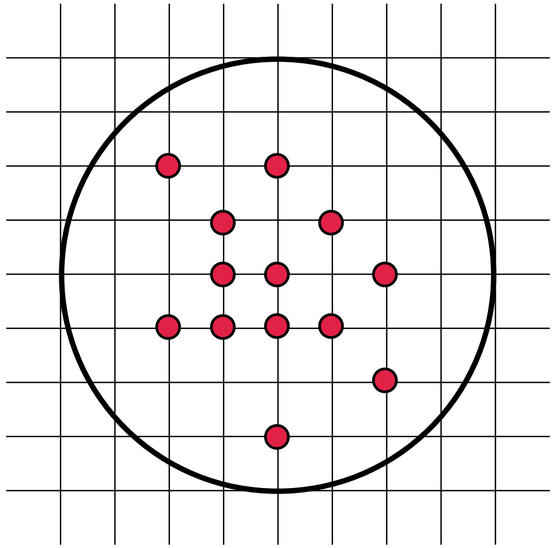
2. $\langle \Psi_{\boldsymbol{K}} | \Psi_{\boldsymbol{K}'} \rangle = \delta_{\boldsymbol{K},\boldsymbol{K}'}$

There will also be interesting twists for physics considerations

Variational autoregressive network for $p(\boldsymbol{K})$

Fermionic occupation in k-space

$$p(\boldsymbol{K}) = p(\boldsymbol{k}_1)p(\boldsymbol{k}_2\,|\,\boldsymbol{k}_1)p(\boldsymbol{k}_3\,|\,\boldsymbol{k}_1,\boldsymbol{k}_2)\cdots$$



$\binom{M}{N}$ probability space

N	# of fermions	# of words
M	Momentum cutoff	Vocabulary

Pauli exclusion: we are modeling a *set of words* with no repetitions and no order

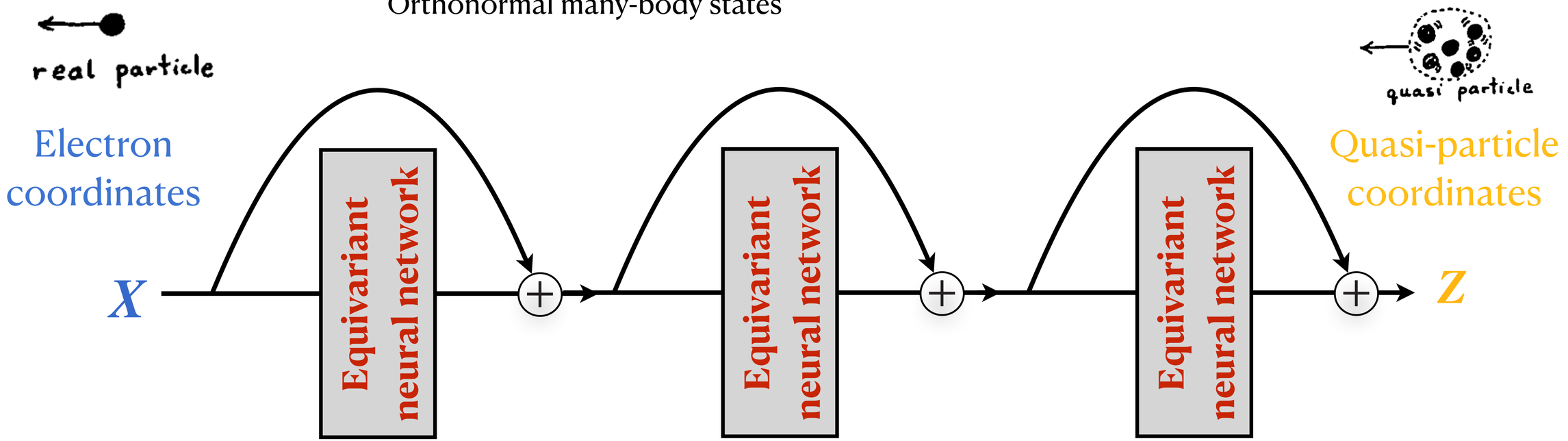
We use masked casual self-attention Vaswani et al 1706.03762; Alternative solution: Hibat-Allah et al, 2002.02793, Barrett et al, 2109.12606

$\sqrt{\text{flow}}$ for $|\Psi_K\rangle$

$$\Psi_K(X) = \frac{\det(e^{ik_i \cdot z_j})}{\sqrt{N!}} \cdot \left| \det\left(\frac{\partial Z}{\partial X}\right) \right|^{\frac{1}{2}}$$

Xie, Zhang, LW, SciPost '23

Orthonormal many-body states



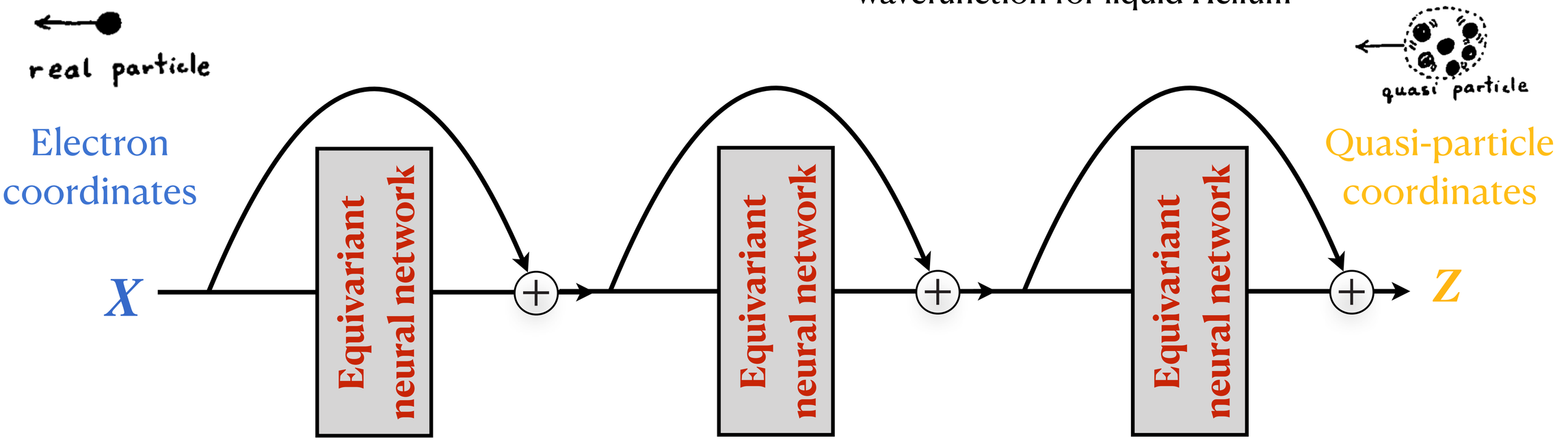
$X \leftrightarrow Z$: unitary backflow between particle and quasiparticle coordinates

Fermion statistics: permutation equivariant flow We use FermiNet layer Pfau et al, 1909.02487

Feynman's backflow in the deep learning era

$$z_i = x_i + \sum_{j \neq i} \eta(|x_i - x_j|)(x_j - x_i)$$

Feynman & Cohen 1956
wavefunction for liquid Helium



Iterative backflow → deep residual network → continuous normalizing flow

Taddei et al, PRB '15 E Commun. Math. Stat 17', Harbor el al 1705.03341, Lu et al 1710.10121, Chen et al, 1806.07366

ANNALS OF PHYSICS: **24,** 63–88 (1963)

Point Transformations and the Many Body Problem*

M. EGER† AND E. P. GROSS

Brandeis University, Waltham, Massachusetts

An investigation is made of possible uses of many dimensional coordinate transformations in the quantum many-body problem. The transformed Hamiltonian is quadratic in the momenta with a space dependent metric. The original potential energy undergoes alteration and an additional "metric" potential energy appears. A relatively complete analysis of the transformed original potential is made, and the coordinate transformation can be used to suppress undesirable features of the original potential. For bosons one can attempt to directly map a complete set of noninteracting states onto approximate eigenstates of the system with interactions. Contact is made with a theory of weakly interacting bosons. In the general case it emerges that a given transformation uniquely fixes all the spatial correlation functions, which can be explicitly computed. The extended point transform can then be used as a link between diverse experimental quantities. The full use of the transformation to compute from first principles requires adequate approximations to the Jacobian and the inverse transform. These problems are not studied.

√flow

materializes this dream

The objective function of variational density matrix

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

$$F = \mathop{\mathbb{E}}_{n\sim p_n}\left[k_B T \ln p_n + \mathop{\mathbb{E}}_{\boldsymbol{R}\sim|\psi_n(\boldsymbol{R})|^2}\left[\frac{H\psi_n(\boldsymbol{R})}{\psi_n(\boldsymbol{R})}\right]\right]$$

Boltzmann distribution

Born probability

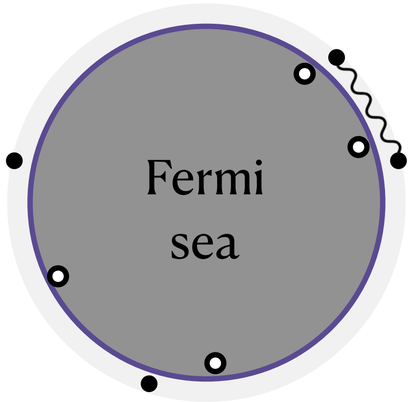
Jointly optimize p_n and $\psi_n(\boldsymbol{R})$ to minimize the variational free energy

The deep variational free energy approach

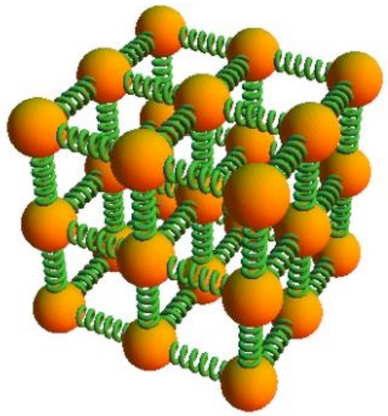
1802.02840, PRL '18 1910.00024, PRX '20 1912.11381, MLST' 21 2201.03156, SciPost '23
1809.10606, PRL '19 2209.06095, PRL '23 2105.08644, JML '22 2403.12518, JCP '24...

Low-temperature properties of interacting electrons
(~50 electrons)

Vibrational spectra of molecules and solids
(~500 atoms)

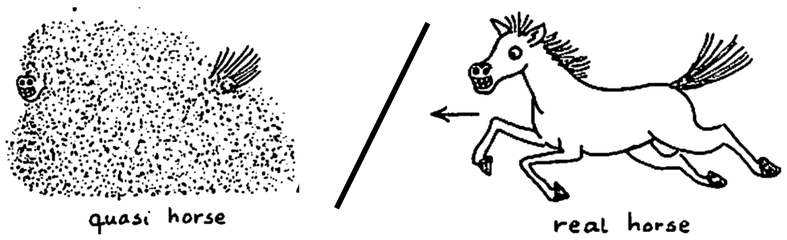


$$\rho = \sum_n U\,|\phi_n\rangle p_n \langle\phi_n|\,U^\dagger$$



A computational framework taking in account of electron correlation, thermal effect, and anharmonic lattices for free energy, entropy, and excitation spectra

Deep variational free energy approach: resolving puzzles



0
e-

Quansi-particle effective mass
contradicting experiments

$m^*/m > 1$

VOLUME 91, NUMBER 4 PHYSICAL REVIEW LETTERS week ending 25 JULY 2003

Spin-Independent Origin of the Strongly Enhanced Effective Mass in a Dilute 2D Electron System

$m^*/m < 1$

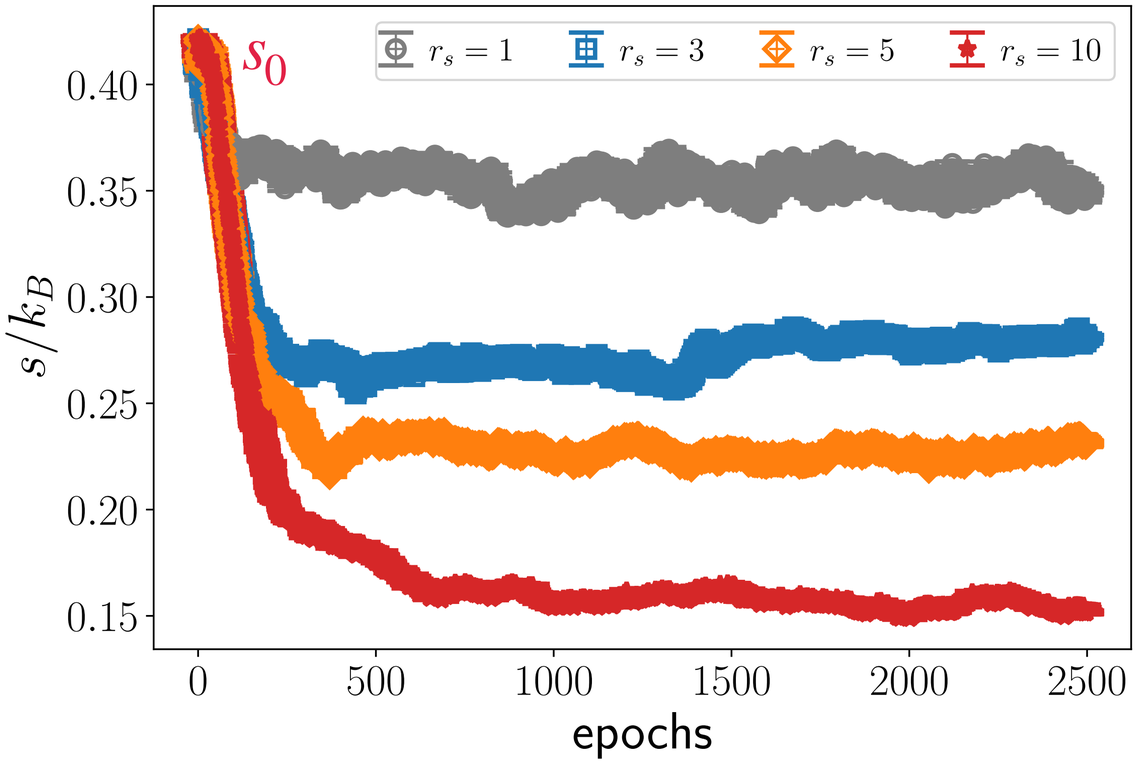
PRL 101, 026402 (2008) PHYSICAL REVIEW LETTERS week ending 11 JULY 2008

Effective Mass Suppression in Dilute, Spin-Polarized Two-Dimensional Electron Systems

Medini Padmanabhan, T. Gokmen, N. C. Bishop, and M. Shayegan
Department of Electrical Engineering, Princeton University, Princeton, New Jersey 08544, USA
(Received 19 September 2007; published 7 July 2008)

Hao Xie et al, SciPost Physics '23
Thermal entropy of 2D electron gas

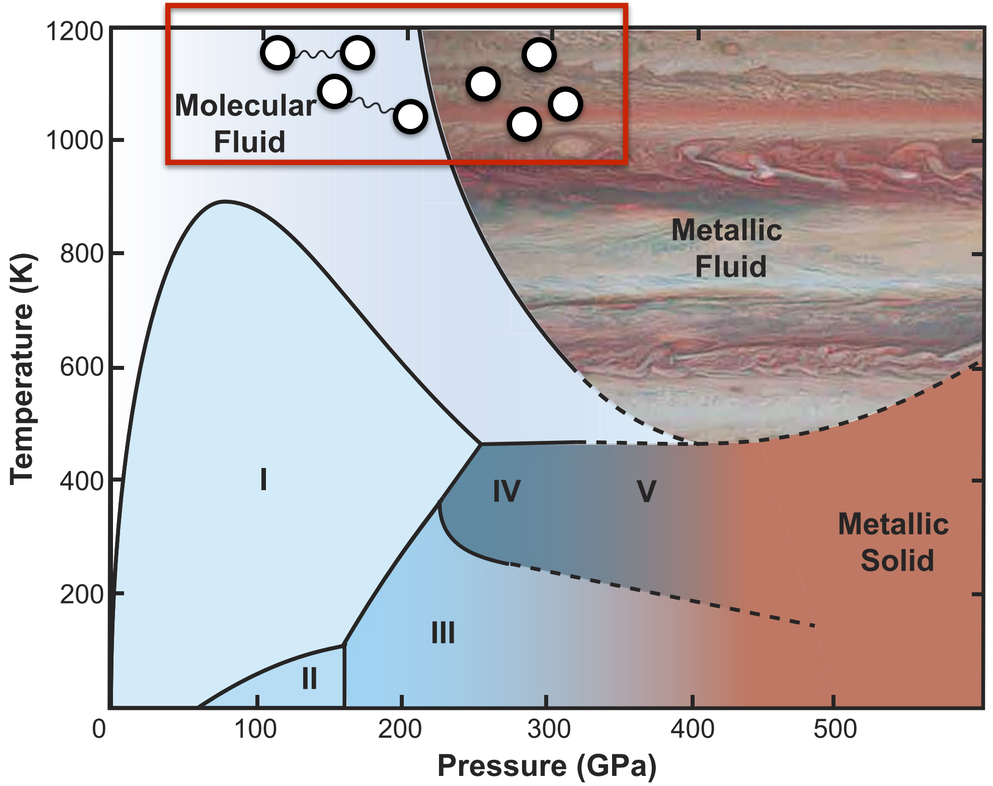
$$\frac{m^*}{m} = \frac{s}{s_0} < 1$$



Deep variational free energy approach: making discoveries

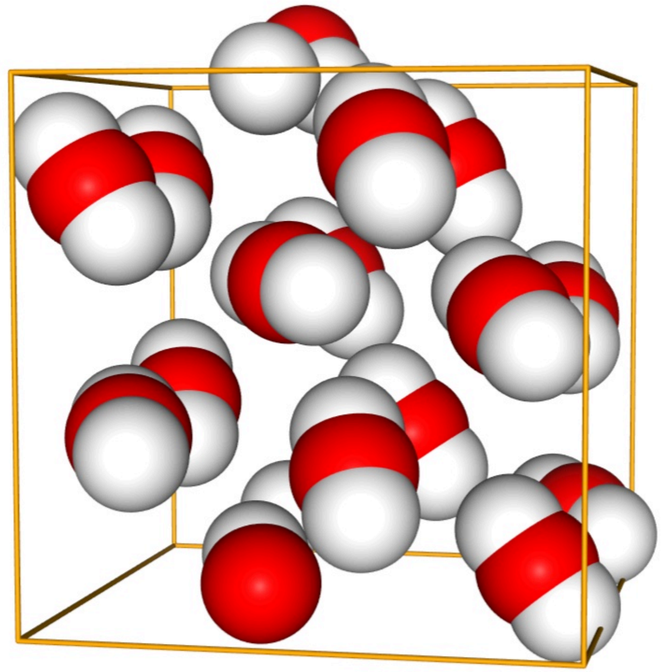
1
H

Knudson et al Science '15, Celliers et al Science '18
Mazzola et al Nat.Comm. '14, Pierleoni et al, PNAS '16
Cheng et al, Nature '20, Karasiev et al, Nature '21



Hydrogen phase diagram

Xinyang Dong et al, 2024



High-temperature solids of dense hydrogen

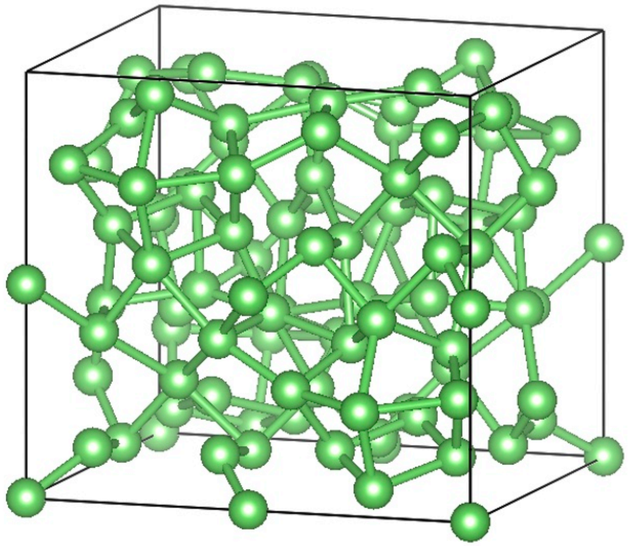
Gregoryanz et al, Matter Radiat. Extremes, 2020

see also Niu et al, PRL '23, Goswami et al, 2411.15665

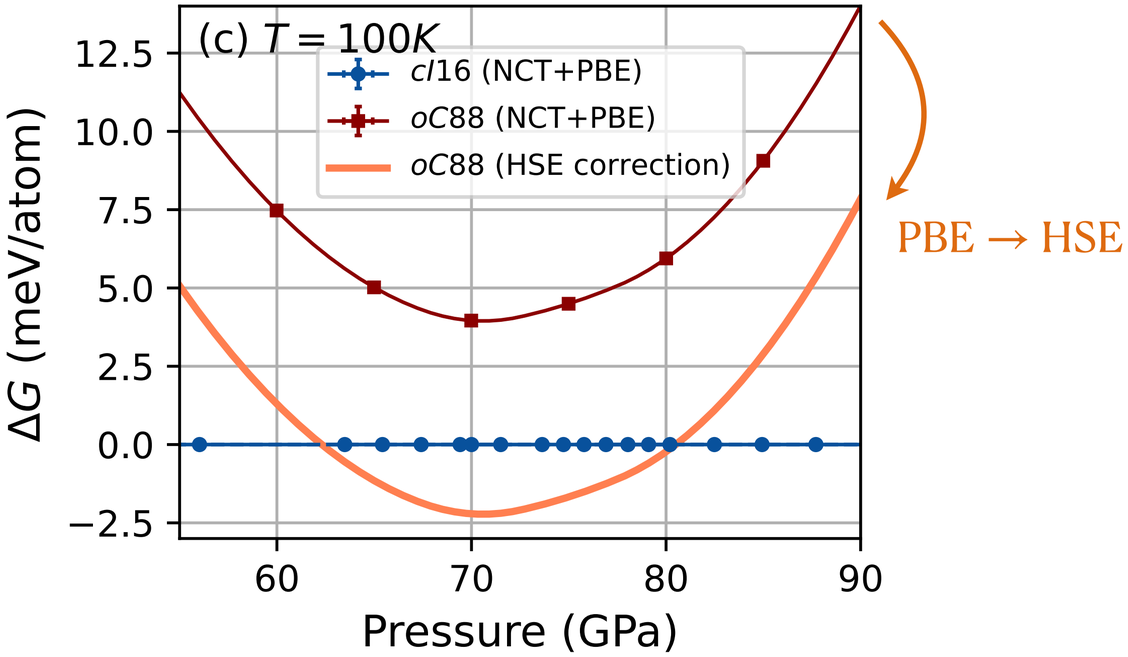
Deep variational free energy approach: clarifying mechanism

3
Li

Guillaume et al, Nature physics, '11
Marqués et al, PRL '11, Gorelli et al PRL '12



Qi Zhang et al, 2412.12451



Our calculation does **NOT** reproduce the experimentally observed Oc88 phase of Lithium, which contradicts consensus

- ☑ Thermal effect
- ☑ Quantum anharmonicity
- ☐ DFT functional error for bad metal

Generative AI for It

1. $$p(\boldsymbol{X} \mid y) \propto p(\boldsymbol{X}) p(y \mid \boldsymbol{X})$$

Matter inverse design

Exploiting intuitions in data

2. $$F[\rho] = E - TS$$

Nature's cost function

Variational free energy is finally practical

Turning physics problems into stochastic optimization

Leverages the deep learning engine

The Universe as a generative model

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

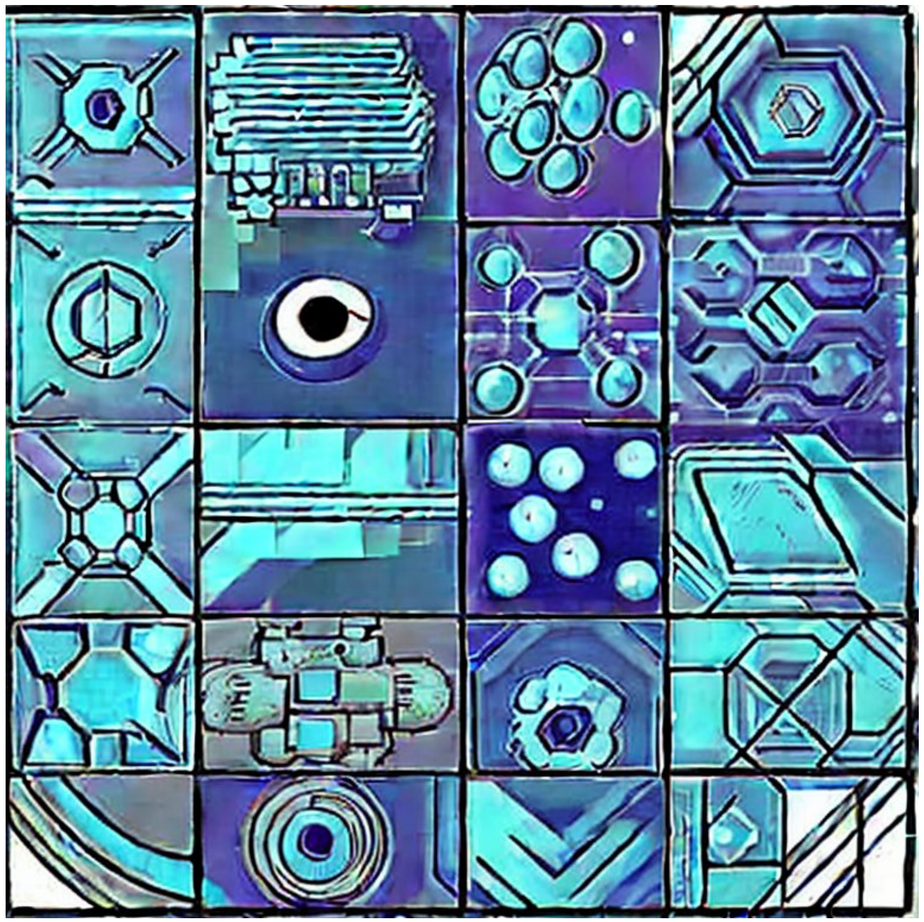
Thank you!

Discovering physical laws: learning the action

Solving physical problems: optimizing the action

A crash course offerred at IOP 2023 spring

2.23	Overview
3.2	Machine learning practices
3.9	A hitchhiker's guide to deep learning
3.16	Research projects hands-on
3.23	Symmetries in machine learning
3.30	Differentiable programming
4.6	Generative models-I
4.13	Generative models-II
4.20	Research projects presentation
4.27	AI for science: why now ?



Machine learning for physicists

https://github.com/wangleiphy/ml4p