

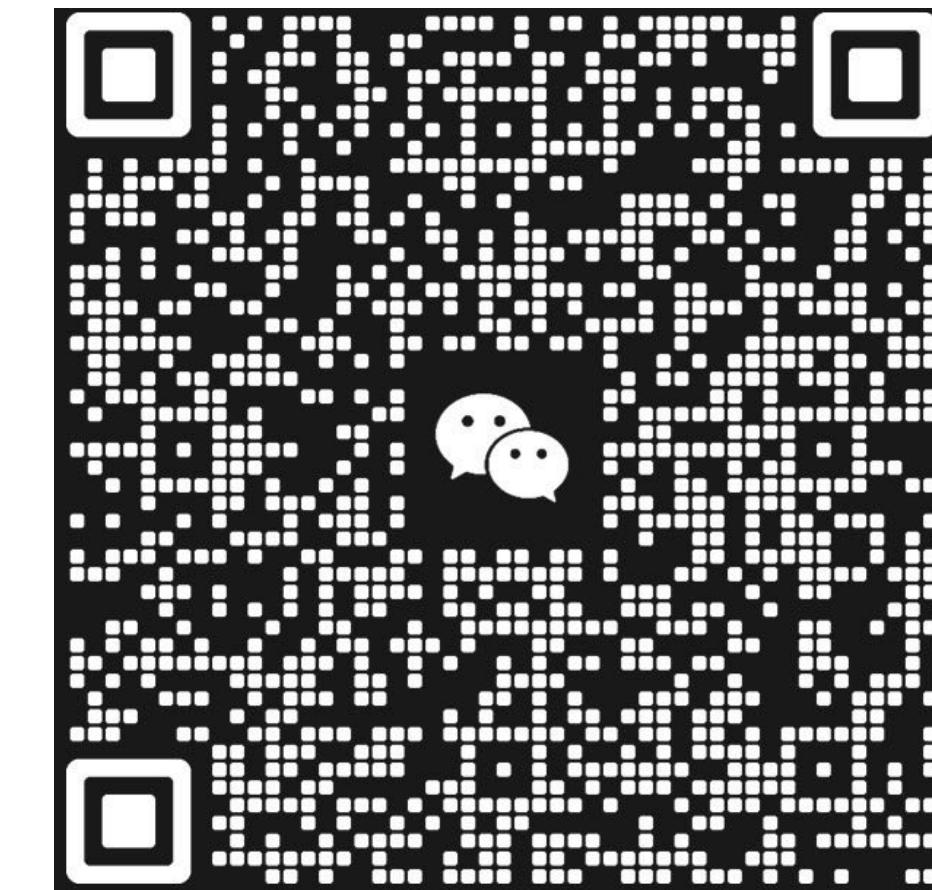
Machine learning for physicists

<https://github.com/wangleiphy/ml4p>

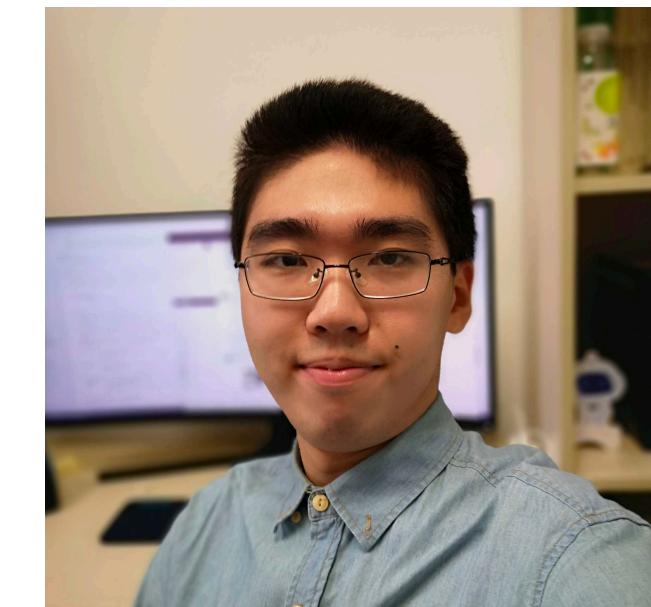
每周四上午10点

课程微信群

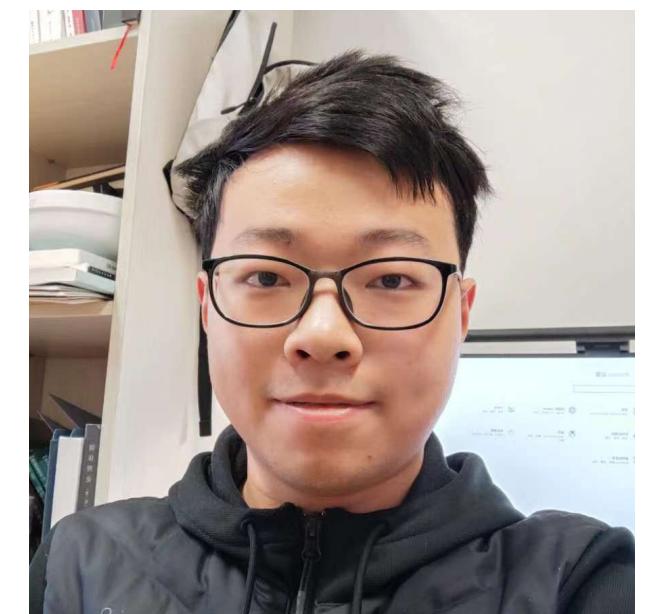
2.23	Overview
3.2	Machine learning practices
3.9	A hitchhiker's guide to deep learning
3.16	Research projects hands-on
3.23	Symmetries in machine learning
3.30	Differentiable programming
4.6	Generative models-I
4.13	Generative models-II
4.20	Research projects presentation
4.27	AI for science: why now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



助教



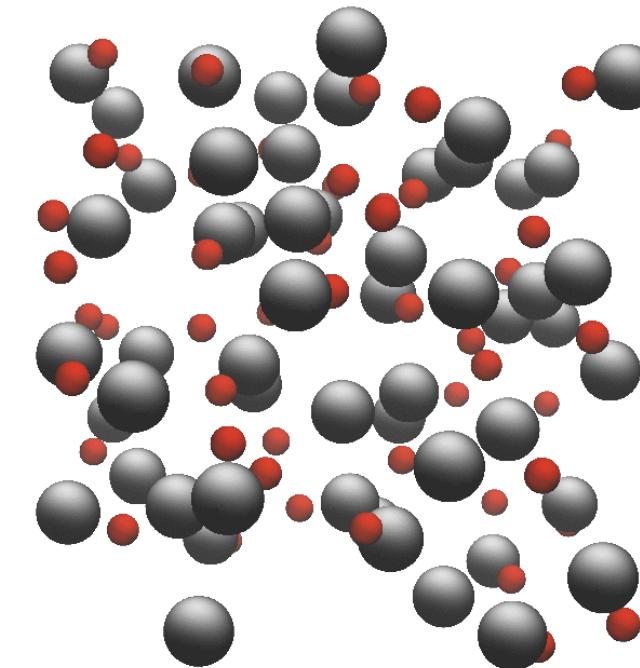
李子航



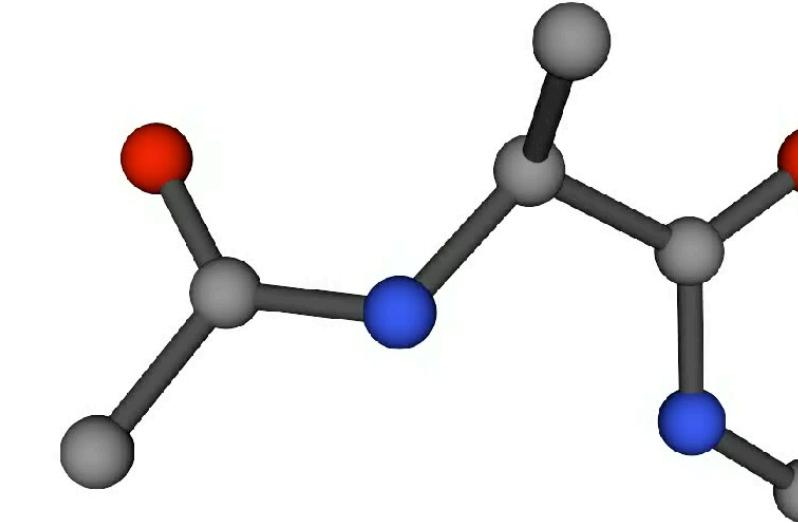
李扬帆

考核方式: project + presentation (1学分)

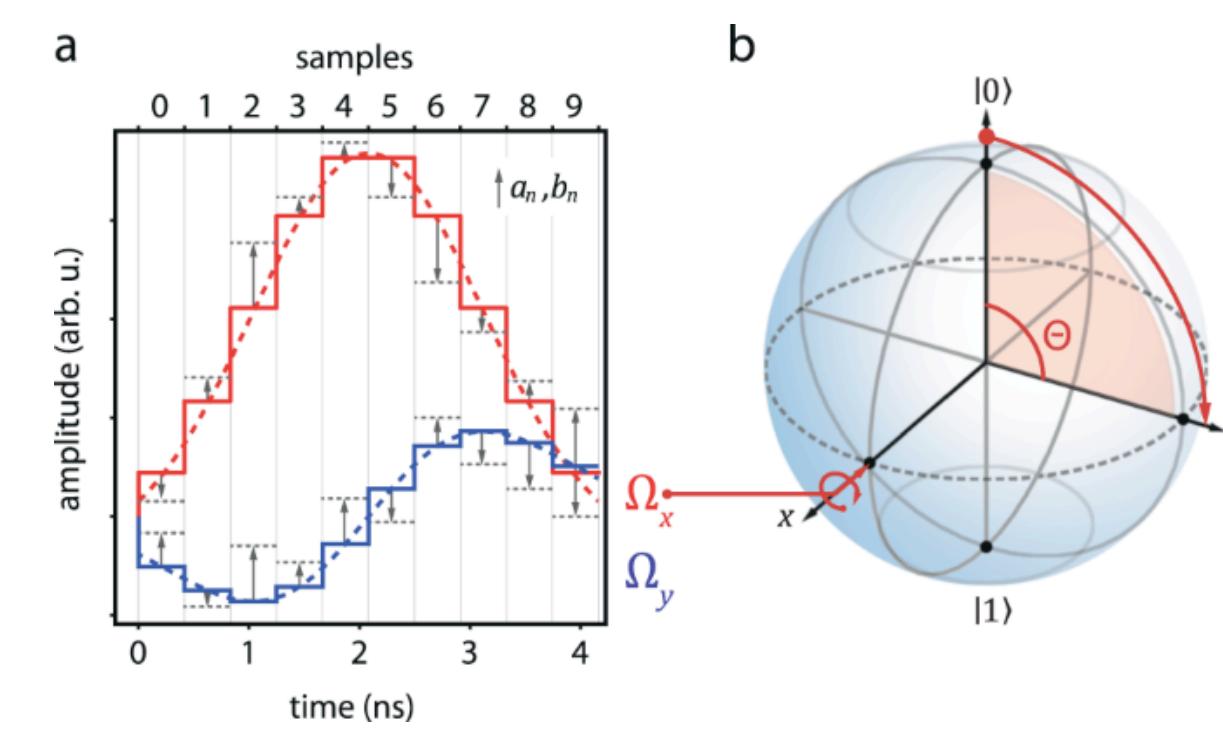
Research projects



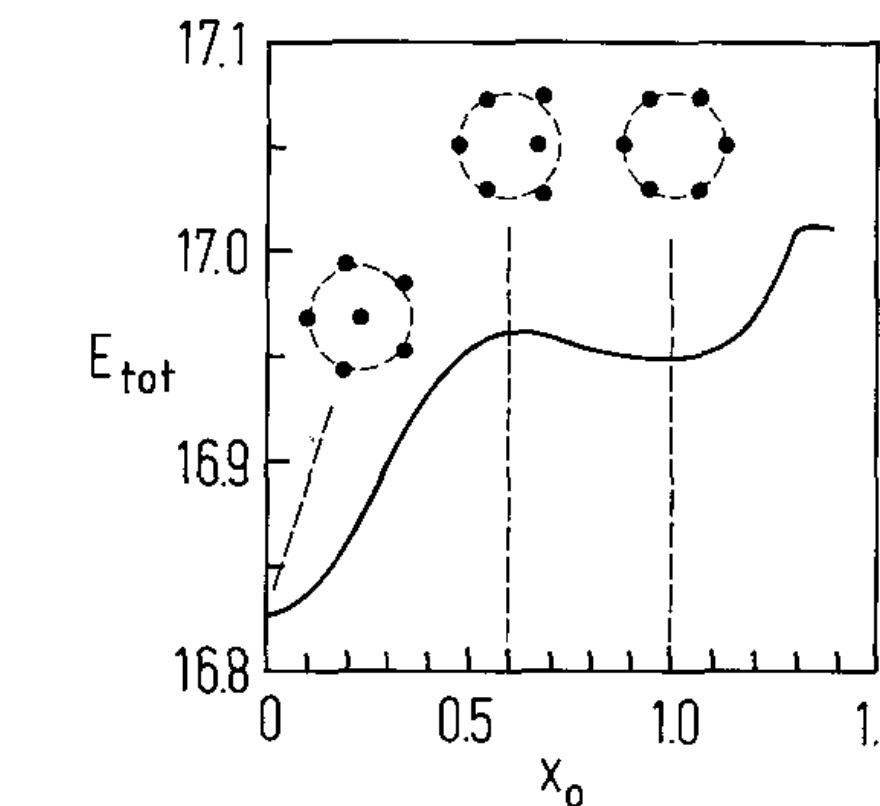
Learning energy of hydrogen atoms in a box



Sample alanine dipeptide conformations



Differentiable quantum optimal control



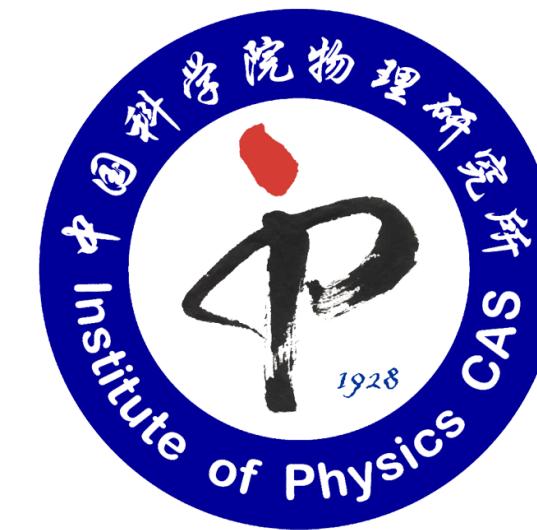
Solving a quantum dot problem: energy, free energy, ..

Or, bring your own...

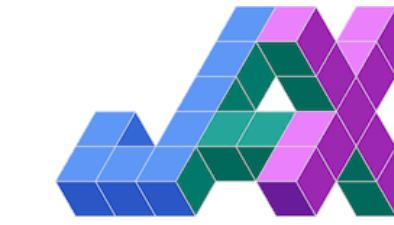
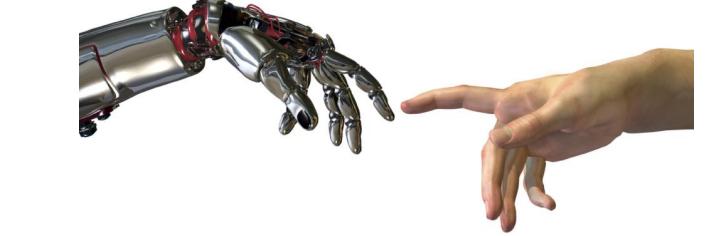
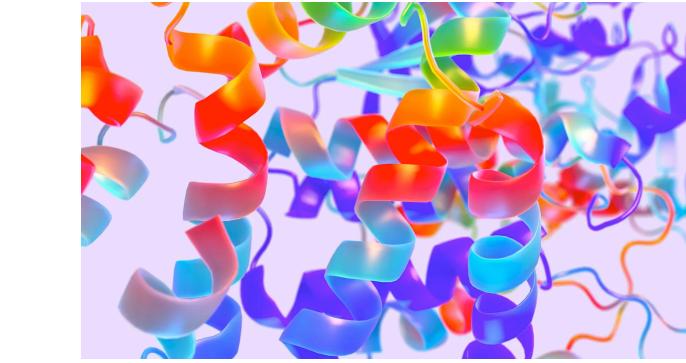
Scientific machine learning with and without data

Lei Wang (王磊)

Institute of Physics, CAS
<https://wangleiphy.github.io>

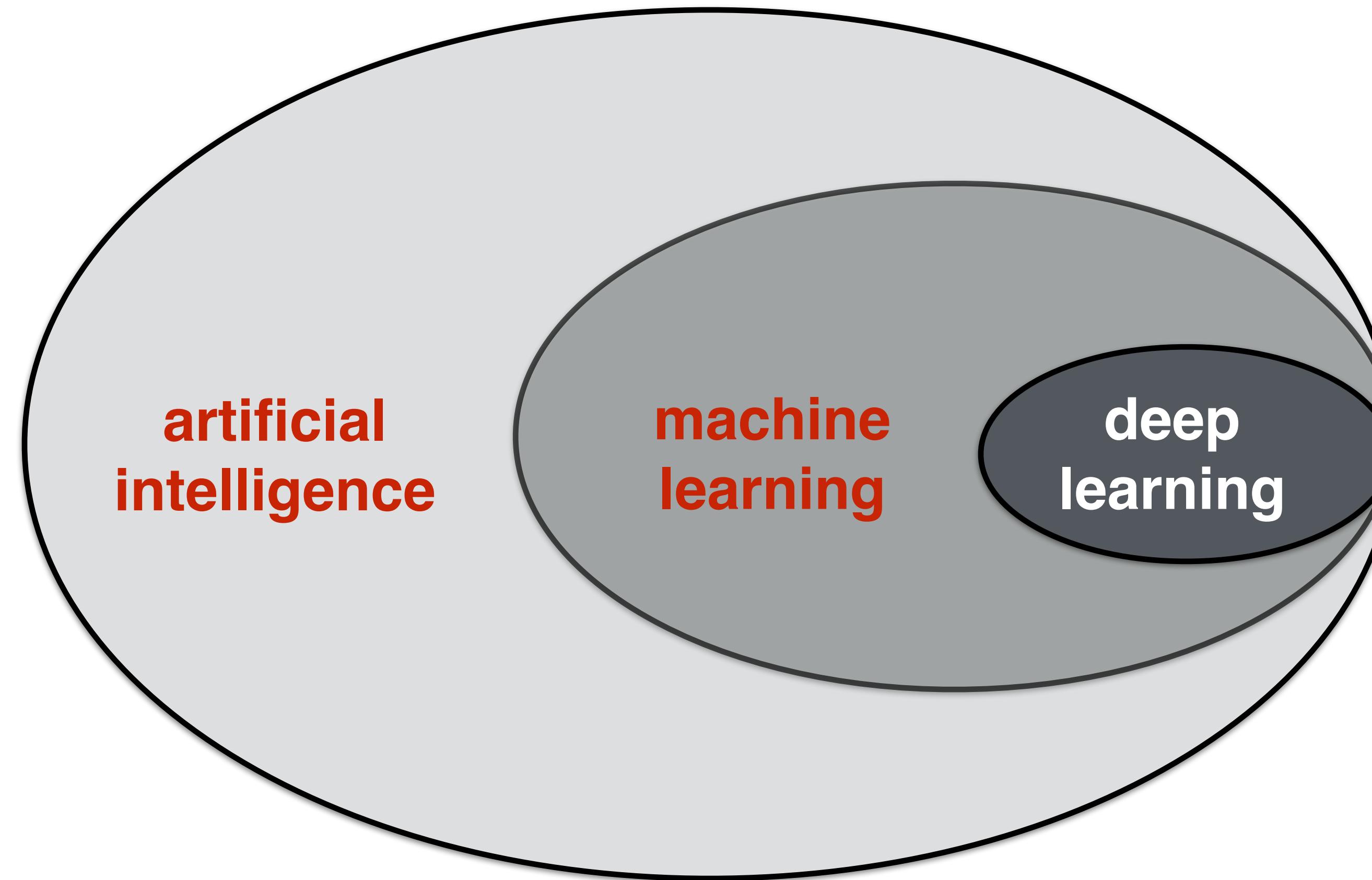


Why machine learning ?



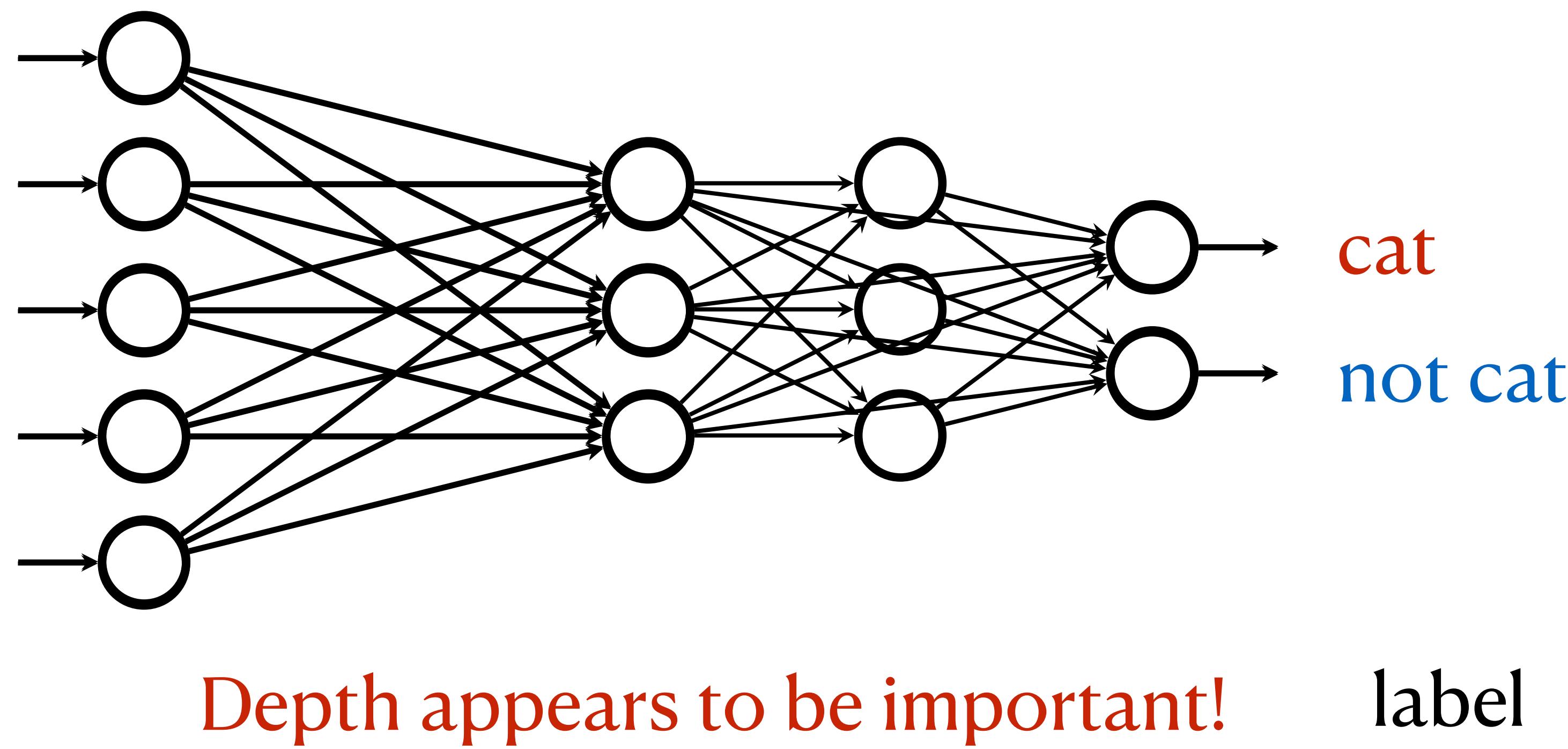
Game-changing technology for human life and scientific research

Why machine learning ?



Game-changing technology for human life and scientific research

Deep learning in 1 slide



Q: Why does deep learning work?

Universal Function Approximator

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

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

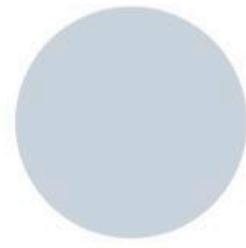
09:44 ↗

4G



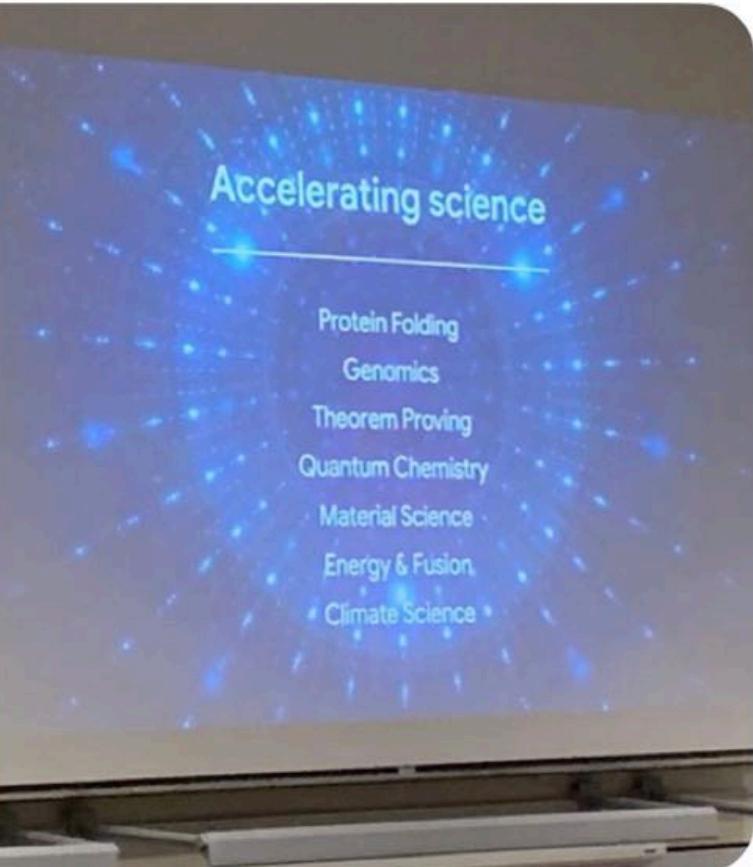
Tweet

♥ Kyle Cranmer and 6 others liked

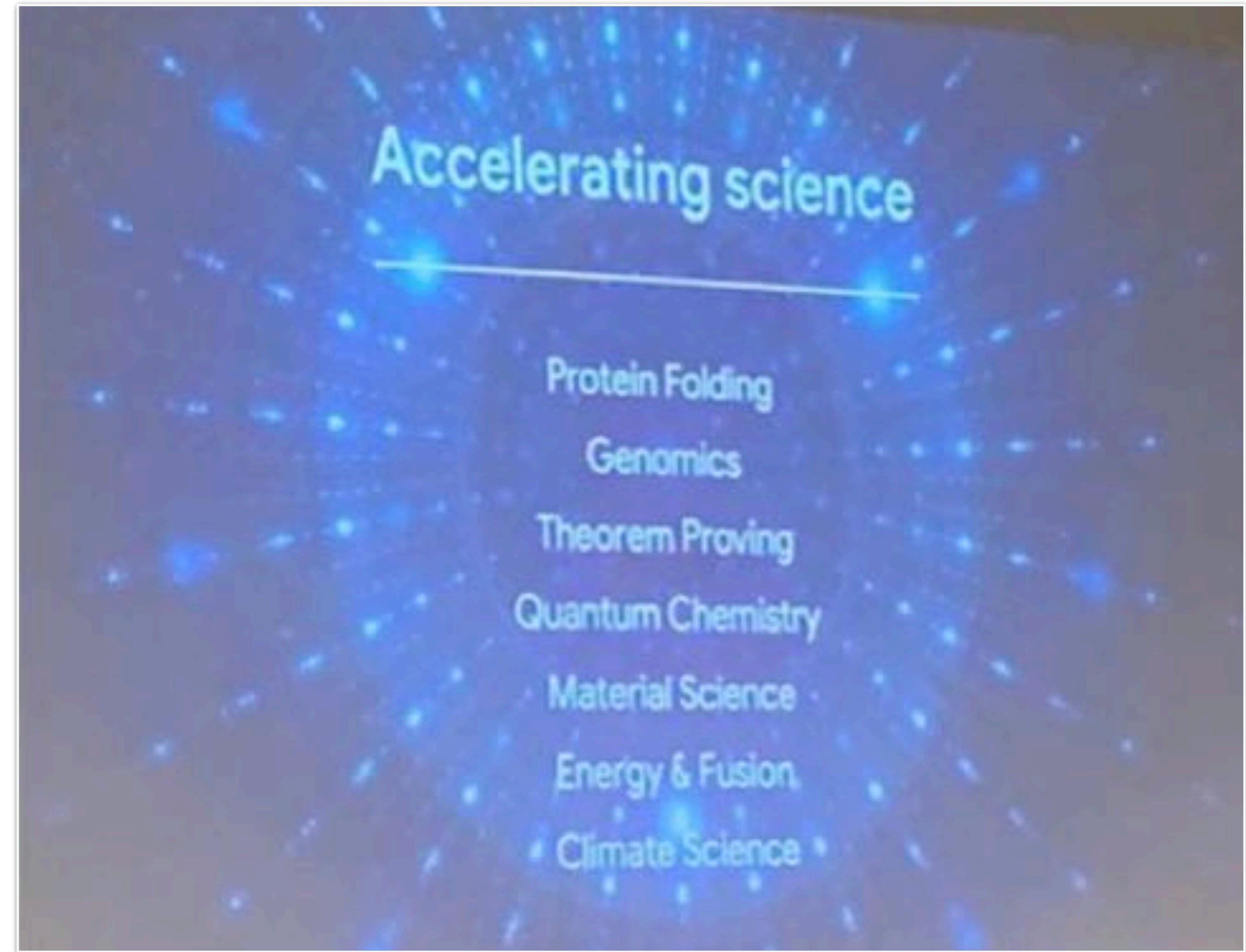


Natasha Jaques
@natashajaques

@demishassabis gave an absolutely packed talk yesterday about how @DeepMindAI is shifting its focus from games -> scientific benefit 🎉🎉



05:33 · 2019/3/22 · Twitter for iPhone



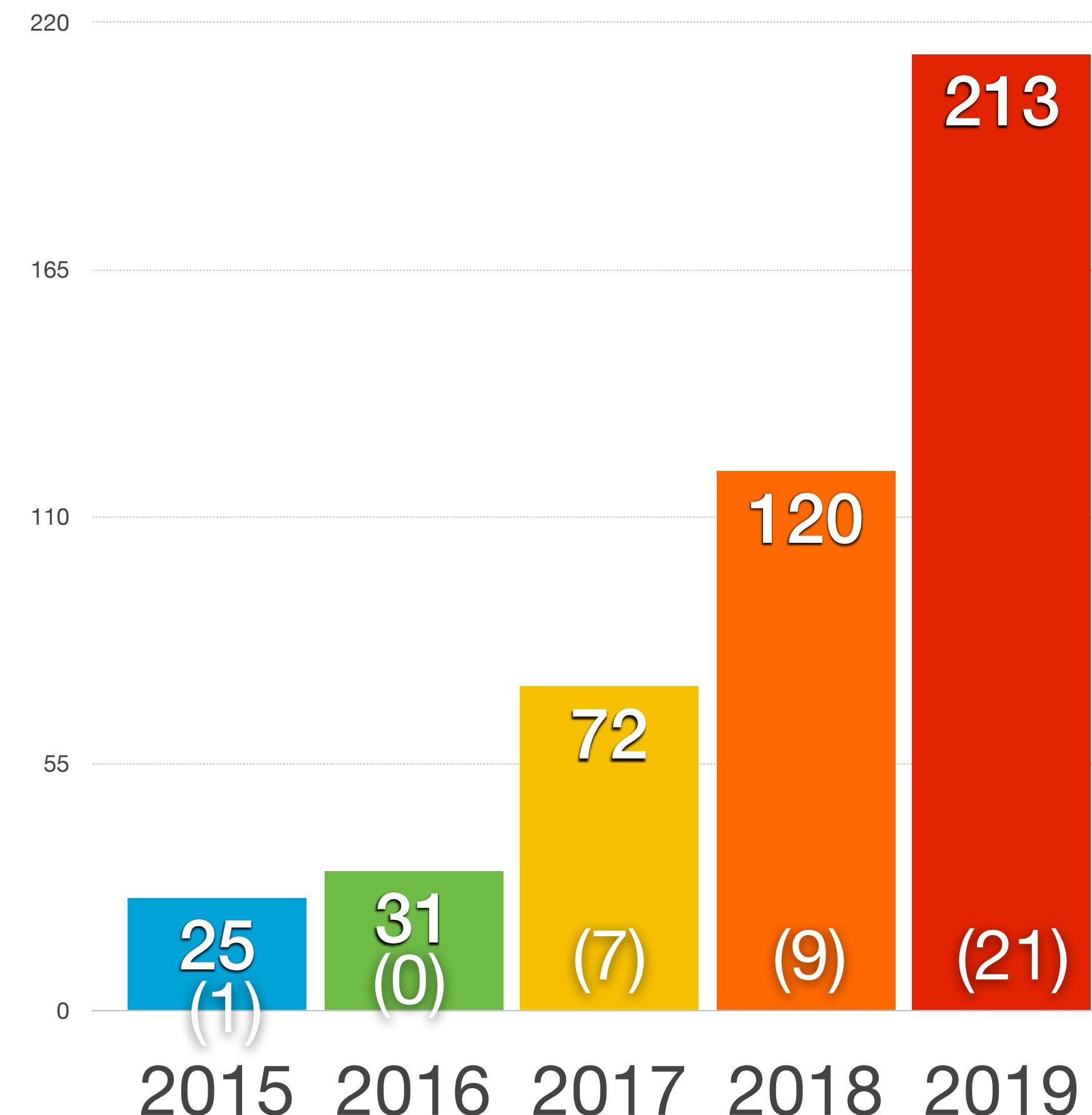
Demis Hassabis, Google DeepMind
2019.3.20@MIT

<https://www.youtube.com/watch?v=cEOAerVz3UU>

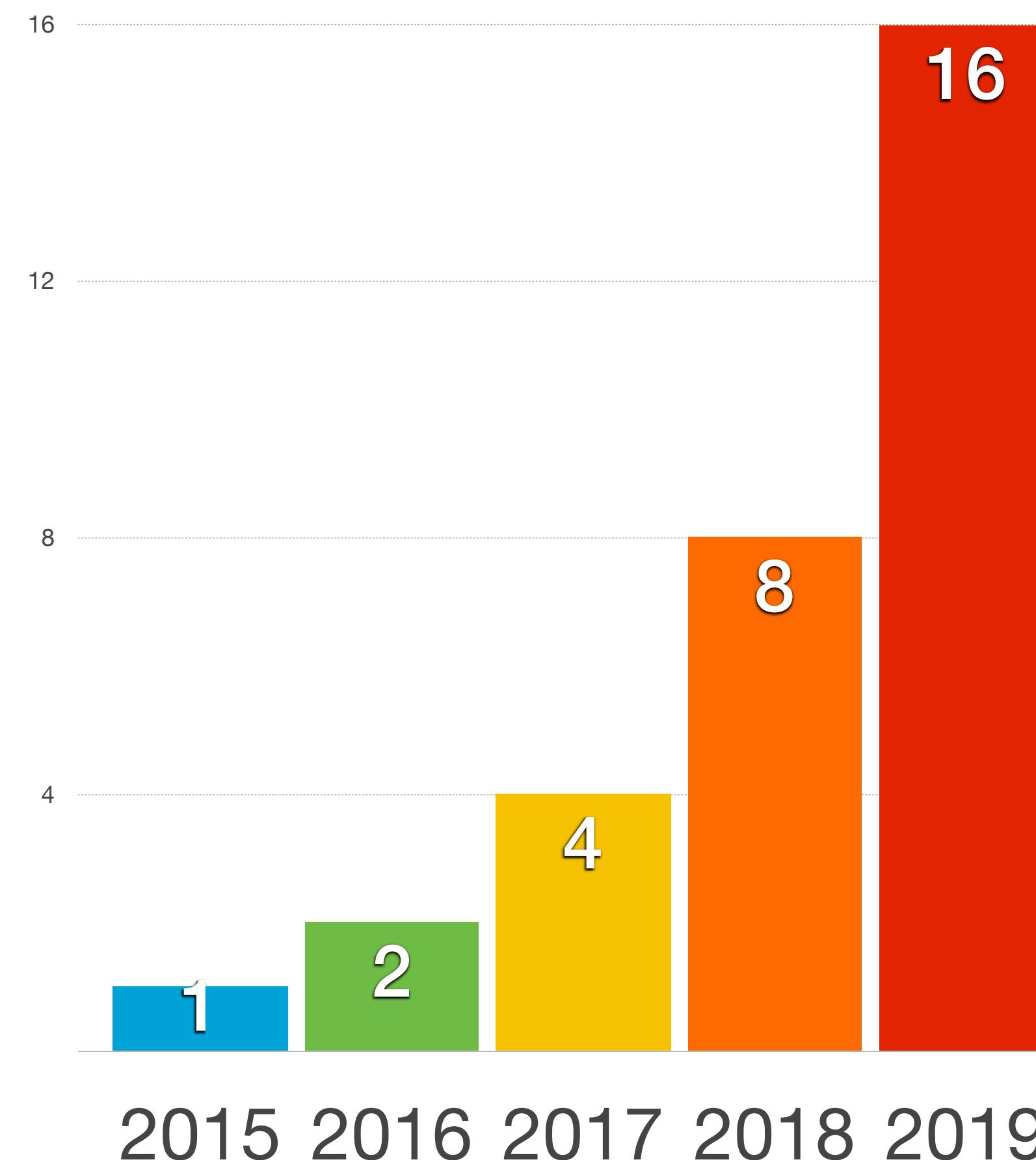
At APS March Meetings

with “machine learning” in the title/abstract

of (invited) talks



of sessions





Max Welling
@wellingmax

...

Proud to share the webpage for the brand-new Microsoft Research Lab Amsterdam where we will work on molecular simulation for green technology and healthcare. We are hiring deep learning for computational chemistry researchers and engineers.



Microsoft Research Lab - Amsterdam - Microsoft Research
“Almost everything around us is made of molecules, and yet it remains challenging to predict the properties of molecule...
[microsoft.com](#)

...



Matthias Troyer
@MatthiasTroyer

We are hiring in machine learning for quantum chemistry. Join a team pushing the frontier of what is possible:



Applied Researcher in Machine Learning for Quantum Che...
Apply for Applied Researcher in Machine Learning for Quantum Chemistry and Materials Science job with ...
[careers.microsoft.com](#)



量子化学与机器学习方向研究员

北京 · 研发 · 社招 · 职位 ID: 9QY1

I 职位描述

- 1、从事量子化学 (quantum chemistry) 和机器学习交叉领域的研究与应用；
- 2、进行基于机器学习的量子化学模型的设计与理论分析；
- 3、进行大规模量子化学模拟计算，验证模型的正确性和实用性。

I 职位要求

- 1、量子化学、机器学习及相关领域的研究人员或博士毕业生；
- 2、从事过量子化学、机器学习或相关领域的研究工作；
- 3、有扎实的数学基础，有一定的计算机编程能力；
- 4、有机器学习背景的优先；
- 5、在顶级会议或期刊发表过论文的优先。

【限制】211204林峰：如何打败DeepMind

在这个时间点讨论这个话题是有意义的。这对我们影响将不仅是能否迎接好眼下AI for Science这一主题下“创新-落地”的事业，还将是能否迎接好未来源源不断的底层算法、技术驱动的“创新-落地”浪潮。

本文中的“我们”泛指深势科技、AISI、DeepModeling开源社区及以此为种子团结的志同道合的同伴们。本文会更加从技术角度讨论相关问题，商业化的部分从略。

2020.9:《落地信》

各位朋友，

克服了一系列困难，今天我终于落地上海。从现在起，我将致力于和大家一起，在中国建立起一个面向未来的、满足国家战略定位的、经得住考验的人工智能及科学计算基础研究和技术创新体系。

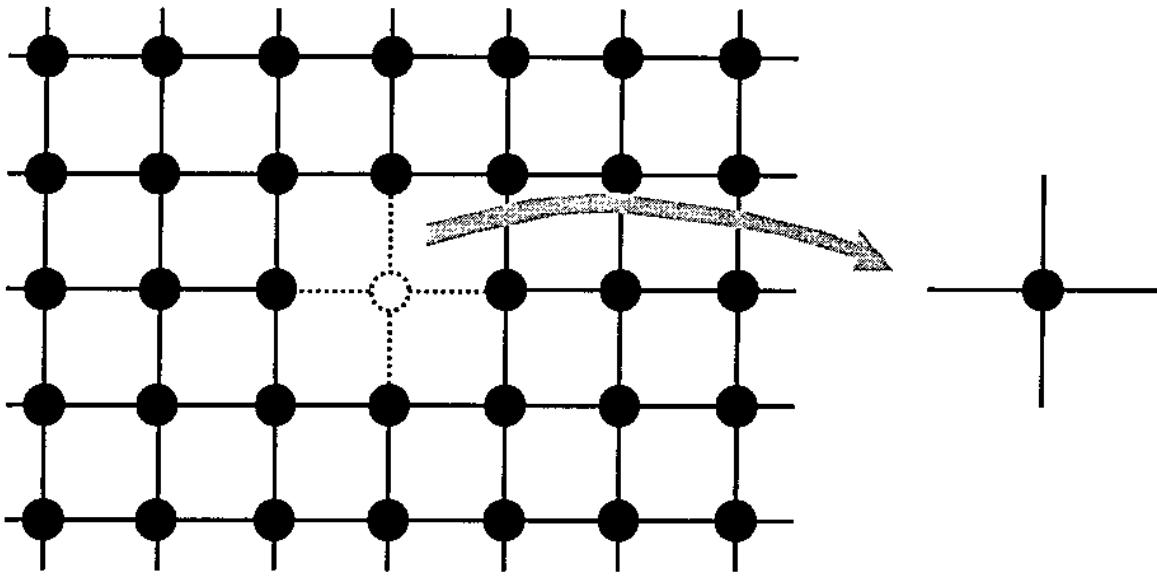
这个体系的建立将基于如下几点认识：

1. 传统科学领域（如物理、化学、生物、材料、工程）应该成为人工智能的主战场。
.....
2. 应用和计算数学将成为人工智能基础模型和算法创新的主要力量。
.....
3. 传统科学计算与机器学习方法的结合，将允许我们全面升级科学计算的能力。
.....

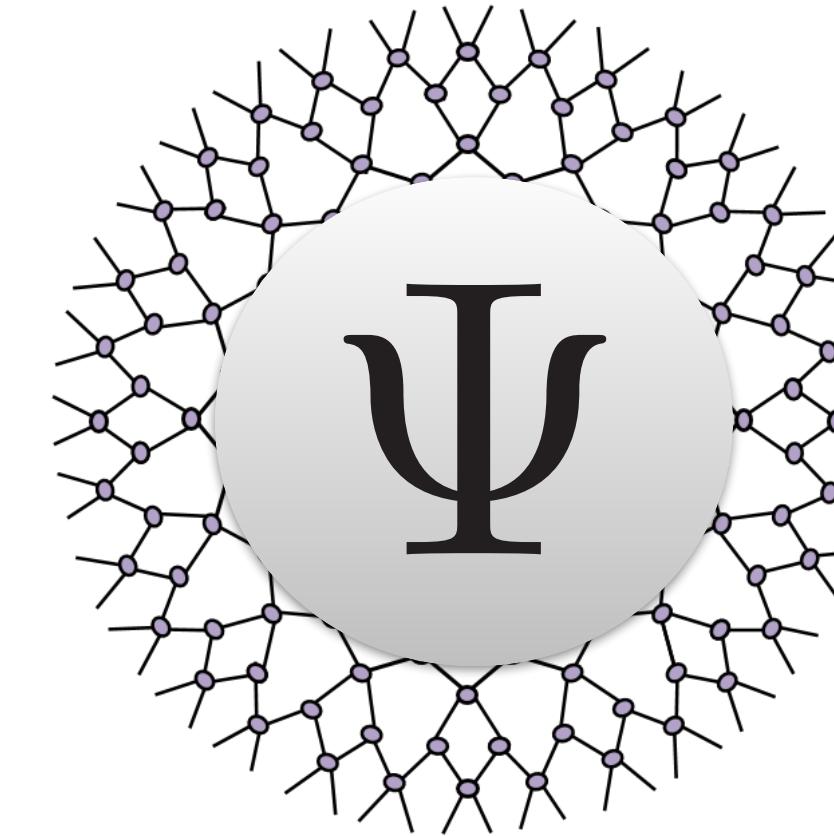


Physicists' gifts to Machine Learning

Mean Field Theory



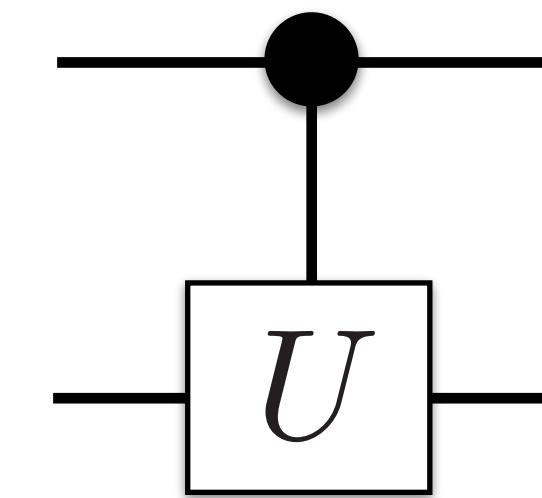
Tensor Networks



Monte Carlo Methods

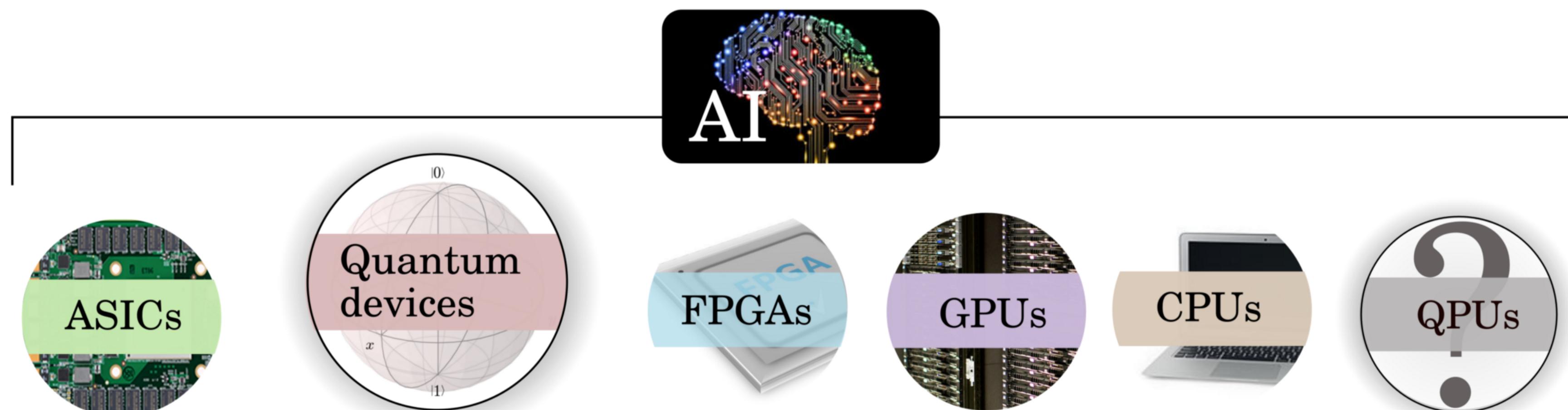


Quantum Computing

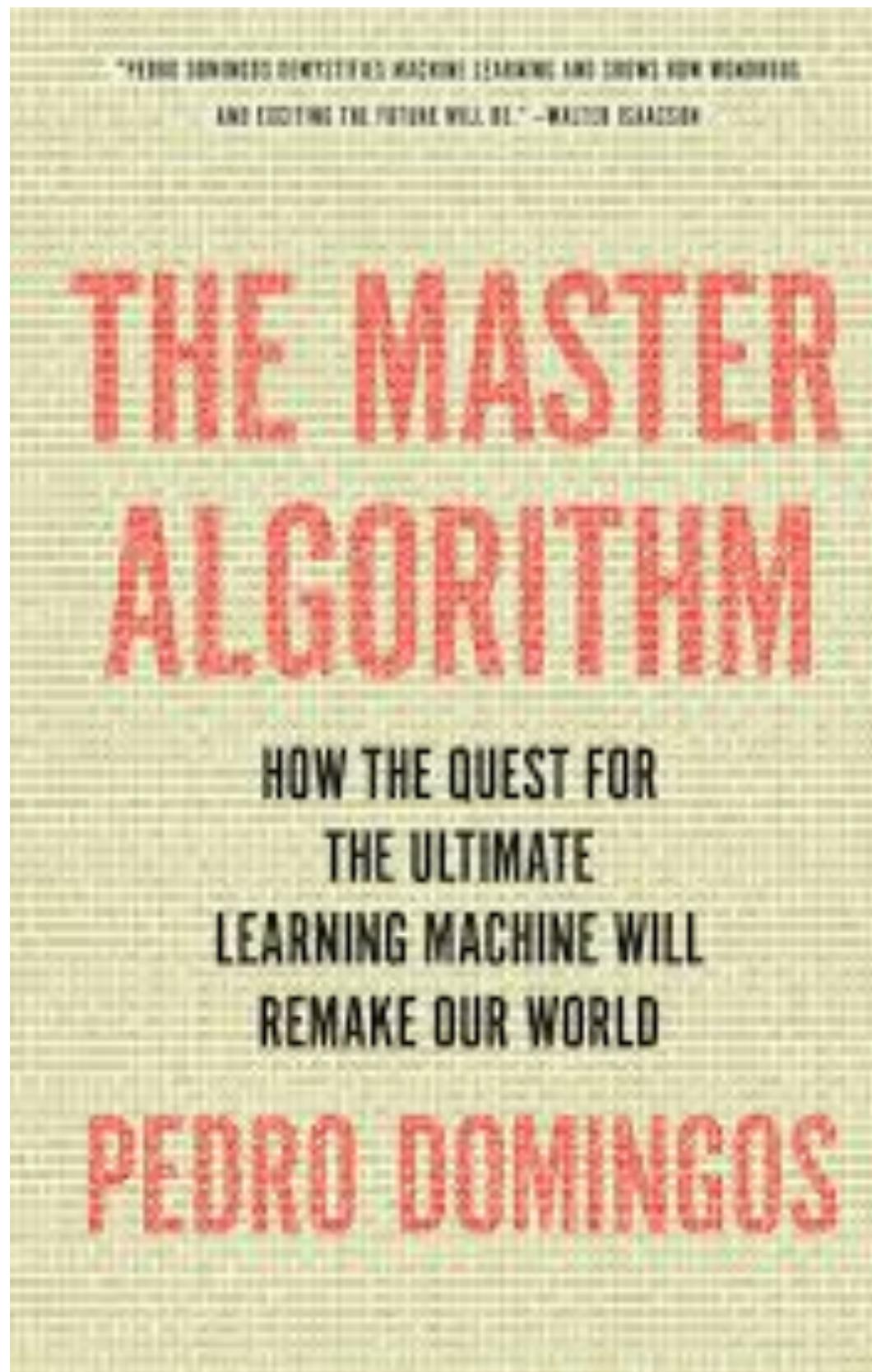


Physicists' gifts to Machine Learning

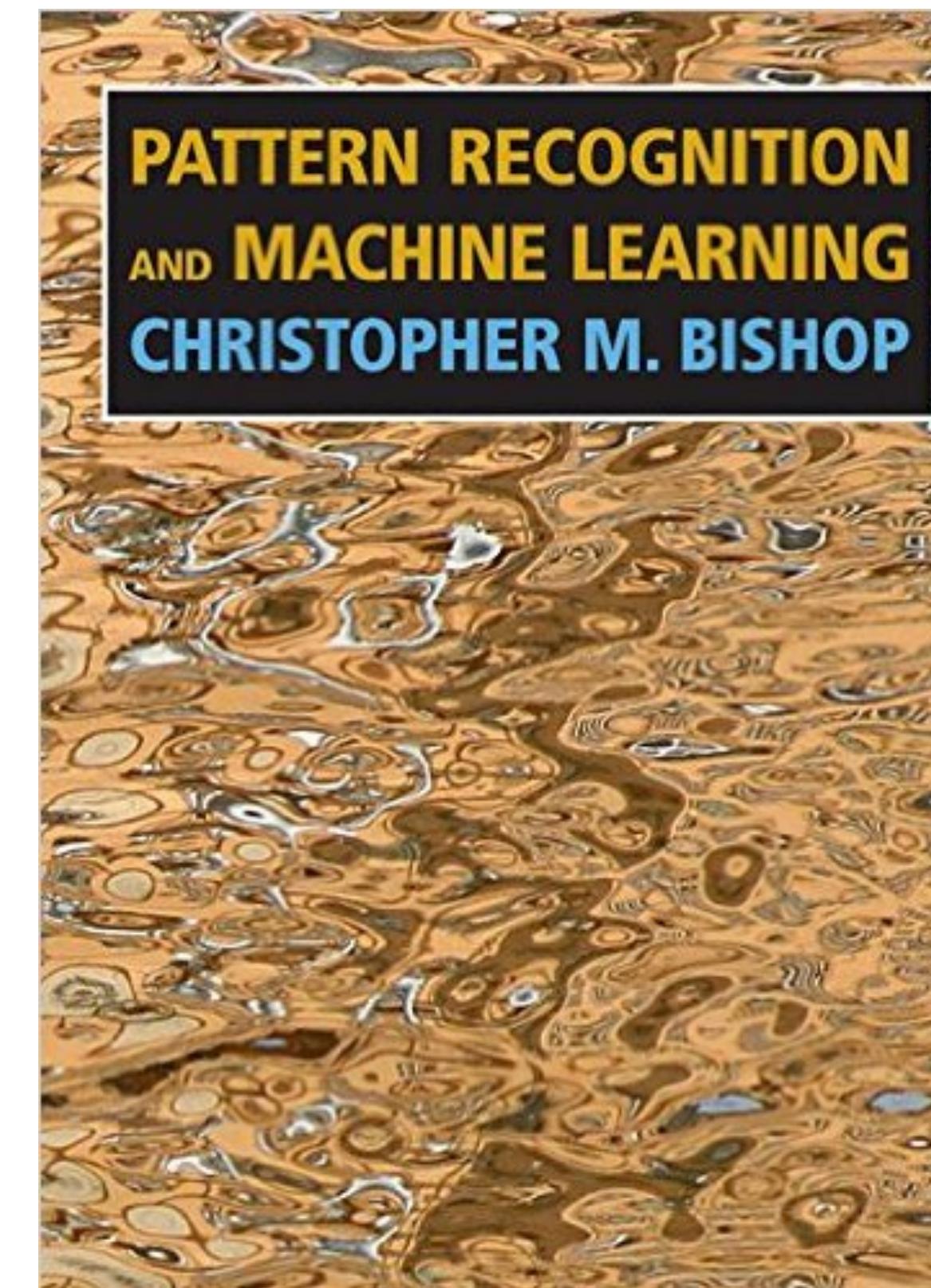
- Tensor network is 21 century's matrix (Mario Szegedy)
- Started as a technique for statistical and quantum many-body physics. Now, entering machine learning.
- Is QPU the next-generation coprocessor that powers AI?



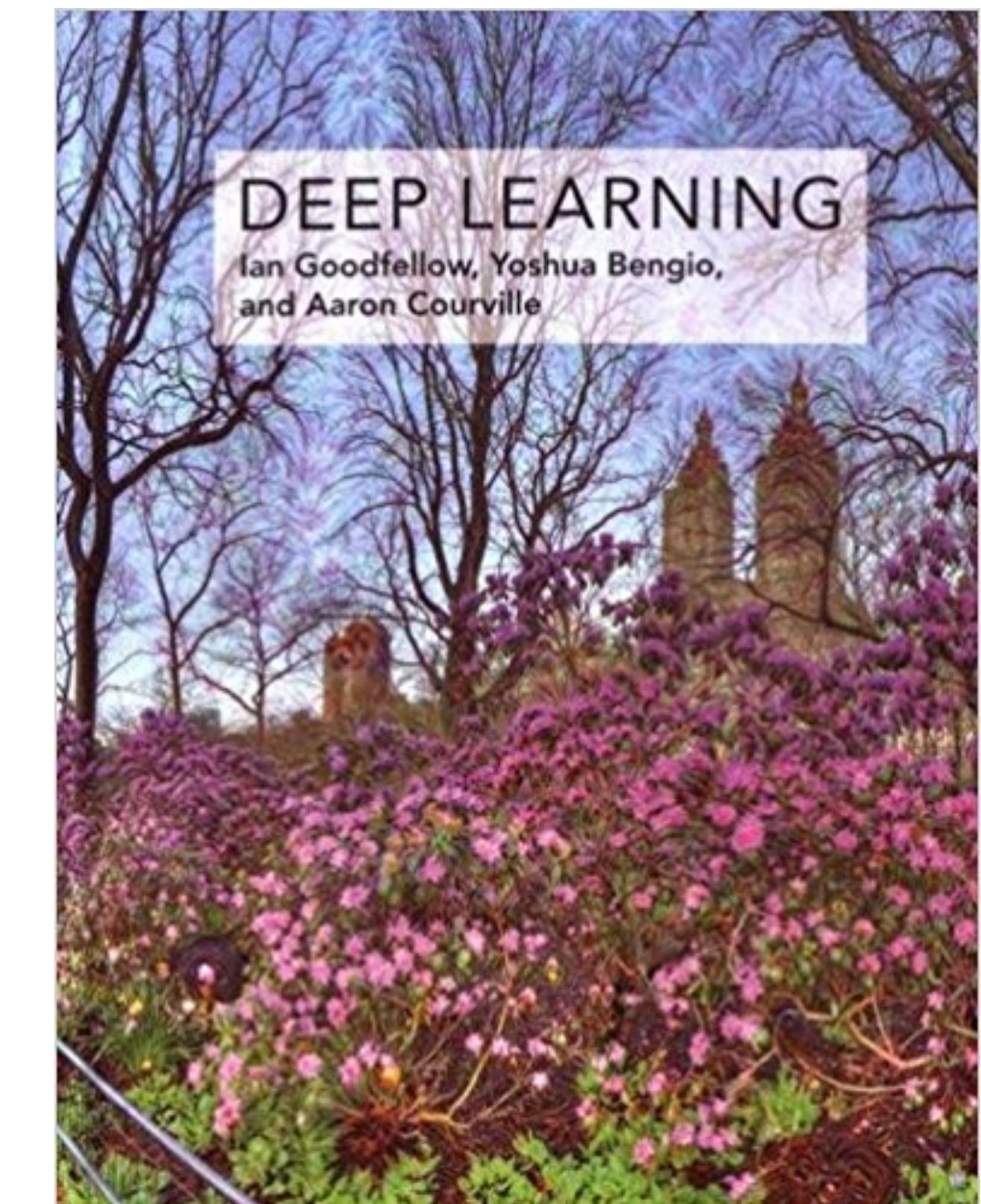
Introductory Books



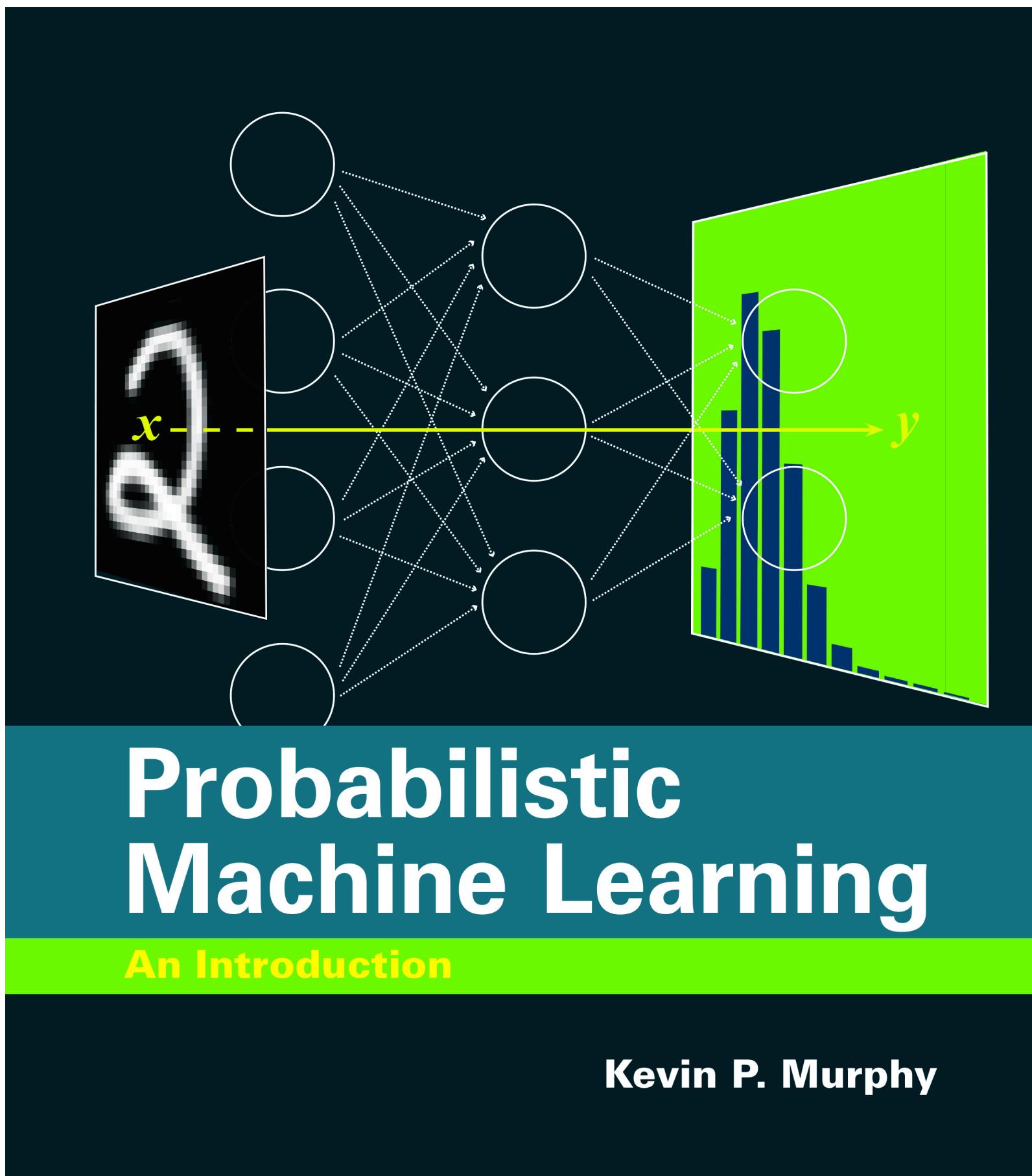
Popular overview
2015



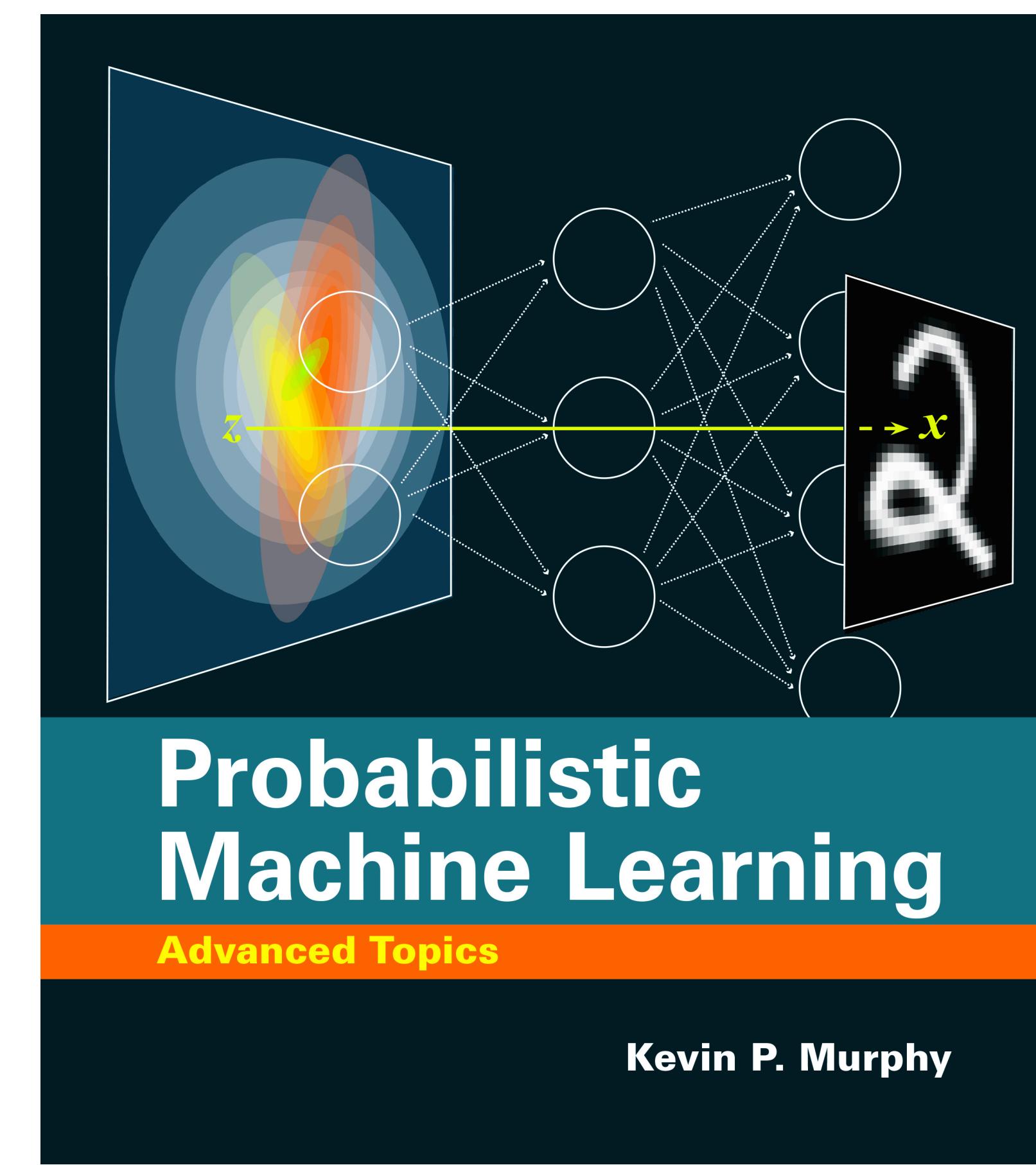
Solid math
2006



“Modern” topics
2016



2022 (855 pages)

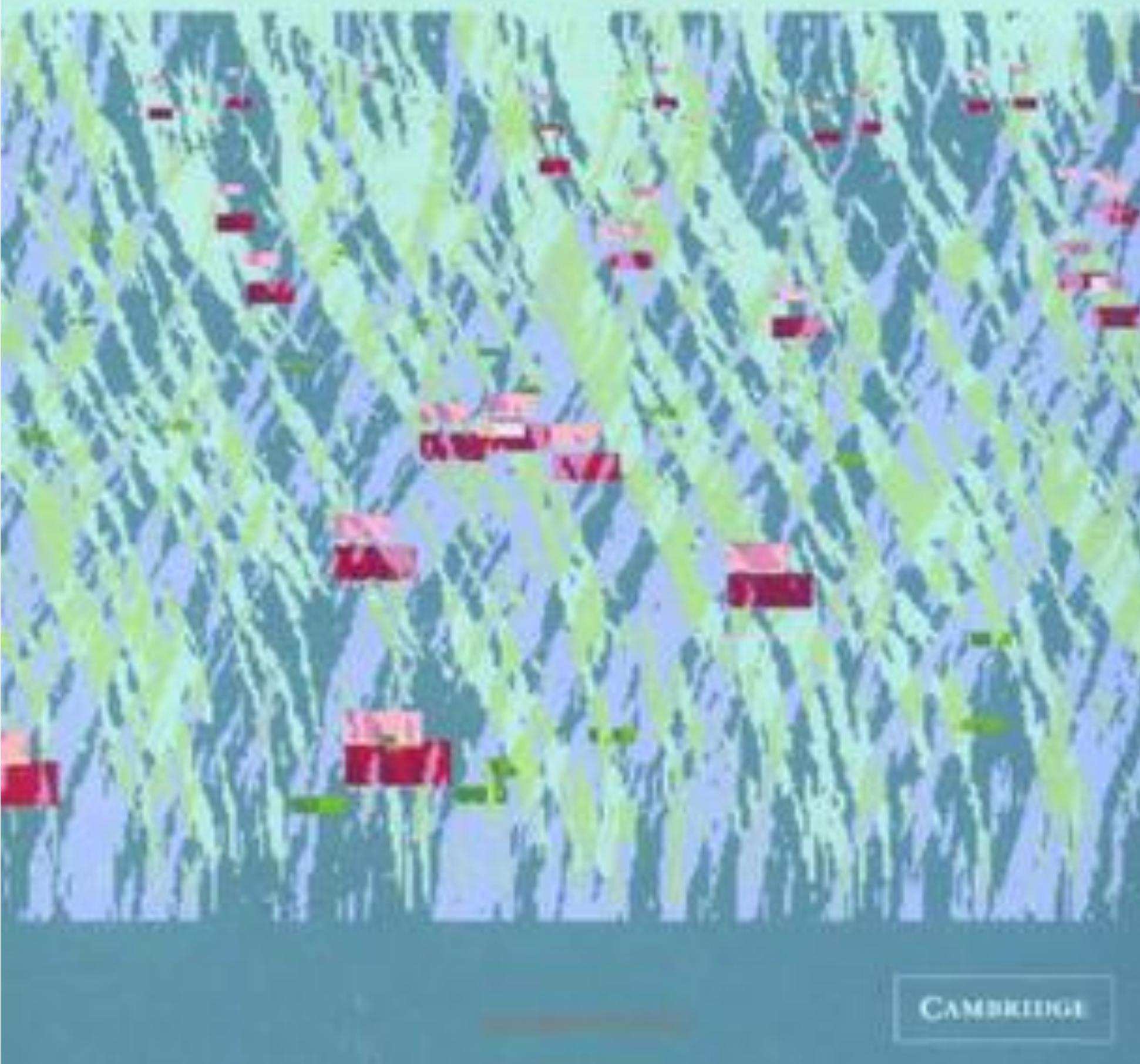


2023 (1352 +332 pages)

<https://probml.github.io/pml-book/>

David J. C. MacKay

Information Theory, Inference, and Learning Algorithms



Insightful, fun to read

$$H(X, Y)$$

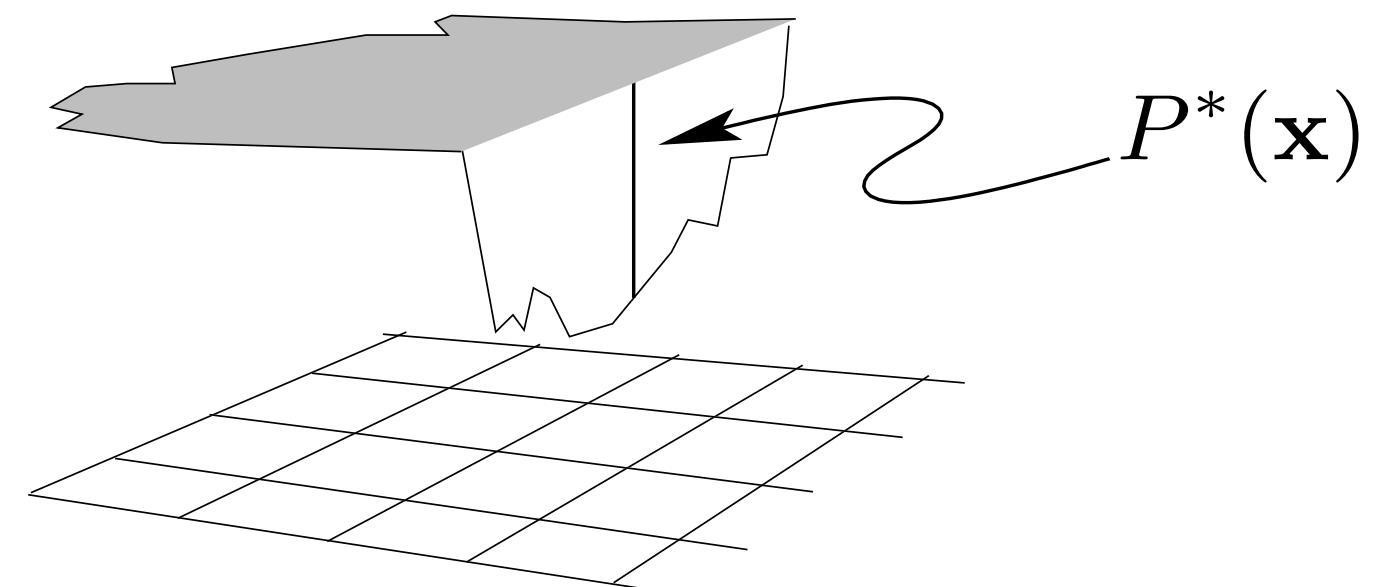
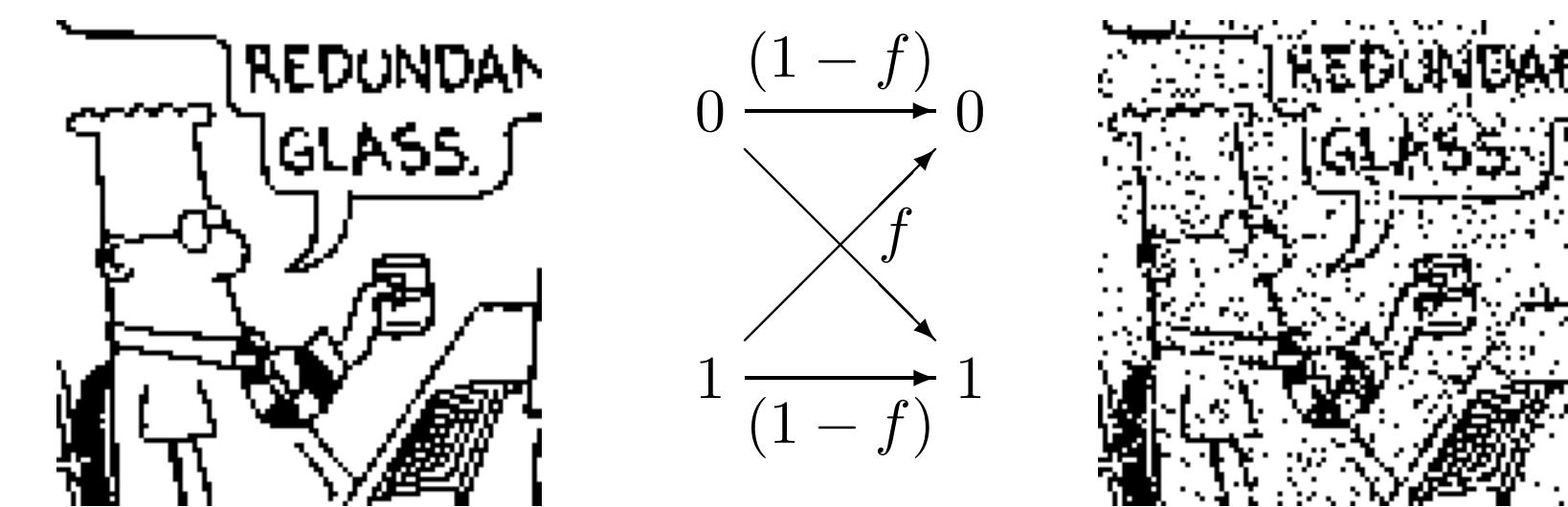
$$H(X)$$

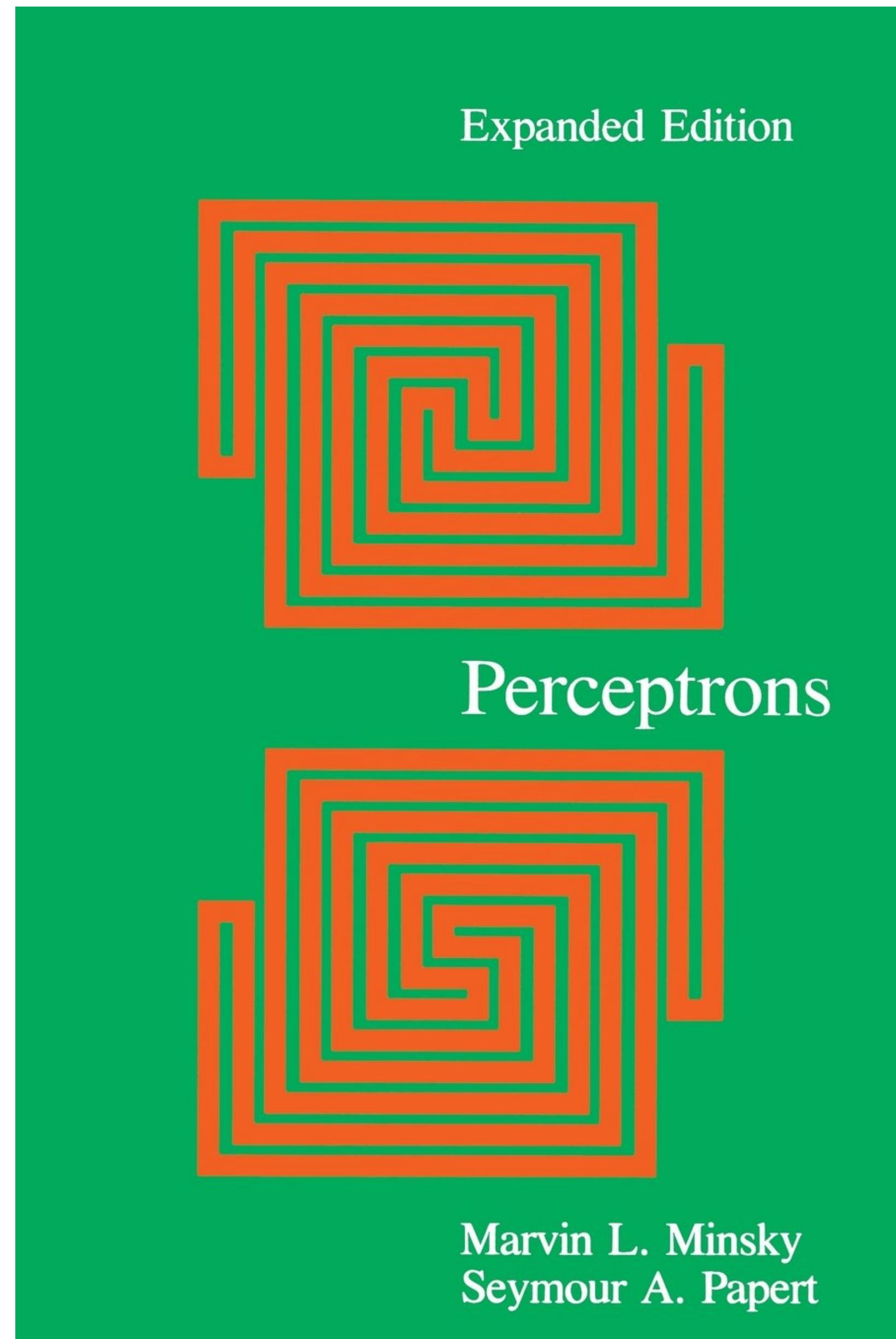
$$H(Y)$$

$$H(X | Y)$$

$$I(X; Y)$$

$$H(Y|X)$$





1st edition 1969
expanded 1972
commented 1988

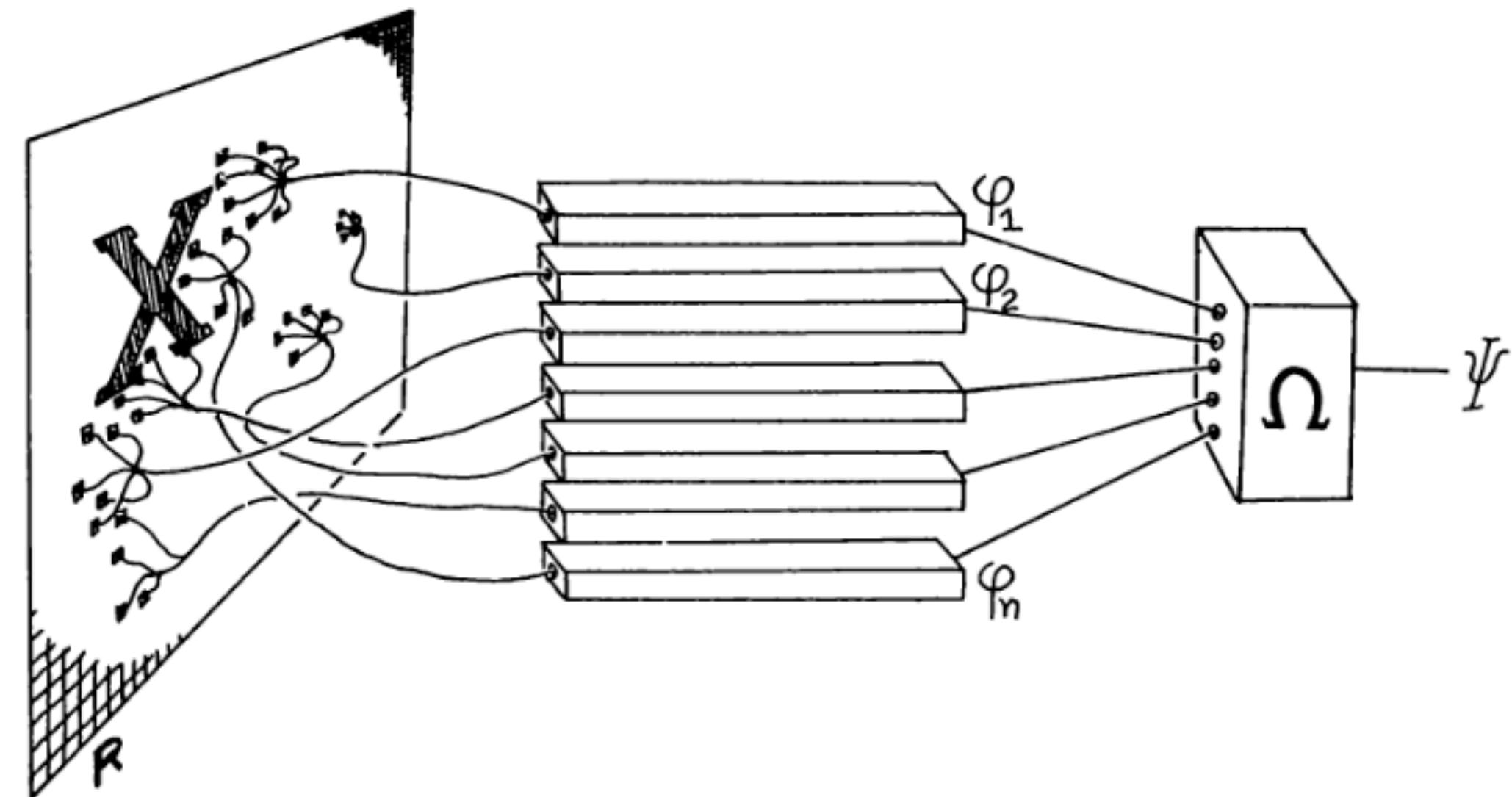
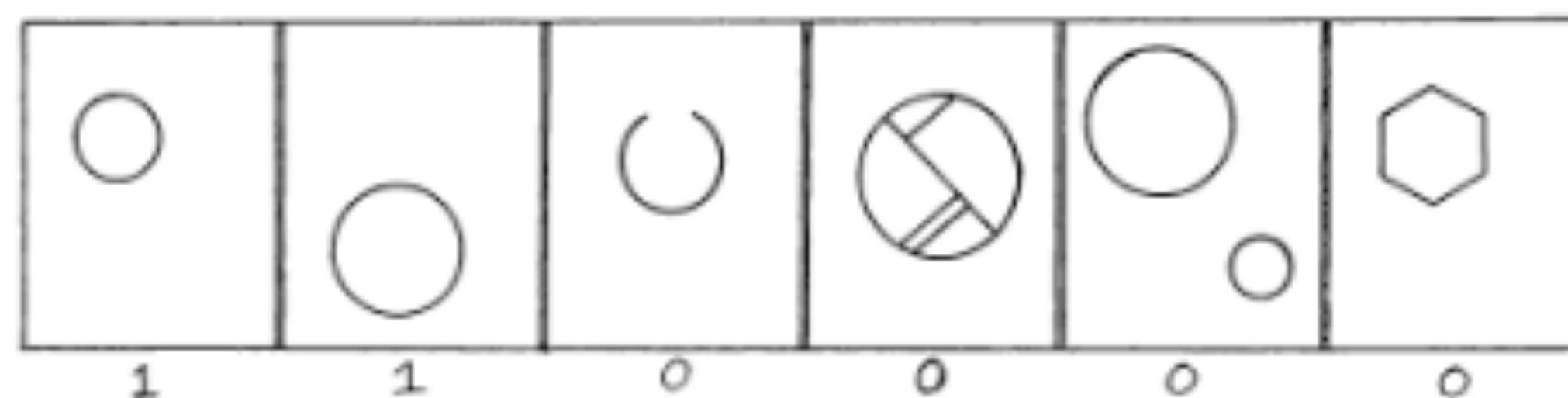


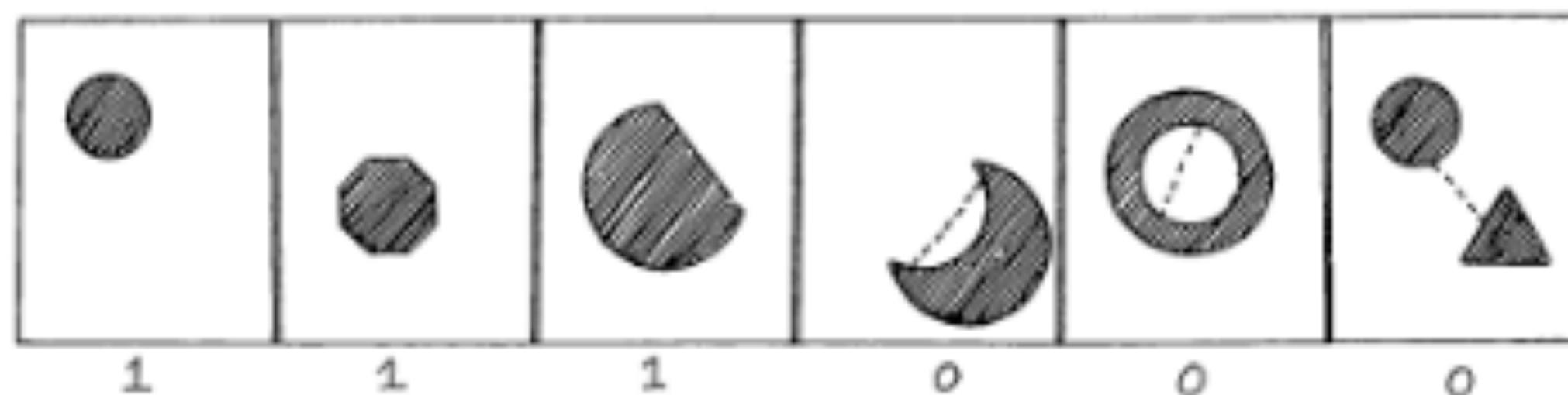
Figure 0.1

on page 6 of
perceptrons

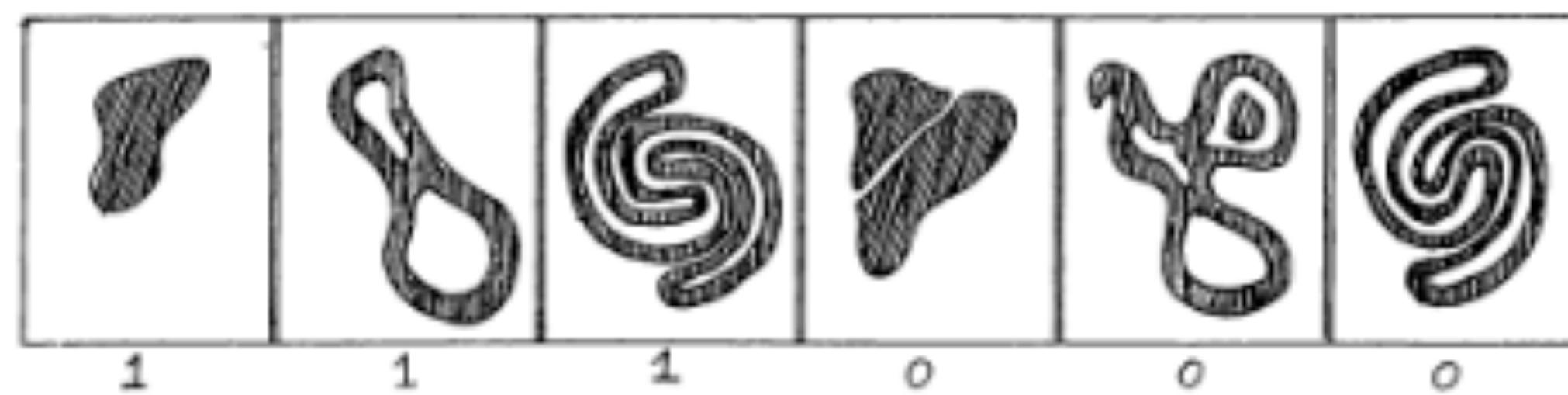
$$\psi_{\text{CIRCLE}}(X) = \begin{cases} 1 & \text{if the figure } X \text{ is a circle,} \\ 0 & \text{if the figure is not a circle;} \end{cases}$$



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



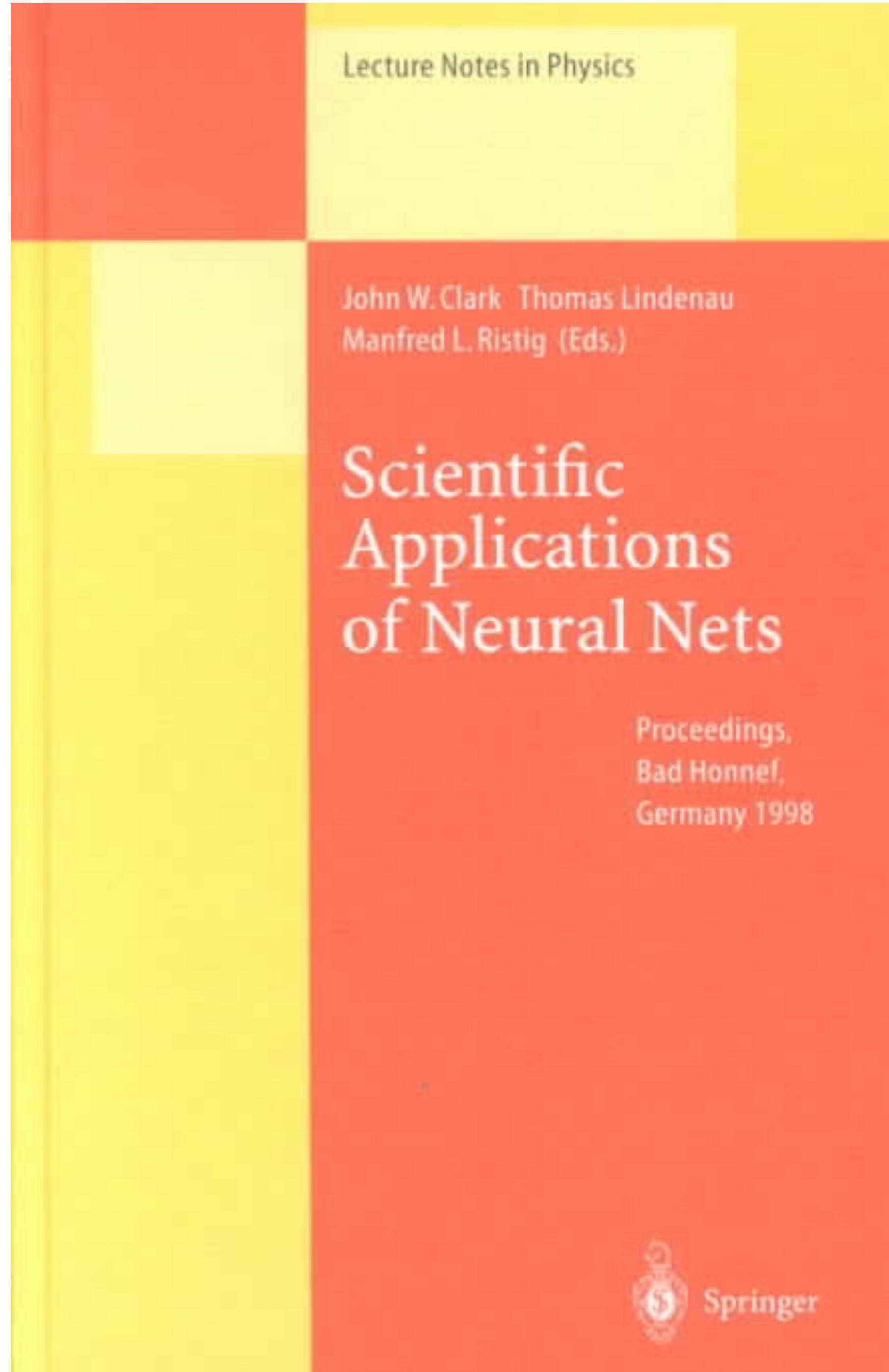
$$\psi_{\text{CONNECTED}}(X) = \begin{cases} 1 & \text{if } X \text{ is a connected figure,} \\ 0 & \text{otherwise.} \end{cases}$$



0.3 Cybernetics and Romanticism

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

Reflection between this and last hype cycle



8 Doing Science With Neural Nets: Pride and Prejudice

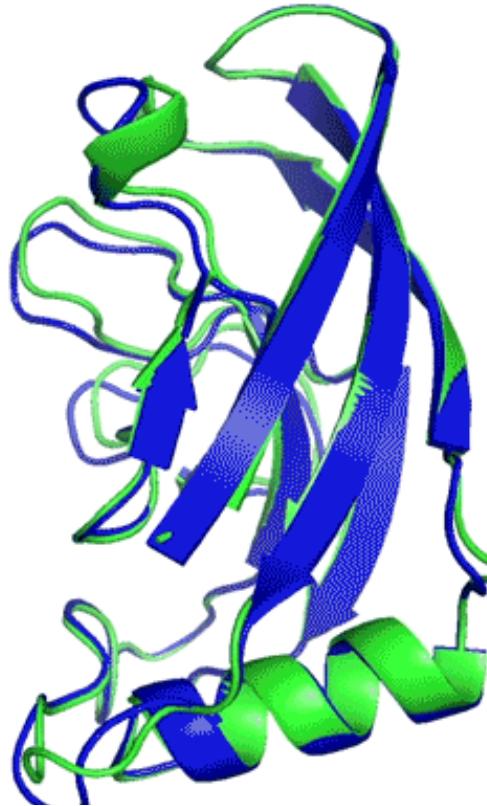
When neural networks re-emerged on the scene in the mid-80s as a new and glamorous computational paradigm, the initial reaction in some sectors of the scientific community was perhaps too enthusiastic and not sufficiently critical. There was a tendency on the part of practitioners to oversell the powers of neural-network or “connectionist” solutions relative to conventional techniques – where conventional techniques can include both traditional theory-rich modeling and established statistical methods. The last five years have seen a correction phase, as some of the practical limitations of neural-network approaches have become apparent, and as scientists have become better acquainted with the wide array of advanced statistical tools that are currently available.

Why now, again ?
What has changed ?
What has not ?

Let's see what have been made possible

Success stories

AlphaFold 2
@DeepMind

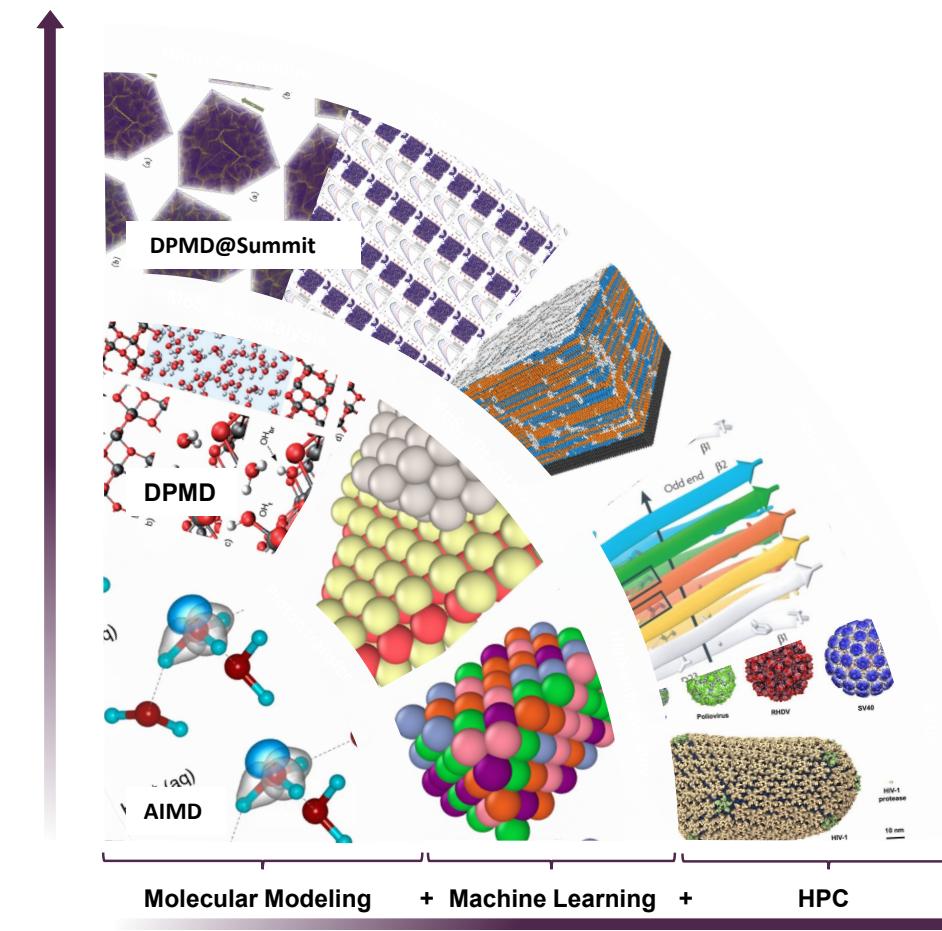


Trained on protein data

<https://deepmind.com/blog/article/alphafold-a-solution-to-a-50-year-old-grand-challenge-in-biology>

Machine learning scientific data

10^8 atoms simulation
@Deep Potential



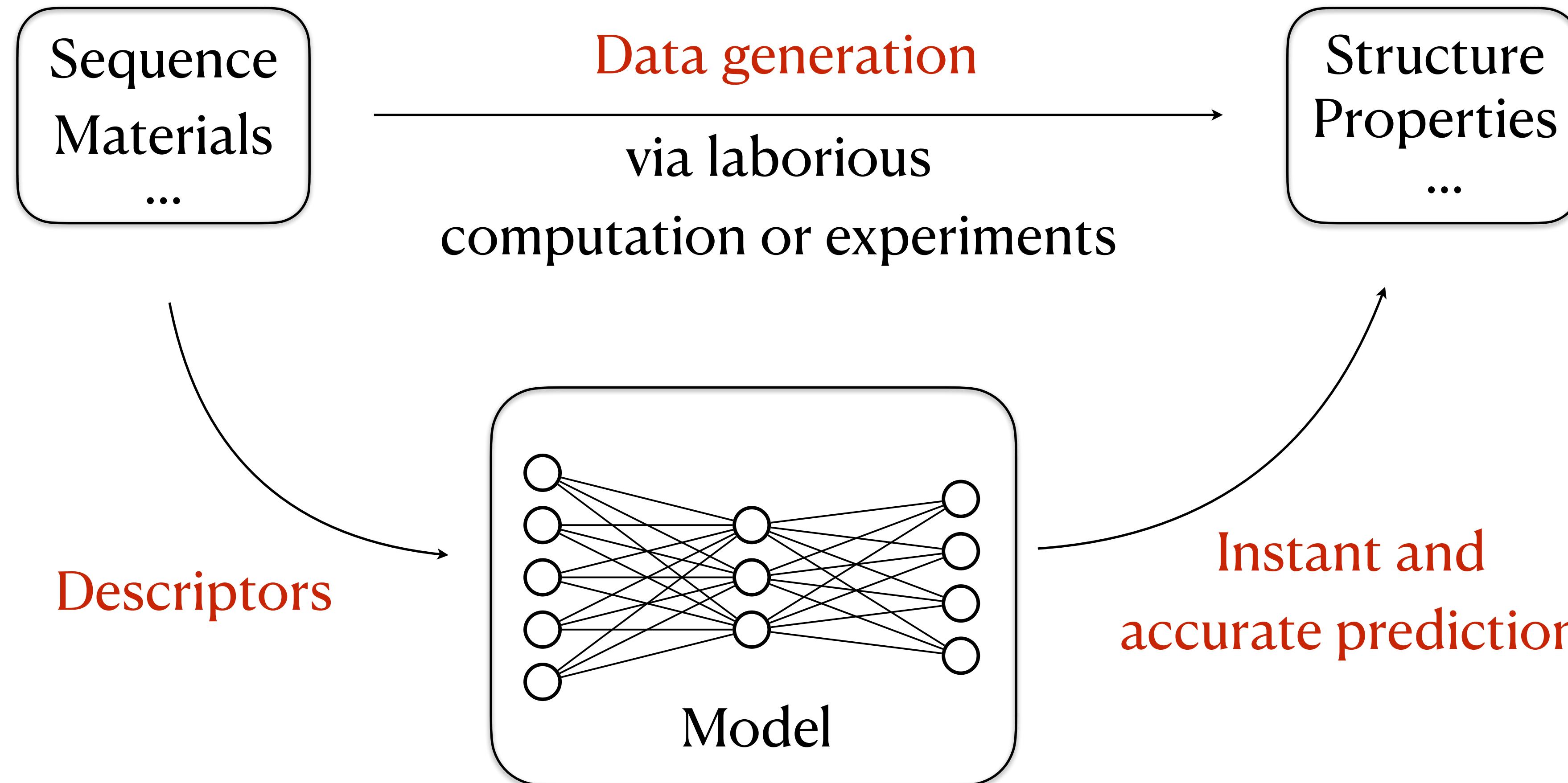
Trained on DFT data

<https://www.acm.org/media-center/2020/november/gordon-bell-prize-2020>

Machine learning aided simulations

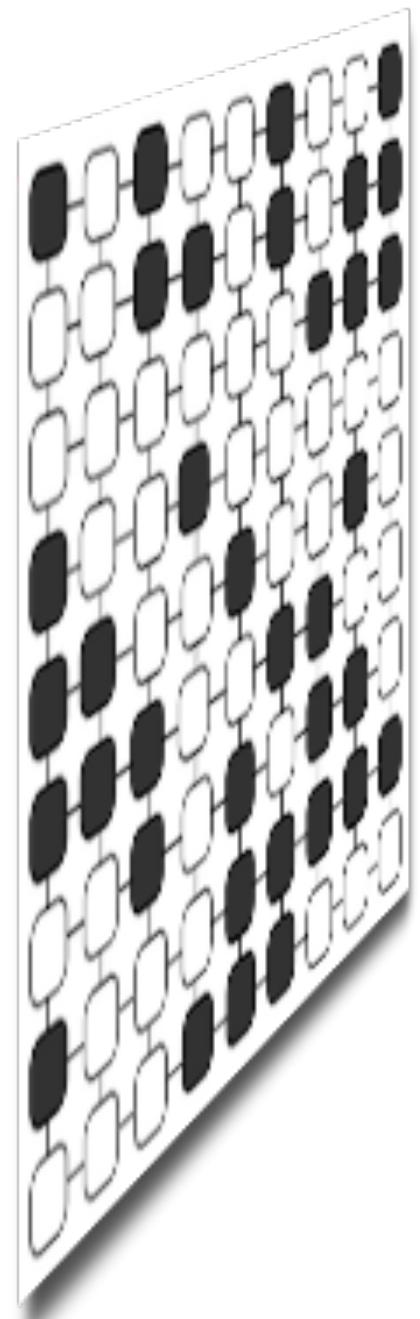
Machine learning scientific data

Machine learning material/chemical/biological properties

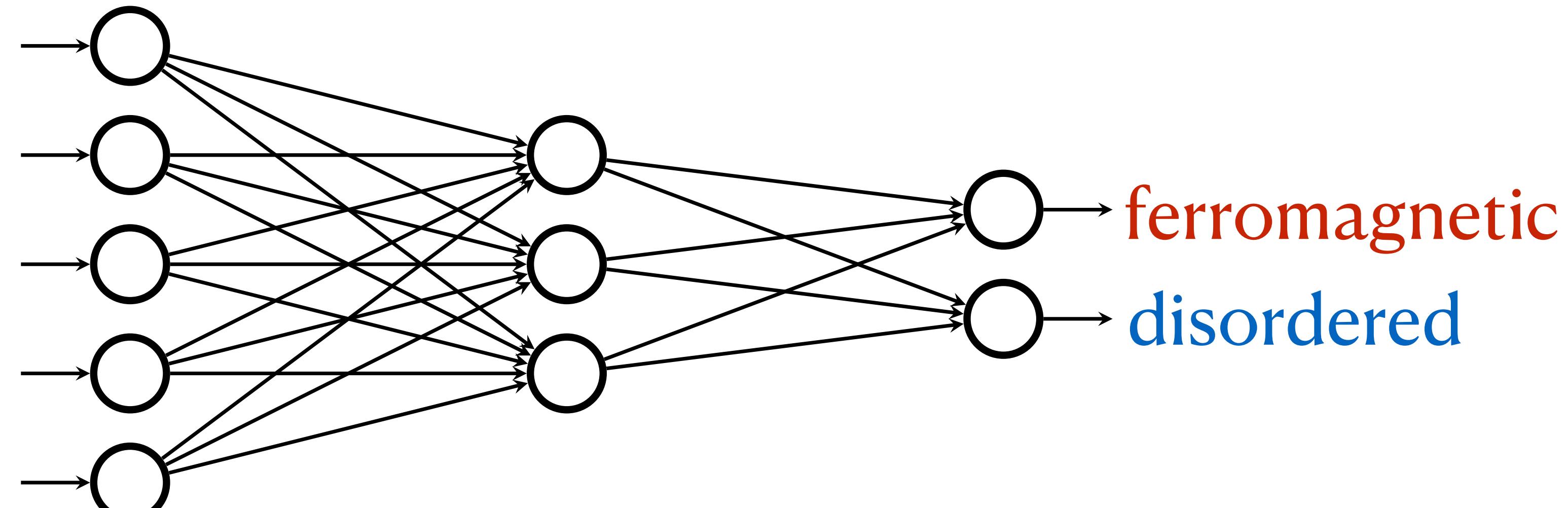


Machine learning phases of matter

Ising spins



data

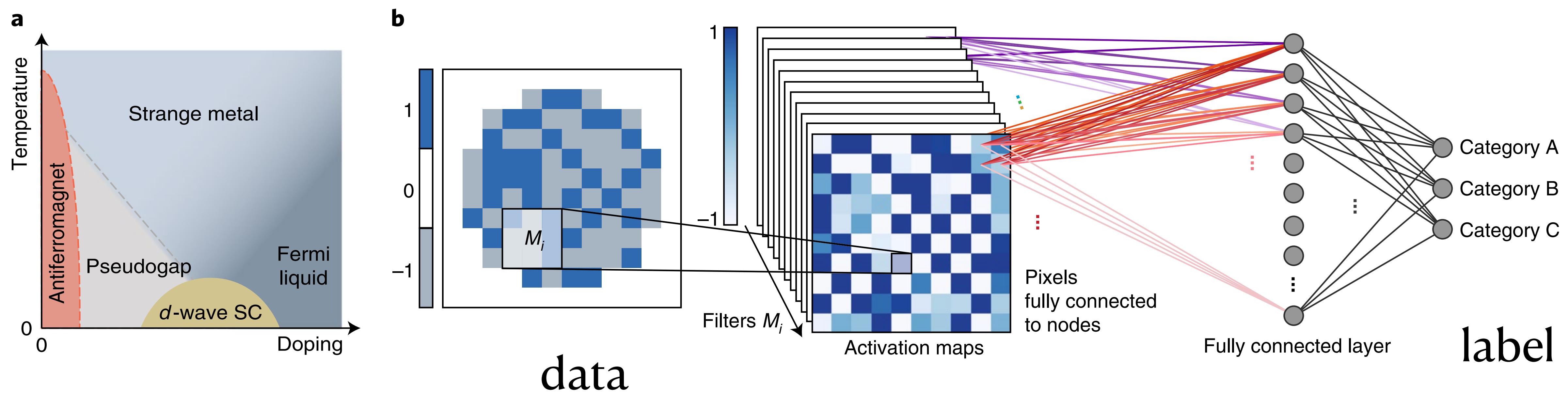


Carrasquilla and Melko, Nature Physics '17

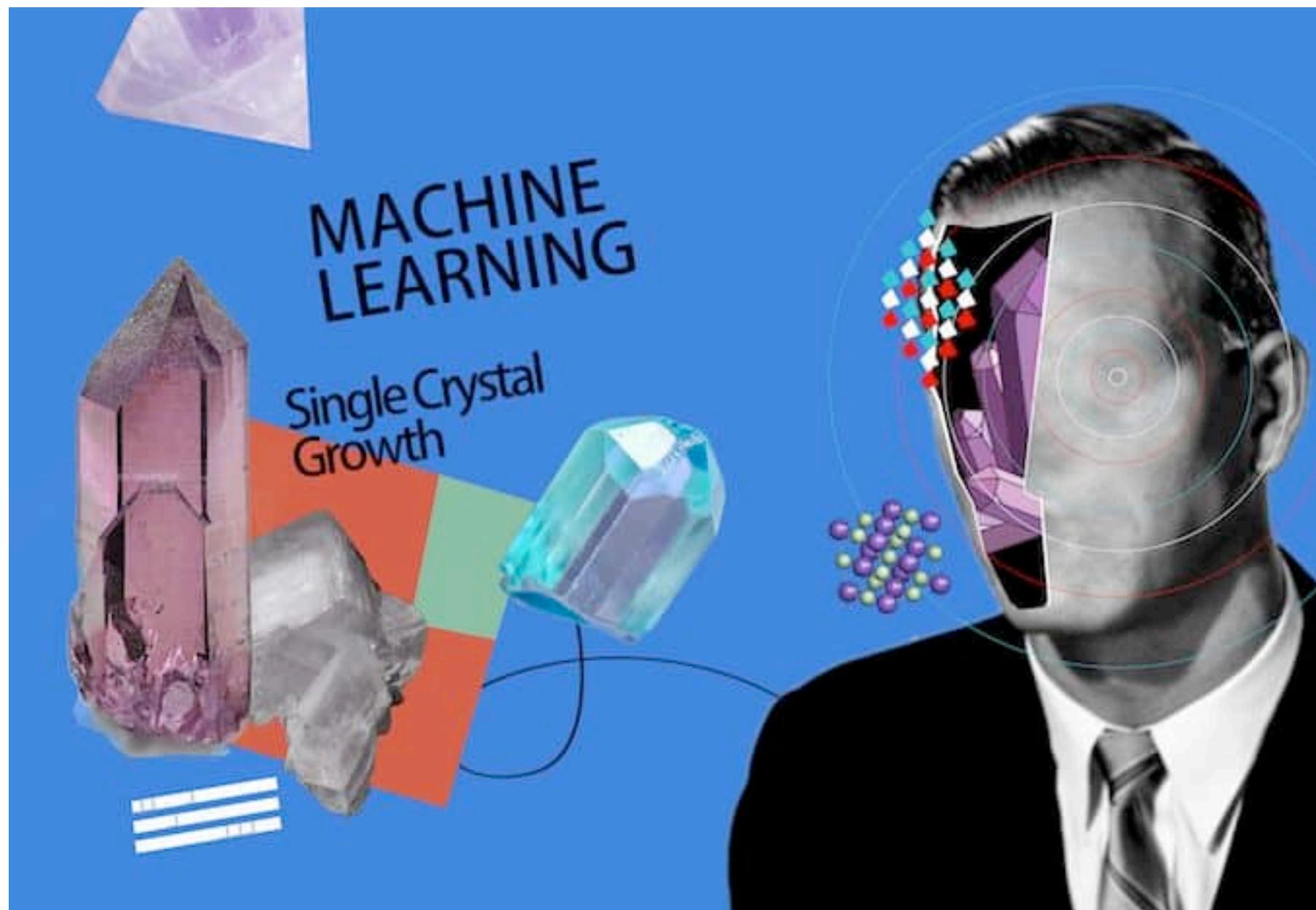
+actual experimental data (ultracold atoms/STM/...)

label

Machine learning phases of matter



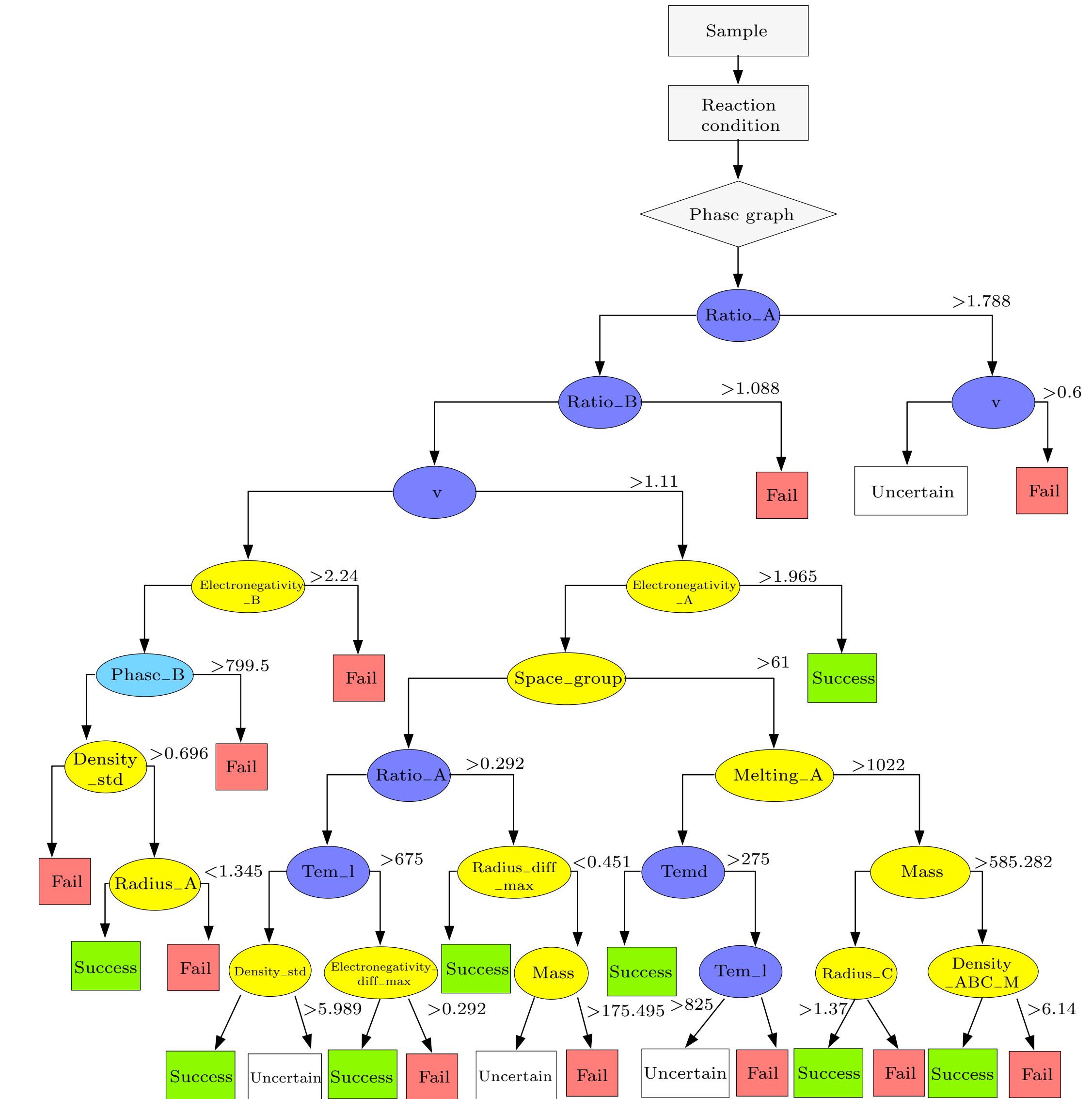
Machine learning crystal growth



<https://physicsworld.com/a/growing-better-single-crystals-with-the-help-of-machine-learning/>

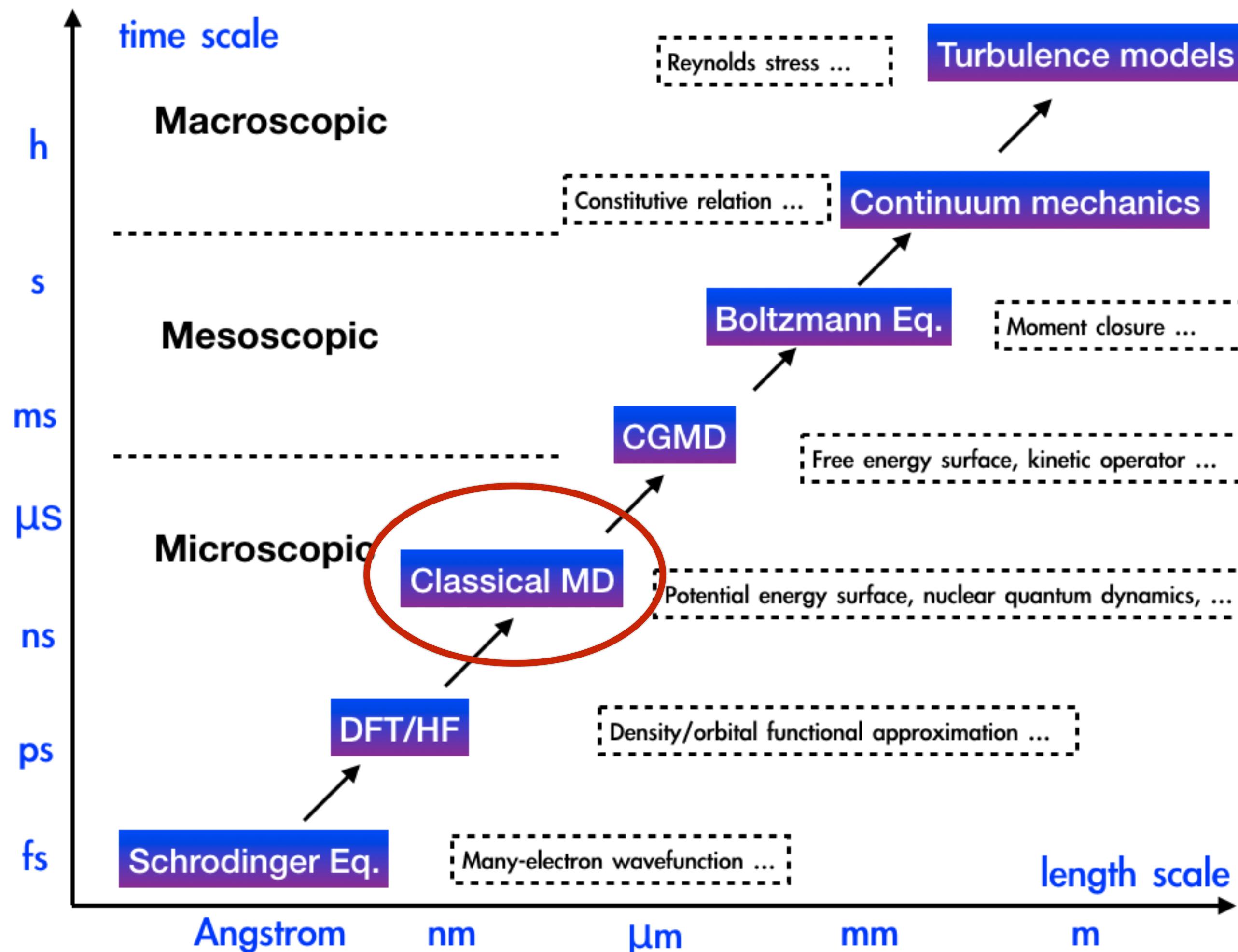
Yao et al, CPL 2019

Trained on digitized lab notebook



Machine learning aided simulations

The simulation dilemma



Molecular dynamics in 1 slide

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

Einstein's first paper, Annalen der Physik, 1901

Folgerungen aus den Capillaritätserscheinungen. 515

ziehungskräfte aus, und prüfte deren Consequenzen auf ihre Uebereinstimmung mit dem Experiment hin. Ich liess mich dabei von der Analogie der Gravitationskräfte leiten.

Sei also das relative Potential zweier Moleküle von der Form:

$$P = P_\infty - c_1 \cdot c_2 \cdot \varphi(r),$$

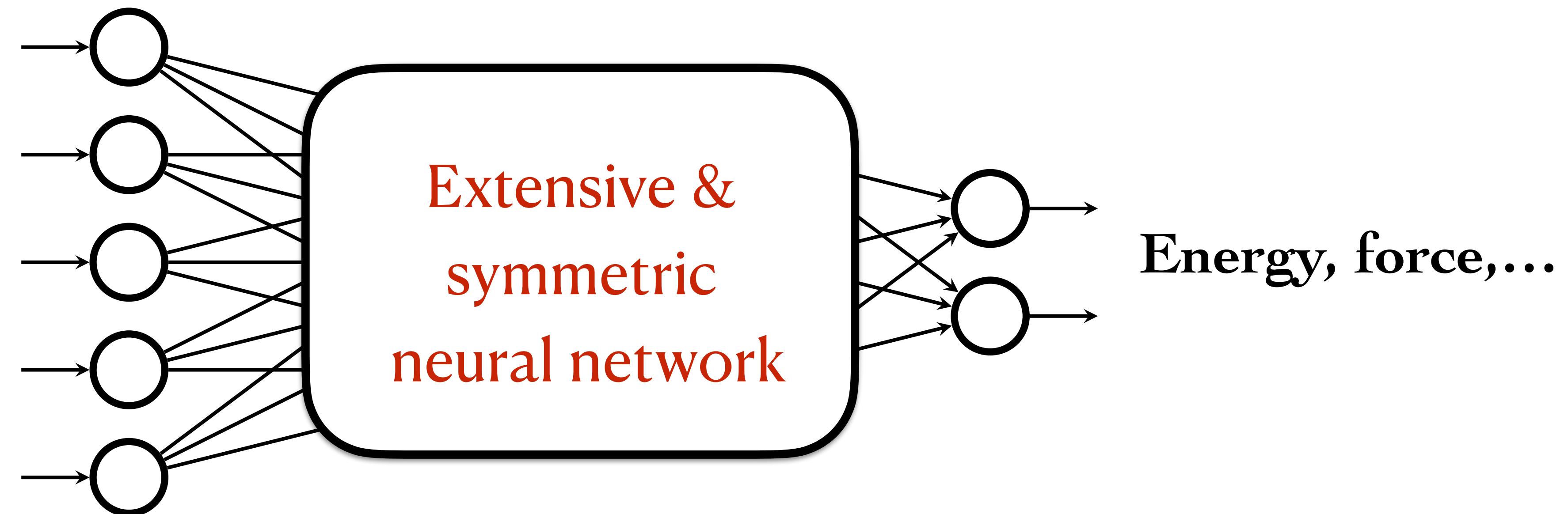
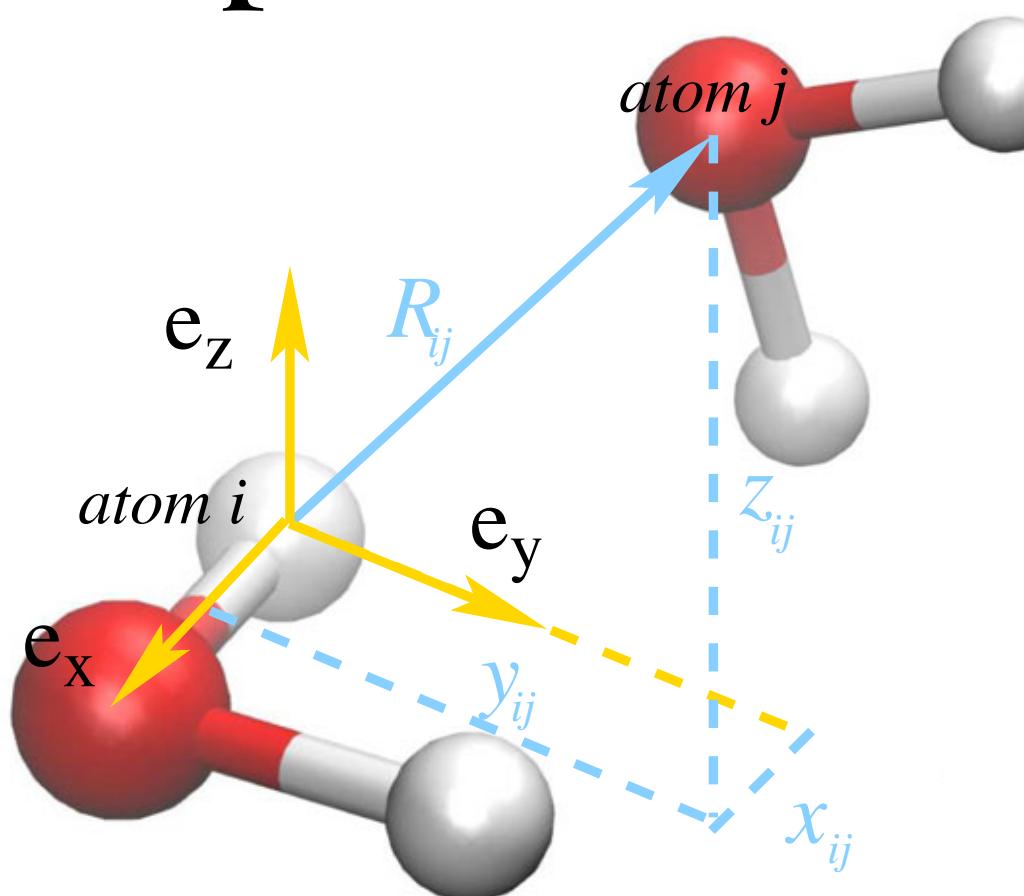
wobei c eine für das betreffende Molekül charakteristische Constante ist, $\varphi(r)$ aber eine vom Wesen der Moleküle unabhängige Function ihrer Entfernung. Wir nehmen ferner

$$m_I \frac{d^2 R_I}{d^2 t} = F_I$$

Integrate Newton's equation for nuclei to study thermal/structural/dynamical properties

Deep Potential Molecular Dynamics

Atom positions

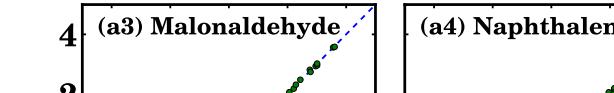
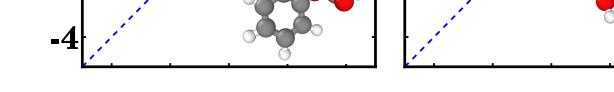
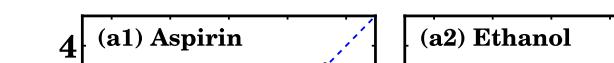


Behler, Parrinello, PRL '07

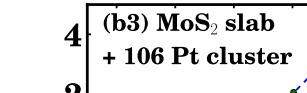
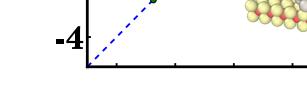
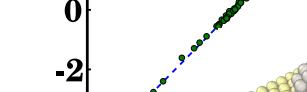
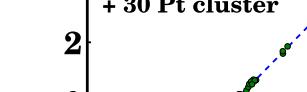
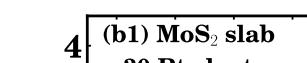
Zhang, Han, Wang, Car, E, PRL '18

- ① Fit neural network with DFT data
- ② Perform molecular dynamics simulation

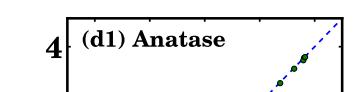
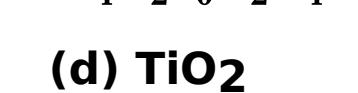
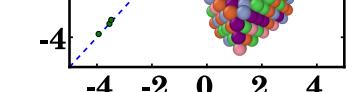
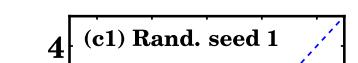
(a) small molecules



(b) $\text{MoS}_2 + \text{Pt}$



(c) CoCrFeMnNi HEA



Why now? What has changed? What has not?

The screenshot shows the GitHub organization page for "Deep Modeling". The page features a dark theme with a blue header bar. The logo is a stylized white "D" inside a blue square. The main title "Deep Modeling" is in white. Below the title, there are navigation links: "Repositories" (23), "Packages", "People" (9), and "Projects". A horizontal line separates this from the main content. The section "Pinned repositories" is labeled in white. It contains six repository cards, each with a small icon, the repository name in blue, a brief description, and statistics (language, stars, forks). The repositories are:

- deepmd-kit**: A deep learning package for many-body potential energy representation and molecular dynamics. (C++, 510 stars, 178 forks)
- dpgen**: The deep potential generator. (Python, 94 stars, 75 forks)
- dpdata**: Manipulating DeePMD-kit, VASP, LAMMPS data formats. (Python, 56 stars, 46 forks)
- deepks-kit**: a package for developing machine learning-based chemically accurate energy and density functional models. (Python, 45 stars, 7 forks)
- rid-kit**: Reinforced dynamics. (Python, 6 stars, 6 forks)
- abacus-develop**: Forked from abacussmodeling/abacus-develop. An electronic structure package based on either plane wave basis or numerical atomic orbitals. (C++, 14 stars, 22 forks)

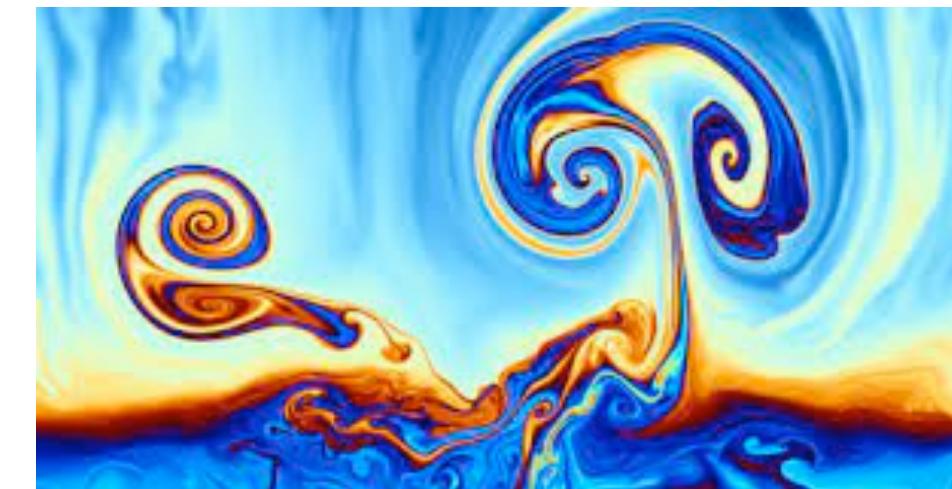
The DeepModeling Manifesto, <https://github.com/deepmodeling/community>

Machine learning **accelerated** simulation

Monte Carlo simulation

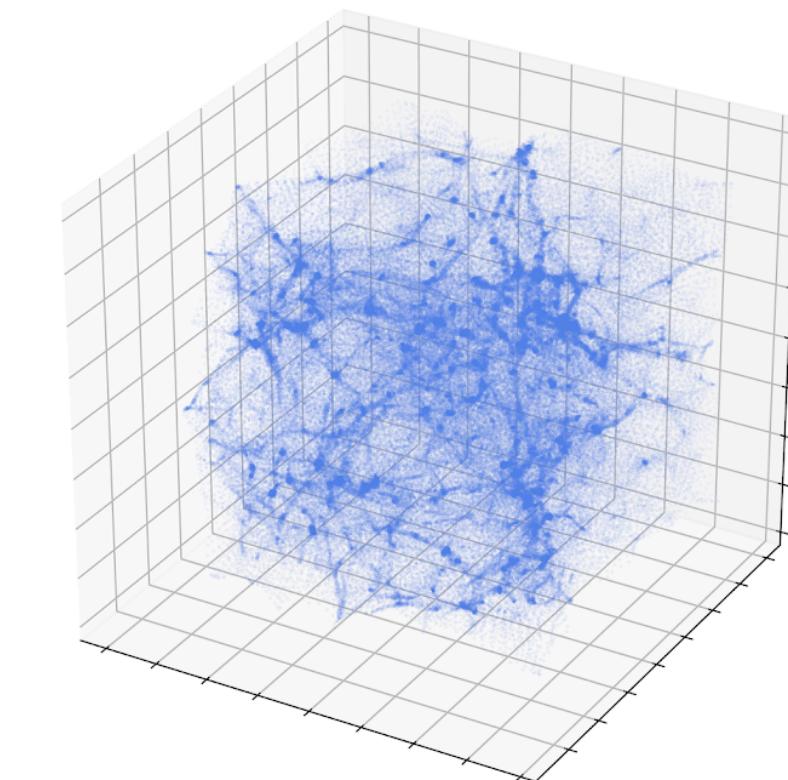


Fluid simulation



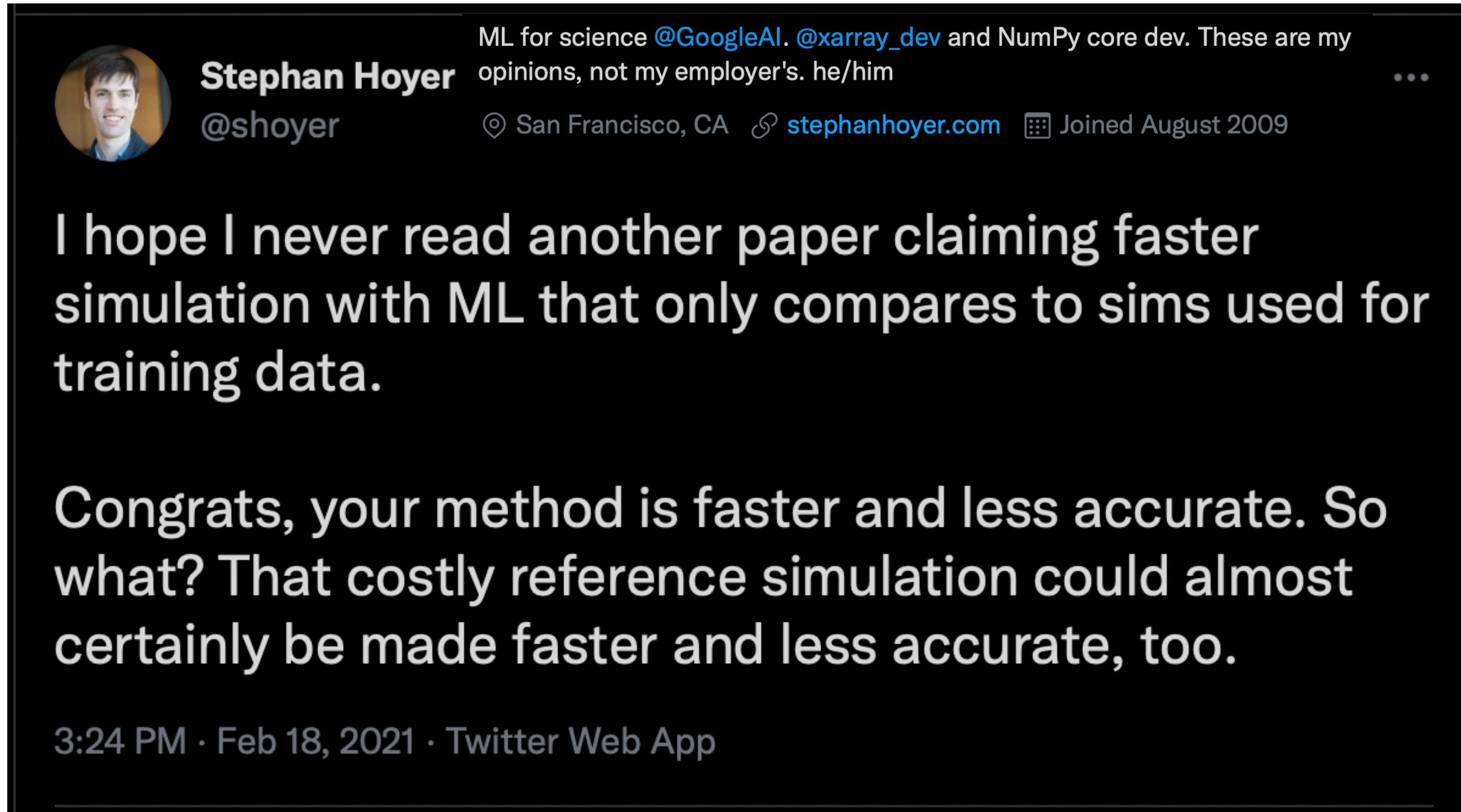
Kochkov, PNAS '21

N-body simulation



[https://blog.tensorflow.org/2020/03/
simulating-universe-in-tensorflow.html](https://blog.tensorflow.org/2020/03/simulating-universe-in-tensorflow.html)

CAUTION!



A screenshot of a Twitter post from Stephan Hoyer (@shoyer). The post features a dark background with white text. It includes a profile picture of Stephan Hoyer, his name and handle (@shoyer), a bio about ML for science, his location (San Francisco, CA), his website (stephanoyer.com), and the date he joined (August 2009). The main text of the post reads: "I hope I never read another paper claiming faster simulation with ML that only compares to sims used for training data." Below this, a second block of text reads: "Congrats, your method is faster and less accurate. So what? That costly reference simulation could almost certainly be made faster and less accurate, too." At the bottom, the timestamp "3:24 PM · Feb 18, 2021 · Twitter Web App" is visible.

ML for science @GoogleAI. @xarray_dev and NumPy core dev. These are my opinions, not my employer's. he/him

...

© San Francisco, CA ⌂ stephanoyer.com Joined August 2009

I hope I never read another paper claiming faster simulation with ML that only compares to sims used for training data.

Congrats, your method is faster and less accurate. So what? That costly reference simulation could almost certainly be made faster and less accurate, too.

3:24 PM · Feb 18, 2021 · Twitter Web App

Cheng et al, Nature 2020

Article

Evidence for supercritical behaviour of high-pressure liquid hydrogen

Even for DFT simulations, which offer a balance between computational cost and efficiency, the systems studied are limited to sizes of few hundreds of atoms and timescales of a few picoseconds^{16–18,21,22}. Given the subtlety of phase-transition phenomena, it is important to overcome the size and timescale limitations, as well as to elucidate the effect of the details of the electronic-structure methods on the location of the LLT, the melting line and the stabilities of the different solid phases^{17,18,23}.

To address these issues, we constructed three sets of machine-learning potentials (MLPs), using the Behler–Parrinello

Karasiev et al, Nature 2021

Matters arising

On the liquid–liquid phase transition of dense hydrogen

significantly larger systems. If the MLP-MD represents the underlying theory (*ab initio* MD) faithfully and if the diagnosis from the MLP-MD of problems in the earlier DFT-MD results is correct, results from the two simulation types should match. To test this, we have carried out much larger, longer DFT-MD calculations. The results are consistent with earlier DFT-MD calculations, and hence qualitatively different from the MLP-MD results. Neither the large-system nor longer-run diagnosis given by the MLP-MD is sustained.

BTW, what's your opinion on this?



Stephan Hoyer @shoyer · Feb 17

...

Learning the laws of physics: 😱 😴

Learning how to solve the laws of physics: 🤔

Learning how to solve the laws of physics more efficiently than SOTA methods from scientific computing: 😍



Martin Trapp @martin_trapp · Feb 17

ML (research) Twitter:

What constitutes an exciting research question for you as opposed to a research question that you find boring or less important?

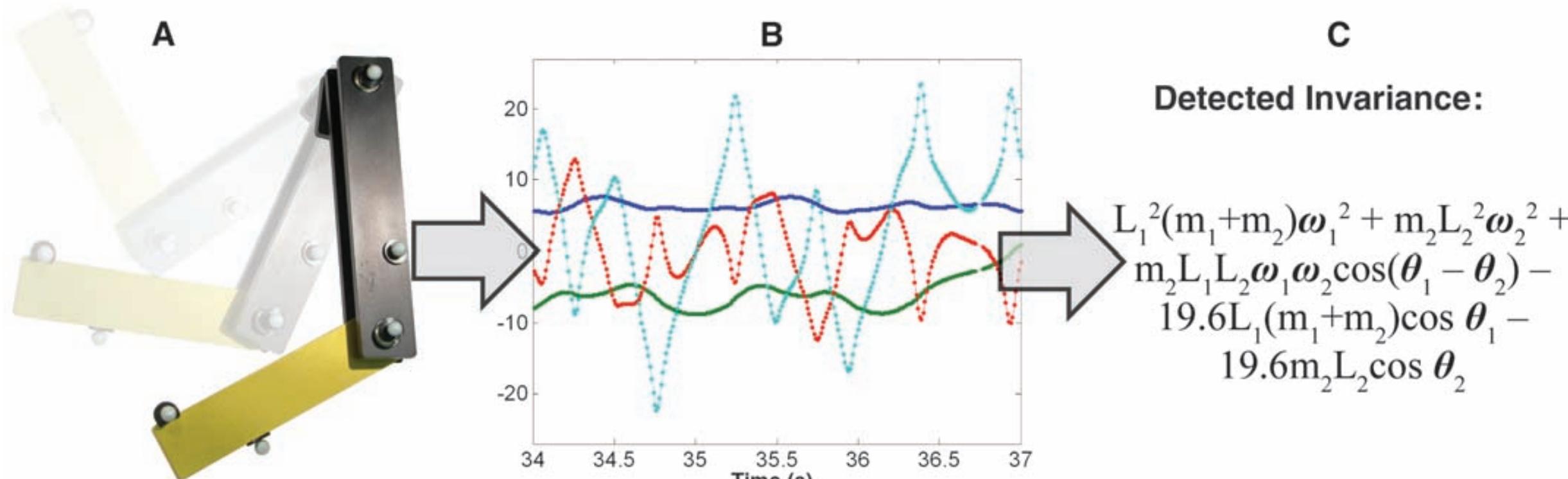


Frequently asked questions

Can machine discover physical law ?

Distilling Free-Form Natural Laws from Experimental Data

Schmidt, Lipson
Science '09



- References and Notes**
1. P. W. Anderson, *Science* **177**, 393 (1972).
 2. E. Noether, *Nachr. d. König Gesellsch. d. Wiss. zu Göttingen, Math-Phys. Klasse* 235 (1918).

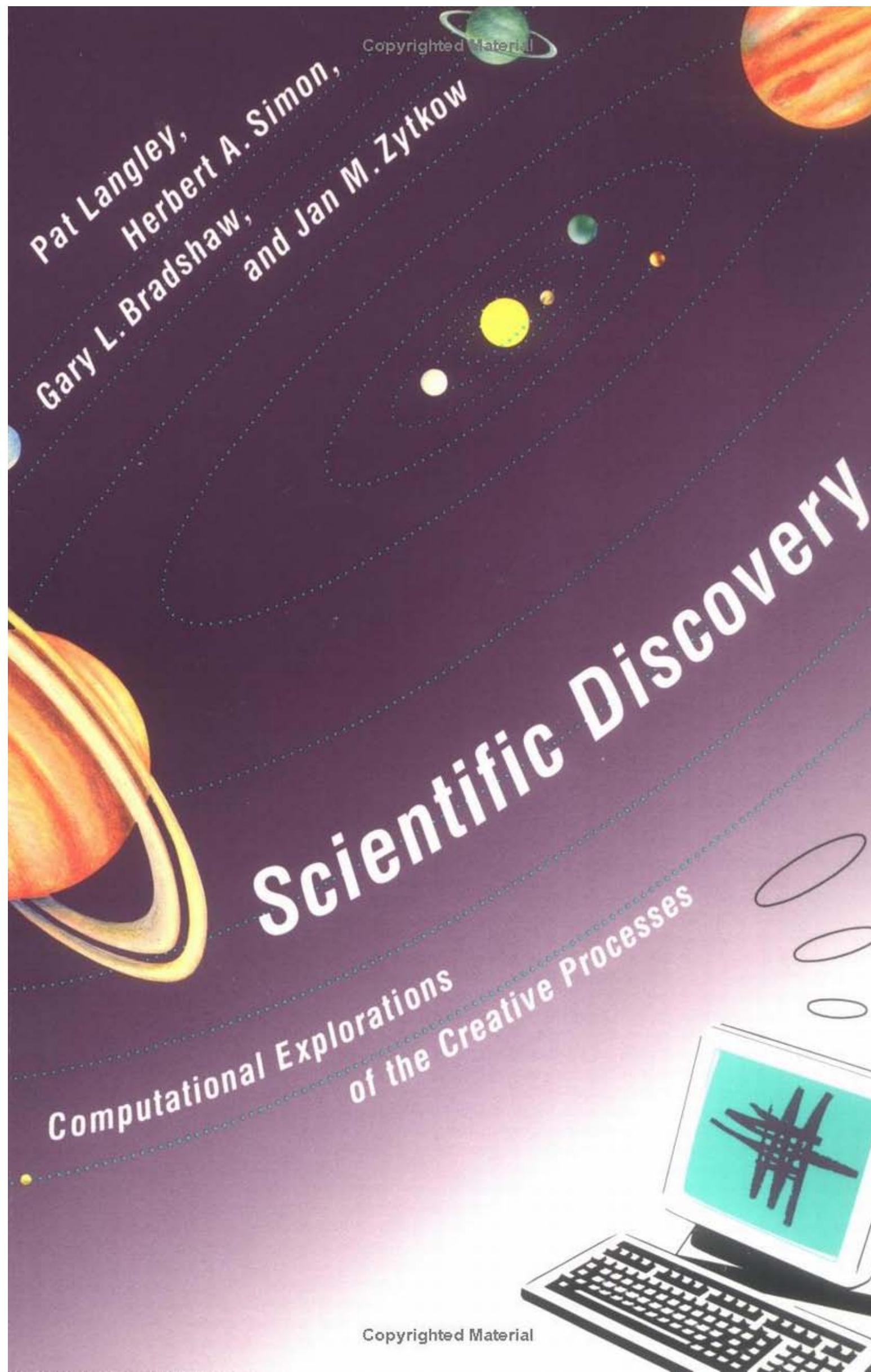
Machines Fall Short of Revolutionary Science

In the Report by Schmidt and Lipson, a machine deduces the equation behind a sample of chaotic motion. The discovery of deterministic chaos is an example of true Kuhnian revolution; others were its application to unexpected fields like meteorology and population biology. In the constrained problem in the Report, the relevant physical law and variables are known in advance; it is hardly a template for the creative, exploratory nature of true science.

PHILIP W. ANDERSON^{1*} AND ELIHU ABRAHAMS²

汤超院士在2022科学智能峰会上的讲话：“关于AI for Science的几层意思”

<https://mp.weixin.qq.com/s/oL7G7ByazbnsgrXDTToPyrw>



2204.01467

On scientific understanding with artificial intelligence

Mario Krenn,^{1, 2, 3, 4,*} Robert Pollice,^{2, 3} Si Yue Guo,² Matteo Aldeghi,^{2, 3, 4} Alba Cervera-Lierta,^{2, 3} Pascal Friederich,^{2, 3, 5} Gabriel dos Passos Gomes,^{2, 3} Florian Häse,^{2, 3, 4, 6} Adrian Jinich,⁷ AkshatKumar Nigam,^{2, 3} Zhenpeng Yao,^{2, 8, 9, 10} and Alán Aspuru-Guzik^{2, 3, 4, 11, †}

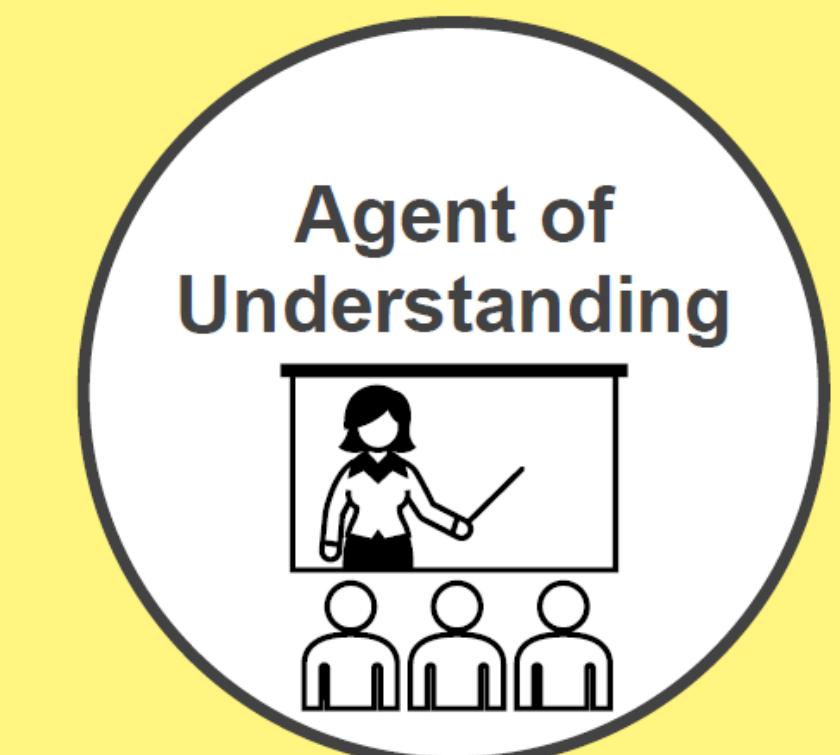
Three Dimensions of Computer-Assisted Scientific Understanding



Computational
Microscope



Resource of
Inspiration



Agent of
Understanding

Frequently asked questions

Can machines discover mathematics?



Timothy Gowers
@wtgowers

An interesting paper by Adam Wagner appeared on arXiv a couple of days ago (thanks to Imre Leader for drawing my attention to it), which uses reinforcement learning to find non-trivial counterexamples to several conjectures in graph theory. 1/
[arxiv.org/pdf/2104.14516...](https://arxiv.org/pdf/2104.14516.pdf)

...

Nature 600, 70 (2021)

Advancing mathematics by guiding human intuition with AI

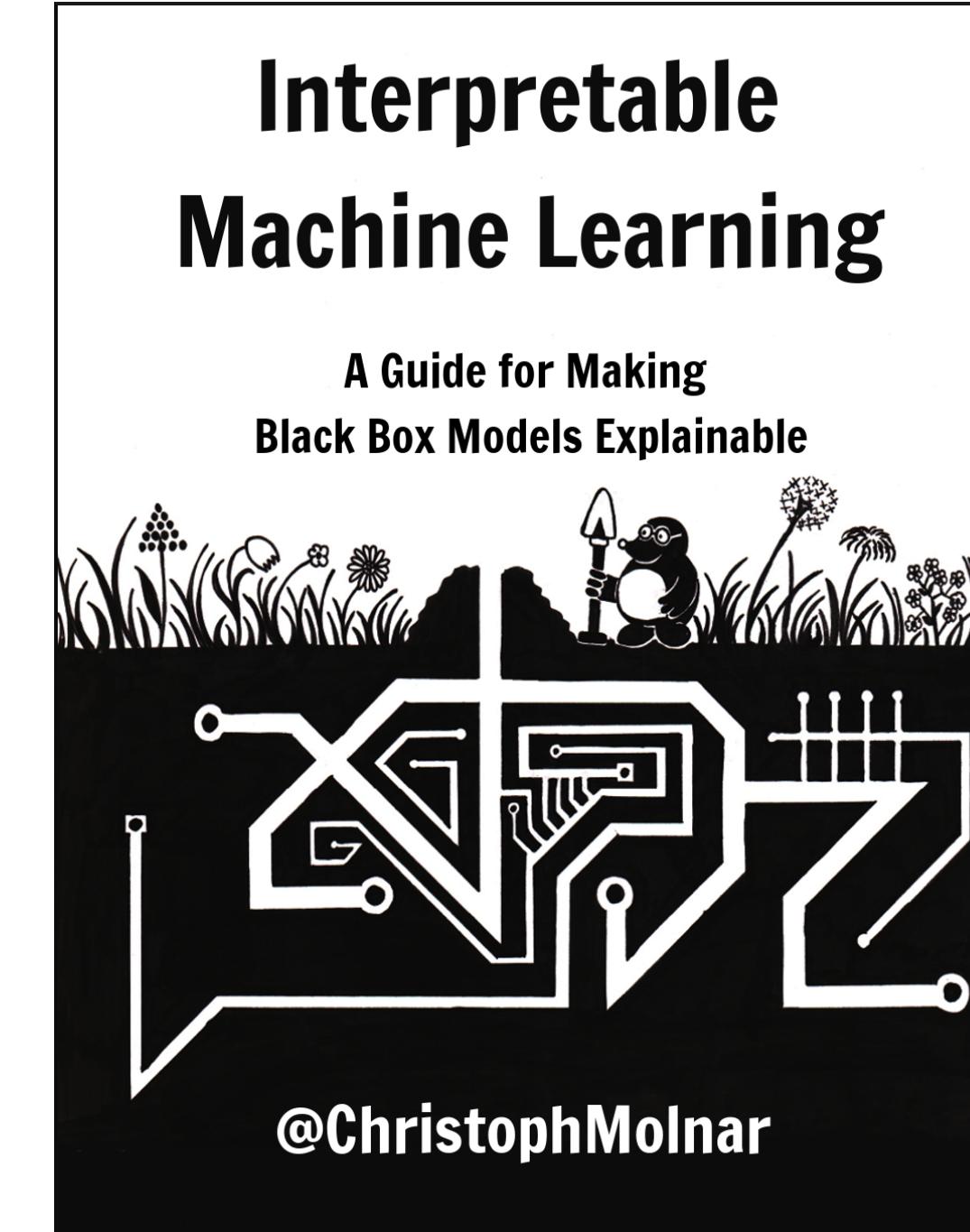
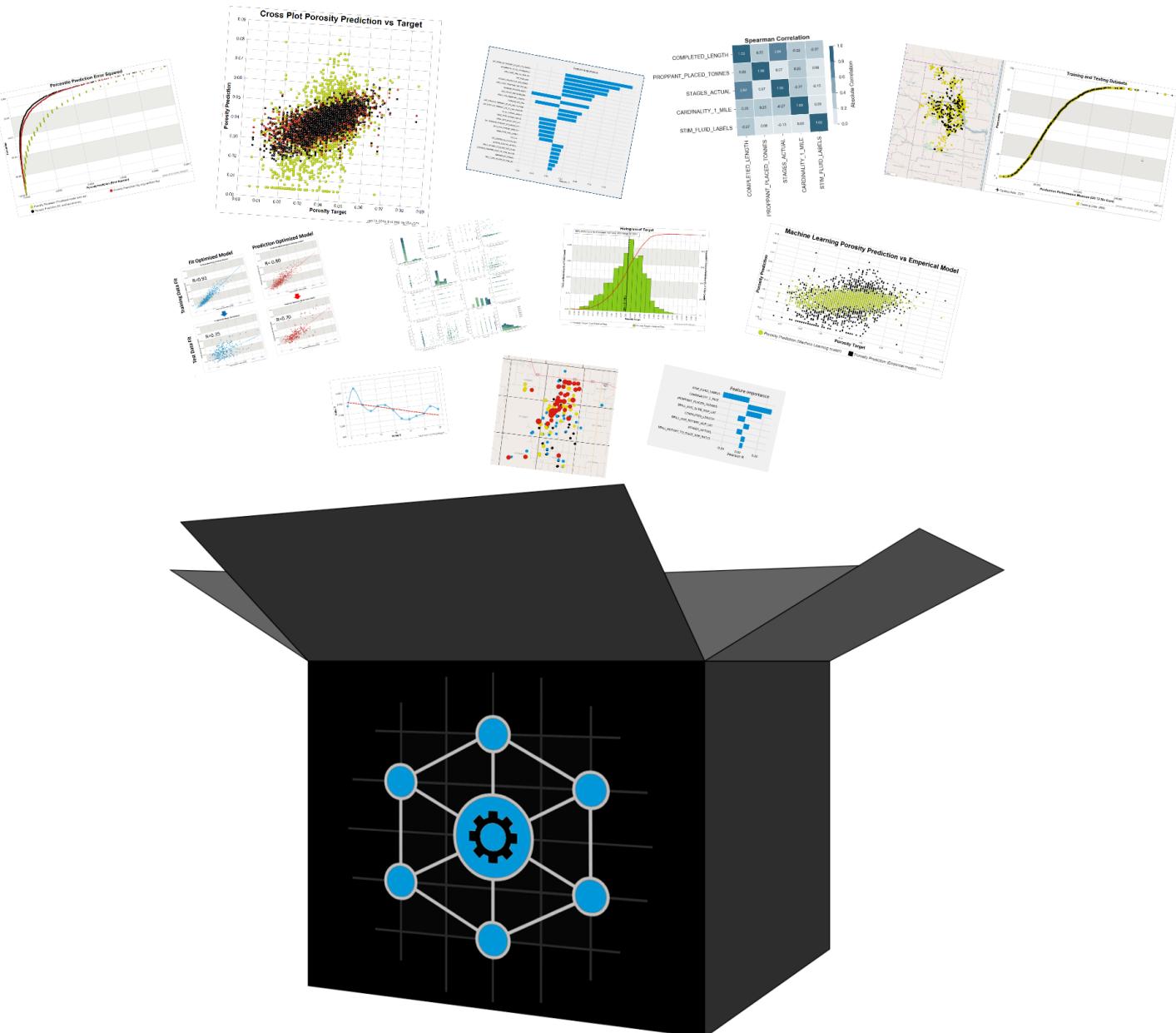
[Alex Davies](#)✉, [Petar Veličković](#), [Lars Buesing](#), [Sam Blackwell](#), [Daniel Zheng](#), [Nenad Tomašev](#), [Richard Tanburn](#), [Peter Battaglia](#), [Charles Blundell](#), [András Juhász](#), [Marc Lackenby](#), [Geordie Williamson](#), [Demis Hassabis](#) & [Pushmeet Kohli](#)✉

Search counter-examples to reject conjectures

Guide human mathematician to propose conjectures

Frequently asked questions

Do we understand what is the machine doing ?



Yes/No/Well, do I have to ?/I don't care...

Frequently asked questions

Are data and compute all you need ?

The screenshot shows a web browser window with two tabs. The active tab displays the title "The Bitter Lesson" by Rich Sutton, dated March 13, 2019. The content discusses the general methods that leverage computation as the most effective approach. A large portion of the text is highlighted in blue, emphasizing the sentence: "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."

<http://www.incompleteideas.net/IncompleteIdeas/BitterLesson.html>

The screenshot shows a web browser window with two tabs. The active tab displays the title "Do we still need models or just more data and compute?" by Max Welling, dated April 20, 2019. The content discusses the polarizing views in the AI community regarding the need for models versus computation. It quotes Rich Sutton's blog and adds context about the debate between symbolic AI and statistical AI, and between model-driven and data-driven AI. The author, Max Welling, expresses his belief in the power of computation and deep learning.

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.

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

Deep learning: Sternstunden der Menschheit



2012

AlexNet

ImageNet Classification with Deep Convolutional Neural Networks

Alex Krizhevsky
University of Toronto

Ilya Sutskever
University of Toronto

Geoffrey E. Hinton
University of Toronto



2016

AlphaGo

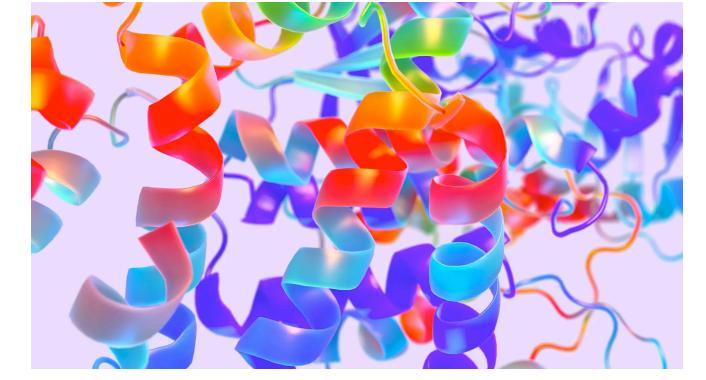
Mastering the game of Go with deep neural networks and tree search

David Silver✉, Aja Huang, Chris J. Maddison, Arthur Guez, Laurent Sifre, George van den Driessche, Julian Schrittwieser, Ioannis Antonoglou, Veda Panneershelvam, Marc Lanctot, Sander Dieleman, Dominik Grewe, John Nham, Nal Kalchbrenner, Ilya Sutskever, Timothy Lillicrap, Madeleine Leach, Koray Kavukcuoglu, Thore Graepel & Demis Hassabis✉



2020

GPT3



2021

AlphaFold2

Language Models are Few-Shot Learners

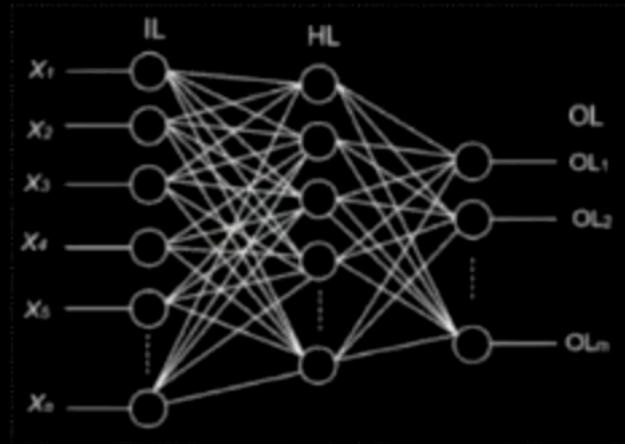
Tom B. Brown, Benjamin Mann, Nick Ryder, Melanie Subbiah, Jared Kaplan, Prafulla Dhariwal, Arvind Neelakantan, Pranav Shyam, Girish Sastry, Amanda Askell, Sandhini Agarwal, Ariel Herbert-Voss, Gretchen Krueger, Tom Henighan, Rewon Child, Aditya Ramesh, Daniel M. Ziegler, Jeffrey Wu, Clemens Winter, Christopher Hesse, Mark Chen, Eric Sigler, Mateusz Litwin, Scott Gray, Benjamin Chess, Jack Clark, Christopher Berner, Sam McCandlish, Alec Radford, Ilya Sutskever, Dario Amodei

The unreasonable effectiveness of deep learning 2022

<https://www.youtube.com/watch?v=BjyZcSiVg5A>

The deep learning formula

Huge neural network



Huge dataset

A screenshot of a terminal window displaying a large amount of binary data. The data consists of several lines of binary digits (0s and 1s), representing a dataset. The terminal window has a dark background with white text.

Huge compute



Ilya Sutskever

Co-founder and Chief Scientist of OpenAI

jaykmody.com

1 https://chat.openai.com | The Bitter Lesson | GPT in 60 Lines... | https://staff.fn...

Jay Mody | Blog | LinkedIn | GitHub

GPT in 60 Lines of NumPy

January 30, 2023

In this post, we'll implement a GPT from scratch in just [60 lines of numpy](#). We'll then load the trained GPT-2 model weights released by OpenAI into our implementation and generate some text.

Note:

- This post assumes familiarity with Python, NumPy, and some basic experience training neural networks.
- This implementation is missing tons of features on purpose to keep it as simple as possible while remaining complete. The goal is to provide a **simple yet complete technical introduction to the GPT as an educational tool**.
- Understanding the GPT architecture is just a small (but vital) piece of the larger LLM puzzle. [\[1\]](#).
- All the code for this blog post can be found at github.com/jaymody/picoGPT.
- [Hacker news thread](#)

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

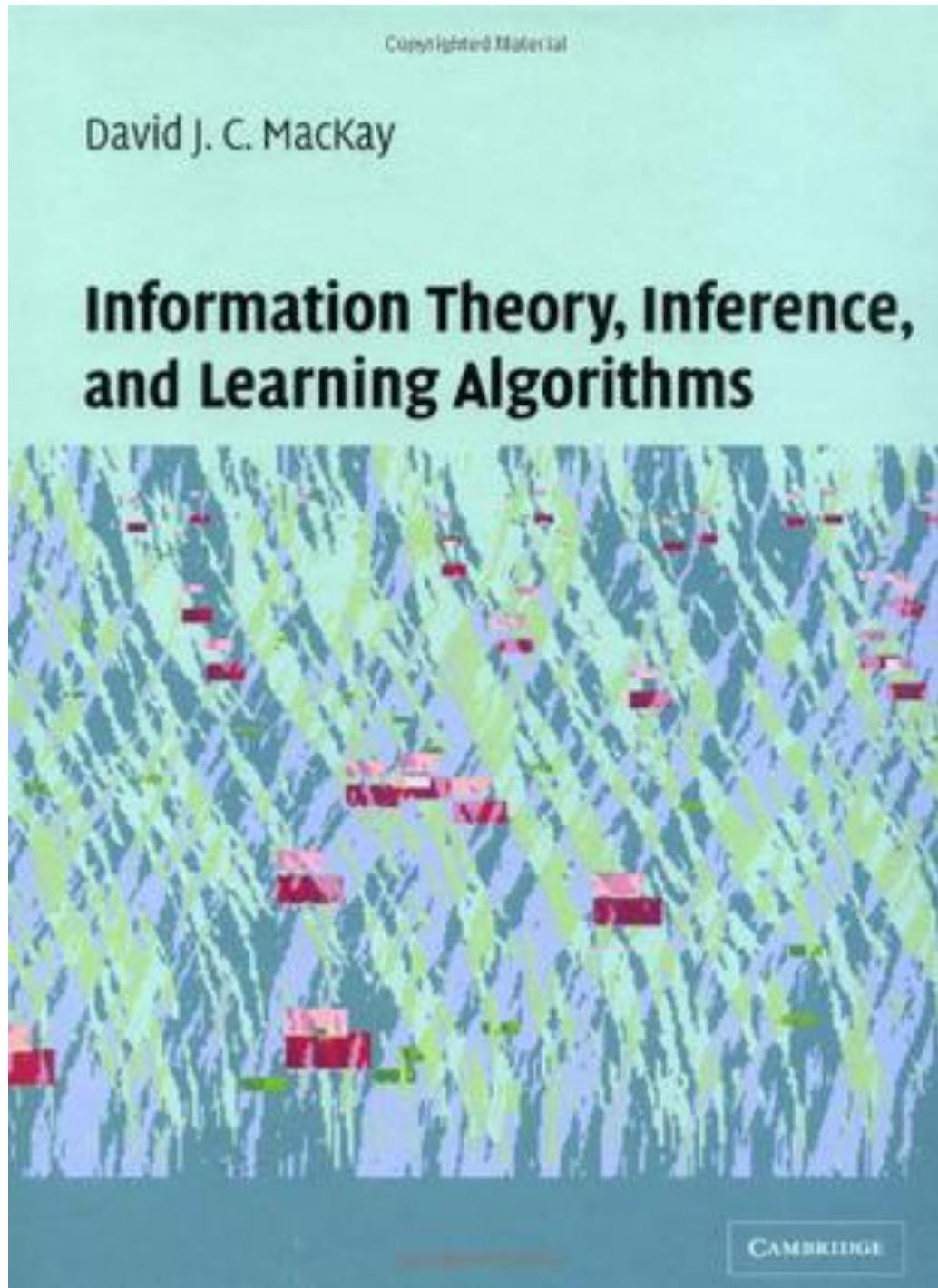
GPT-3: 175B parameters ~1000 GPUs train for months

<p>Supplementary Information for: Highly accurate protein structure prediction with AlphaFold</p> <p>John Jumper^{1,2*}, Richard Figsenkov^{1*}, Olaf Ronneberg¹, Augustin Žídek^{1*}, Anna Korshenina¹, A A Kohl^{1*}, Andrew J Baker¹, Stanislav Nikolov^{1*}, Richard Pacholska¹, David Reiman¹, Ellen C Williams¹, Tamas Bertalan¹, Vinyals¹, Andrew W Senior¹</p> <p>¹School of Biological Sciences, University College London, London, UK *Correspondence to: j.jumper@ucl.ac.uk</p> <p>Contents</p> <p>1 Supplementary Methods</p> <ul style="list-style-type: none"> 1.1 Notation 1.2 Data pipeline 1.2.1 Parsing 1.2.2 Genetic search 1.2.3 Template search 1.2.4 Training data 1.2.5 Filtering 1.2.6 MSA block deletion 1.2.7 MSA clustering 1.2.8 Residue cropping 1.2.9 Featureization and embeddings 1.3 Self-distillation dataset 1.4 AlphaFold Inference 1.5 Input embeddings 1.6 Evoformer blocks 1.6.1 MSA row-wise gated self-attention 1.6.2 MSA column-wise gated self-attention 1.6.3 MSA transition 1.6.4 Outer product mean 1.6.5 Triangular multiplicative update 1.6.6 Triangular self-attention 1.6.7 Transition in the pair stack 1.7 Additional inputs 1.7.1 Template stack 1.7.2 Unclustered MSA stack 1.8 Structure module 1.8.1 Construction of frames from templates 1.8.2 Invariant point attention (IPA) 1.8.3 Backbone update 1.8.4 Compute all atom coordinates 1.8.5 Rename symmetric ground states 1.8.6 Amber relaxation 1.9 Loss functions and auxiliary heads 1.9.1 Side chain and backbone torsions 1.9.2 Frame aligned point error (FAPE) 1.9.3 Chiral properties of AlphaFold 1.9.4 Configurations with FAPE 1.9.5 Metric properties of FAPE 1.9.6 Model confidence prediction 1.9.7 TM-score prediction 1.9.8 Distogram prediction 1.9.9 Masked MSA prediction 1.9.10 “Experimentally resolved” 1.9.11 Structural violations 1.10 Recycling iterations 1.11 Training and inference details 1.11.1 Training stages 1.11.2 MSA resampling and ensemble 1.11.3 Optimization details 1.11.4 Parameters initialization 1.11.5 Loss clamping details 1.11.6 Dropout details 1.11.7 Evaluator setup 1.11.8 Reducing the memory consumption 1.12 CASP14 assessment 1.12.1 Training procedure 1.12.2 Inference and scoring 1.13 Ablation studies 1.13.1 Architectural details 1.13.2 Procedure 1.13.3 Results 1.14 Network probing details 1.15 Novel fold performance 1.16 Visualization of attention 1.17 Additional results 	<p><i>Suppl. Material for Jumper et al. (2021): Highly accurate protein structure prediction with AlphaFold</i> 2</p> <p>1.6.2 MSA column-wise gated self-attention 15 1.6.3 MSA transition 16 1.6.4 Outer product mean 17 1.6.5 Triangular multiplicative update 18 1.6.6 Triangular self-attention 19 1.6.7 Transition in the pair stack 20 1.7 Additional inputs 21 1.7.1 Template stack 22 1.7.2 Unclustered MSA stack 23 1.8 Structure module 24 1.8.1 Construction of frames from templates 25 1.8.2 Invariant point attention (IPA) 26 1.8.3 Backbone update 27 1.8.4 Compute all atom coordinates 28 1.8.5 Rename symmetric ground states 29 1.8.6 Amber relaxation 30 1.9 Loss functions and auxiliary heads 31 1.9.1 Side chain and backbone torsions 32 1.9.2 Frame aligned point error (FAPE) 33 1.9.3 Chiral properties of AlphaFold 34 1.9.4 Configurations with FAPE 35 1.9.5 Metric properties of FAPE 36 1.9.6 Model confidence prediction 37 1.9.7 TM-score prediction 38 1.9.8 Distogram prediction 39 1.9.9 Masked MSA prediction 40 1.9.10 “Experimentally resolved” 41 1.9.11 Structural violations 42 1.10 Recycling iterations 43 1.11 Training and inference details 44 1.11.1 Training stages 45 1.11.2 MSA resampling and ensemble 46 1.11.3 Optimization details 47 1.11.4 Parameters initialization 48 1.11.5 Loss clamping details 49 1.11.6 Dropout details 50 1.11.7 Evaluator setup 51 1.11.8 Reducing the memory consumption 52 1.12 CASP14 assessment 53 1.12.1 Training procedure 54 1.12.2 Inference and scoring 55 1.13 Ablation studies 56 1.13.1 Architectural details 57 1.13.2 Procedure 58 1.13.3 Results 59 1.14 Network probing details 60 1.15 Novel fold performance 61 1.16 Visualization of attention 62 1.17 Additional results 63</p> <p>List of Supplementary Figures</p> <p>1 Input feature embeddings 11 2 MSA row-wise gated self-attention with pair bias 15 3 MSA column-wise gated self-attention 16 4 MSA transition layer 16 5 Outer product mean 17 6 Triangular multiplicative update using “outgoing” edges 18 7 Triangular self-attention around starting node 19 8 Invariant Point Attention Module 27 9 Accuracy distribution of a model trained with dRMSD instead of the FAPE loss 36 10 Accuracy of ablations depending on the MSA depth 50 11 Performance on a set of novel structures 53 12 Visualization of row-wise pair attention 54 13 Visualization of attention in the MSA along sequences 55 14 Median all-atom RMSD₉₅ on the CASP14 set 56 15 RMSD histograms on the template-reduced recent PDB set 57</p> <p>List of Supplementary Tables</p> <p>1 Input features to the model 8 2 Rigid groups for constructing all atoms from given torsion angles 24 3 Ambiguous atom names due to 180°-rotation-symmetry 31 4 AlphaFold training protocol 43 5 Training protocol for CASP14 models 46 6 Quartiles of RMSD distributions on the template-reduced recent PDB set 57</p> <p>List of Algorithms</p> <p>1 MSABlockDeletion MSA block deletion 6 2 Inference AlphaFold Model Inference 12 3 InputEmbedder Embeddings for initial representations 13 4 relpos Relative position encoding 13 5 one_hot One-hot encoding with nearest bin 13 6 EvoformerStack Evoformer stack 14 7 MSARowAttentionWithPairBias MSA row-wise gated self-attention with pair bias 15 8 MSAColumnAttention MSA column-wise gated self-attention 16 9 MSATransition Transition layer in the MSA stack 17 10 OuterProductMean Outer product mean 17 11 TriangleMultiplicationOutgoing Triangular multiplicative update using “outgoing” edges 18 12 TriangleMultiplicationIncoming Triangular multiplicative update using “incoming” edges 18 13 TriangleAttentionStartingNode Triangular gated self-attention around starting node 19 14 TriangleAttentionEndingNode Triangular gated self-attention around ending node 20 15 PairTransition Transition layer in the pair stack 20 16 TemplatePairStack Template pair stack 21 17 TemplatePointwiseAttention Template pointwise attention 21 18 ExtraMsaStack Extra MSA stack 22 19 MSAColumnGlobalAttention MSA global column-wise gated self-attention 22 20 StructureModule Structure module 25</p>
--	--

60 pages supplementary of the AlphaFold2 paper

Alphafold2: 93M parameters, 16TPUv3 (~100-200 GPUs) train for a few weeks

Frequently asked questions



Is this all fitting ?

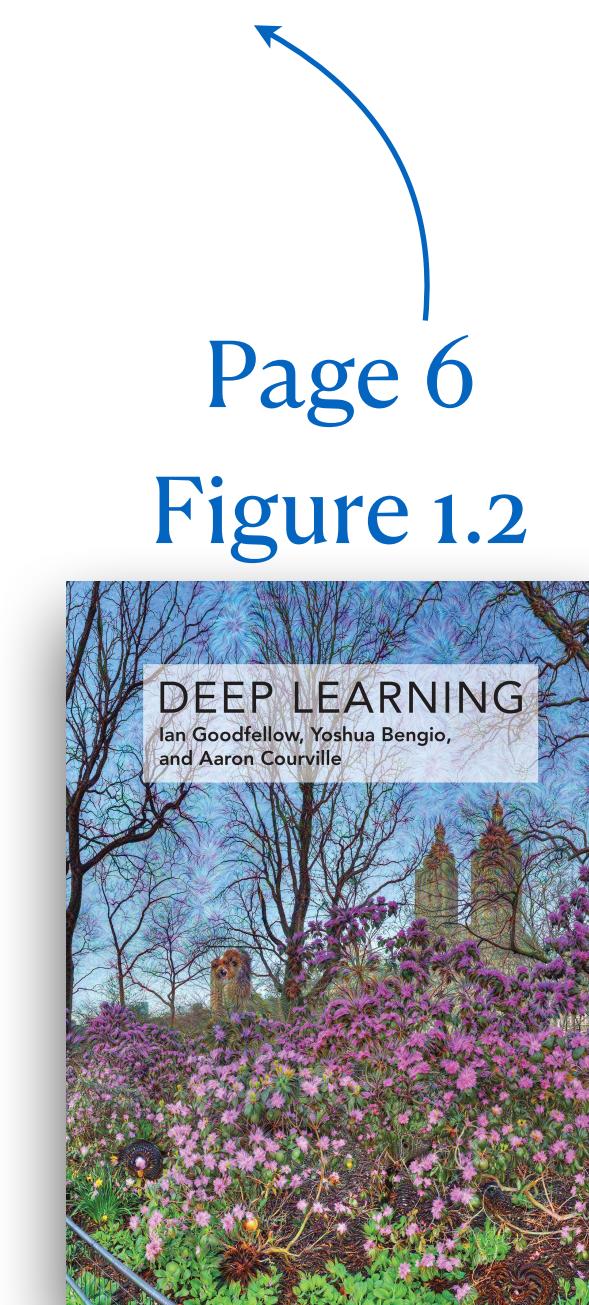
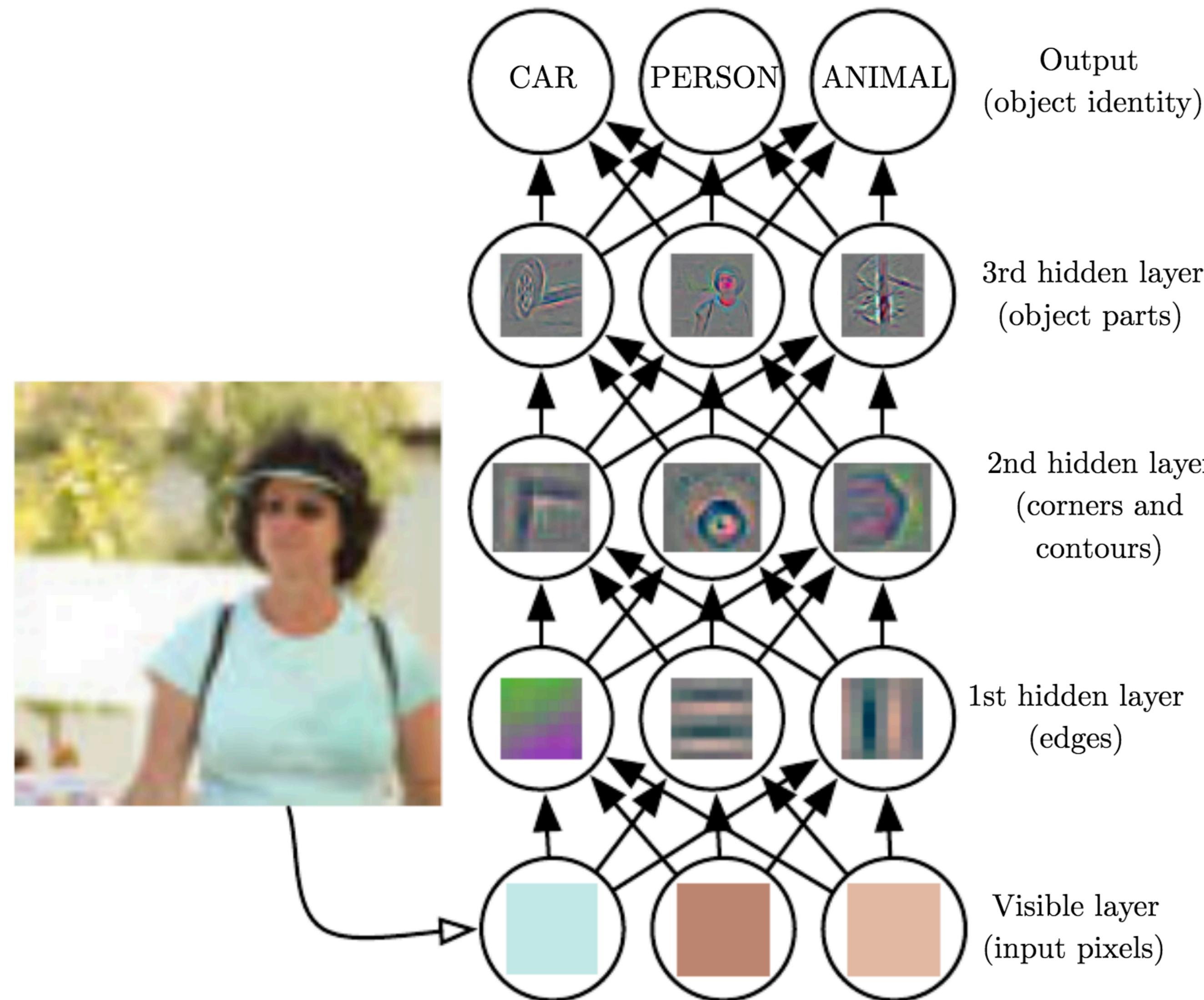
One of my students, Robert, asked:

Maybe I'm missing something fundamental, but supervised neural networks seem equivalent to fitting a pre-defined function to some given data, then extrapolating – what's the difference?

I agree with Robert. The supervised neural networks we have studied so far are simply parameterized nonlinear functions which can be fitted to data.

True for supervised learning, which is hugely successful for real-world applications.
But that is not the whole story, especially for scientific applications.

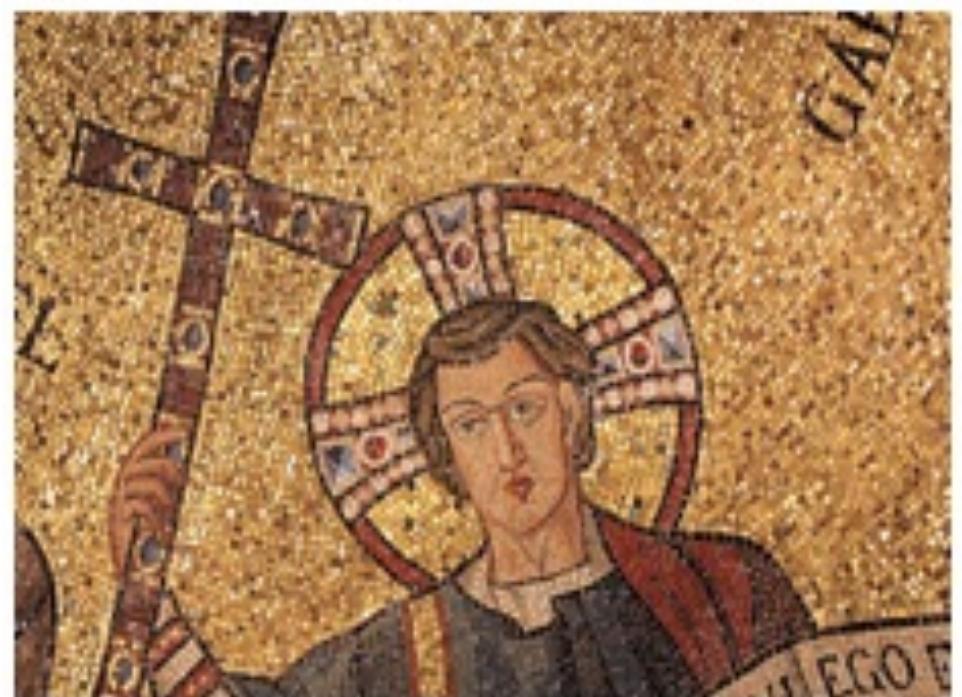
Representation Learning



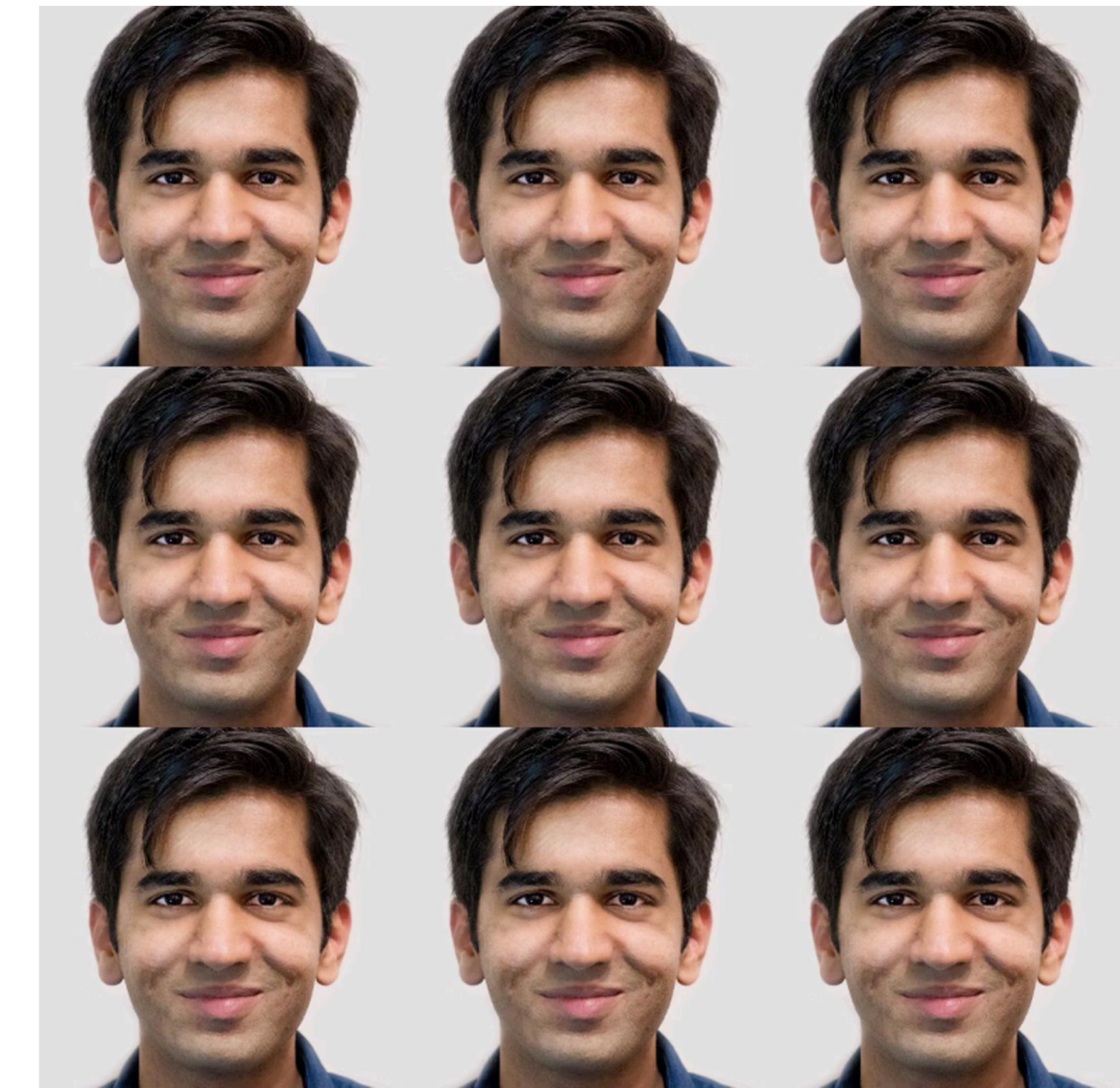
Page 6
Figure 1.2

Magic of learning representations

Neural style transfer



Latent space interpolation



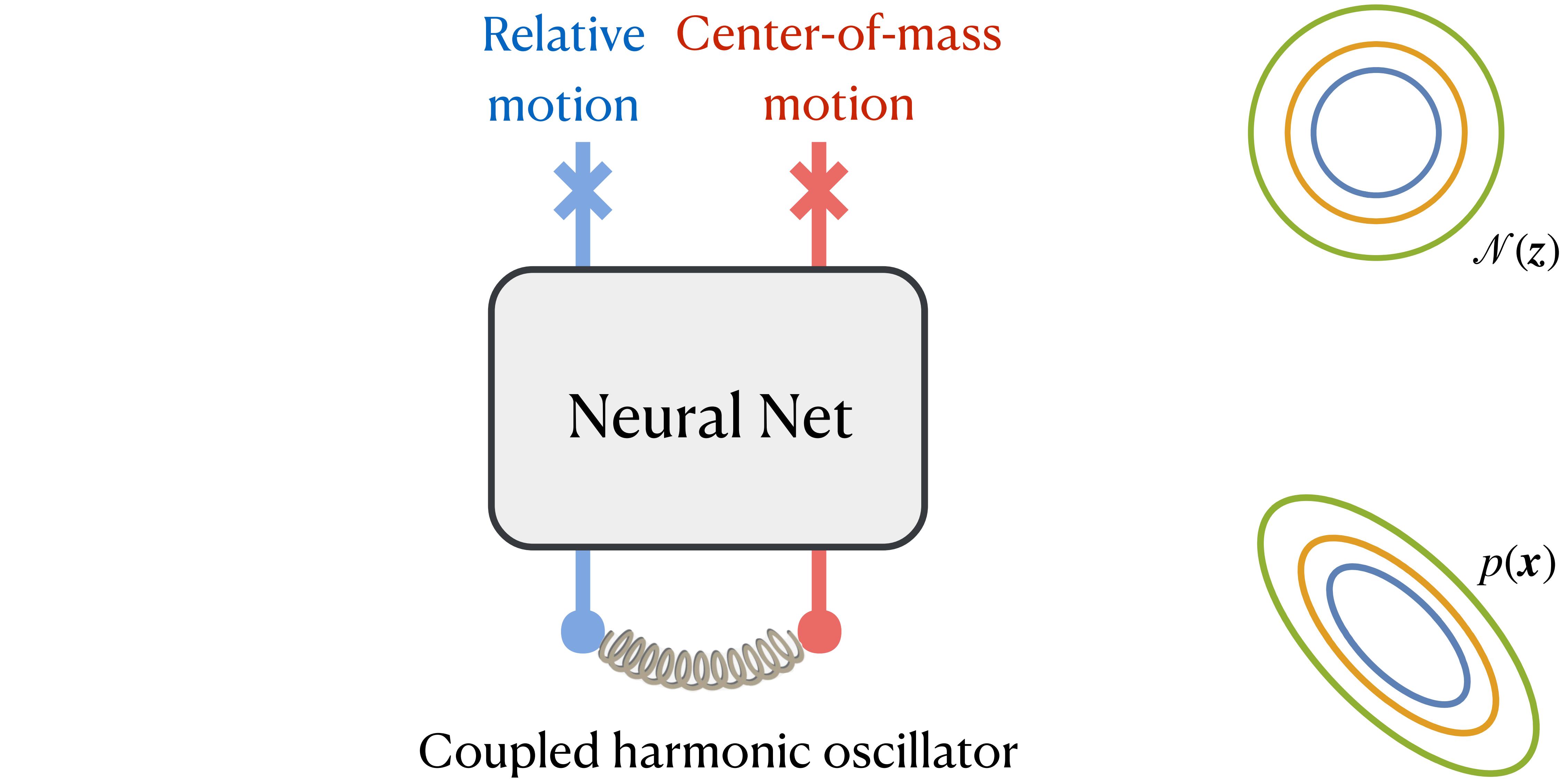
Gatys et al, 1508.06576



Glow 1807.03039

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

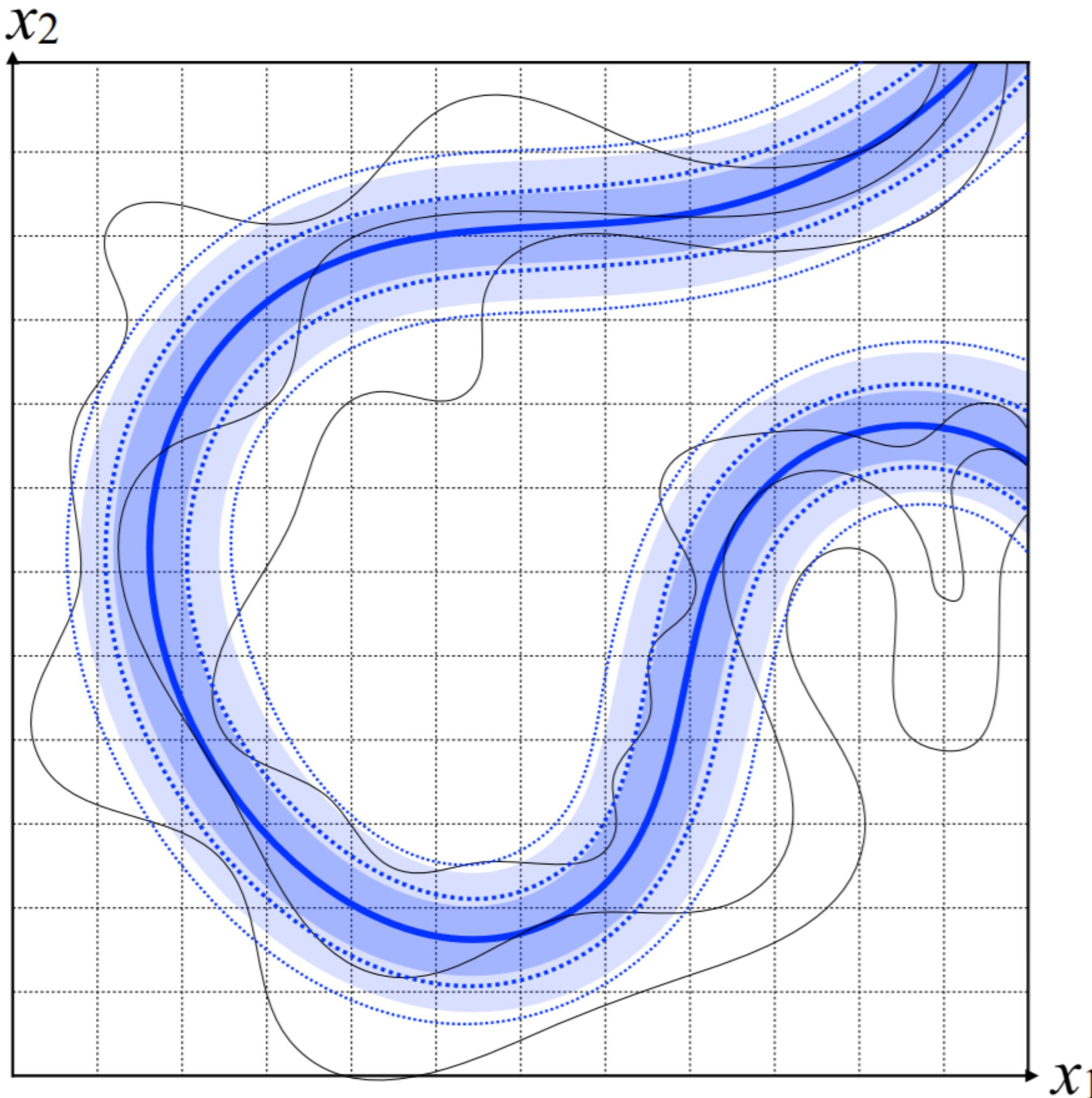
Representation in physics



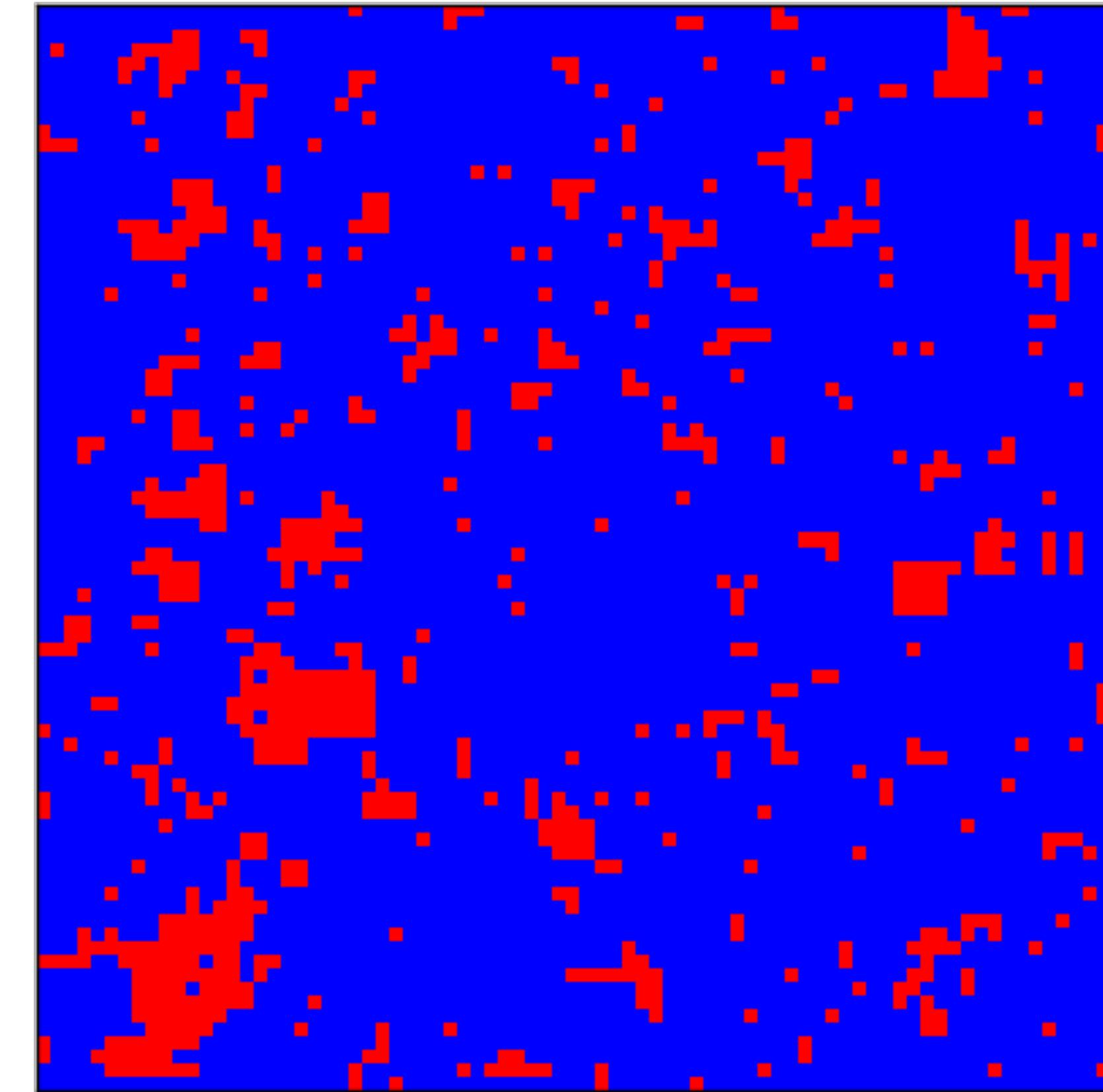
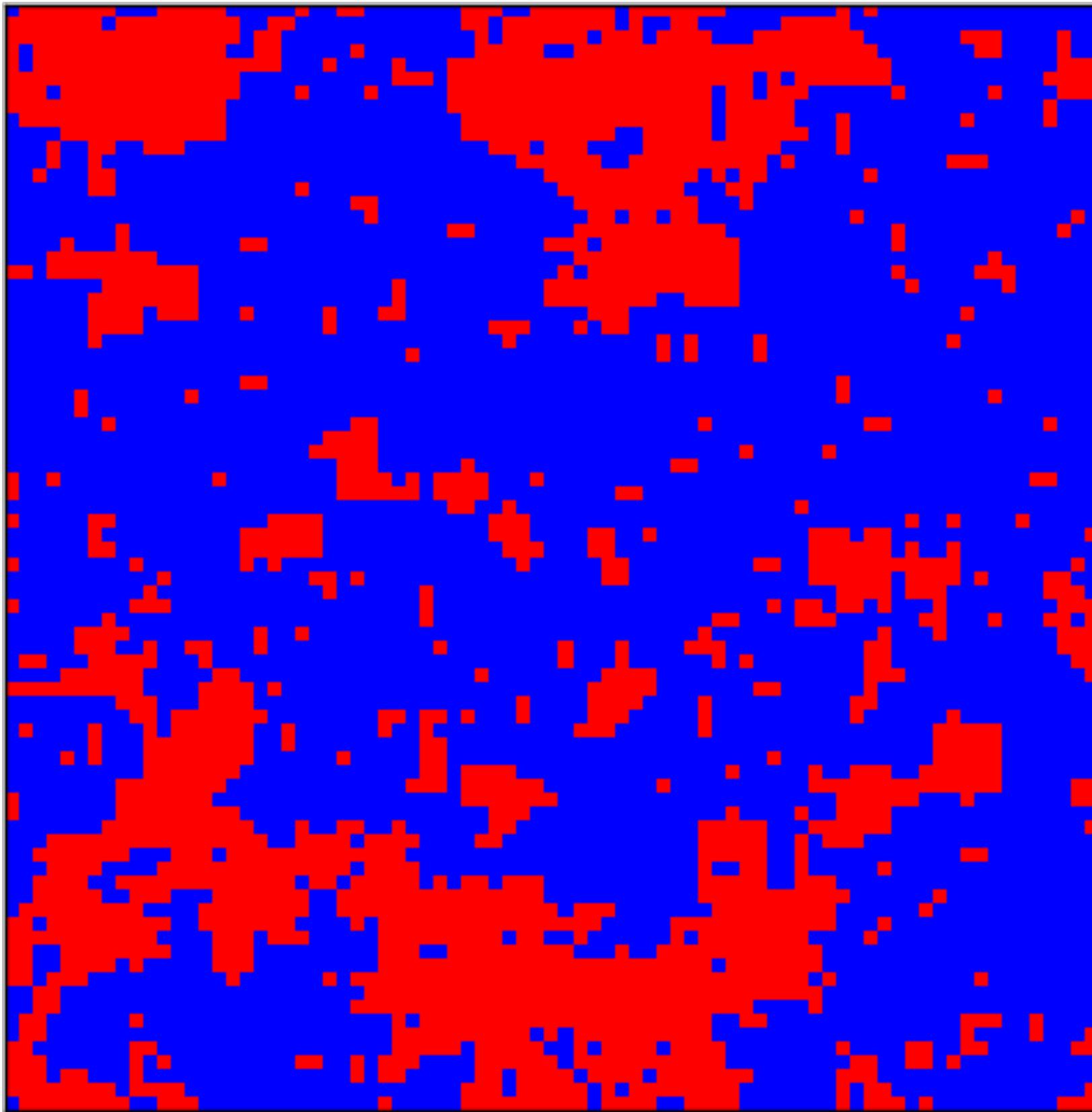
How to unfold the manifold ?



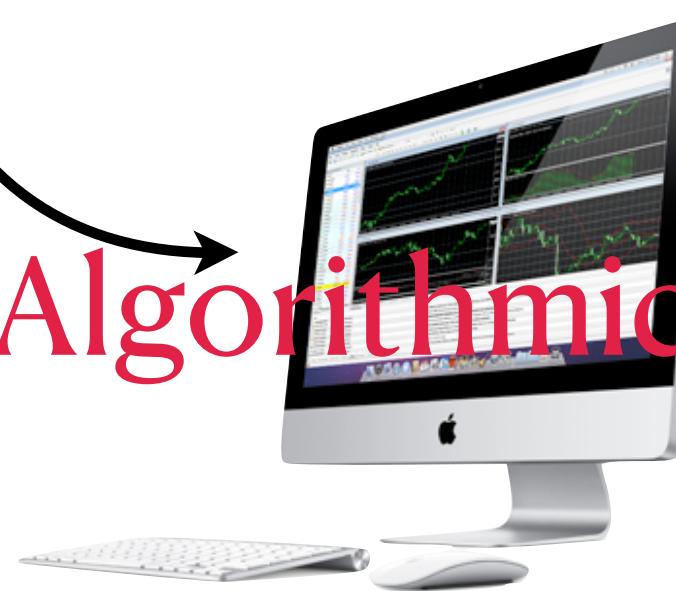
Monte Carlo sampling
of the physical d.o.f.
can be difficult since
they live on a manifold



Two different Monte Carlo update policies



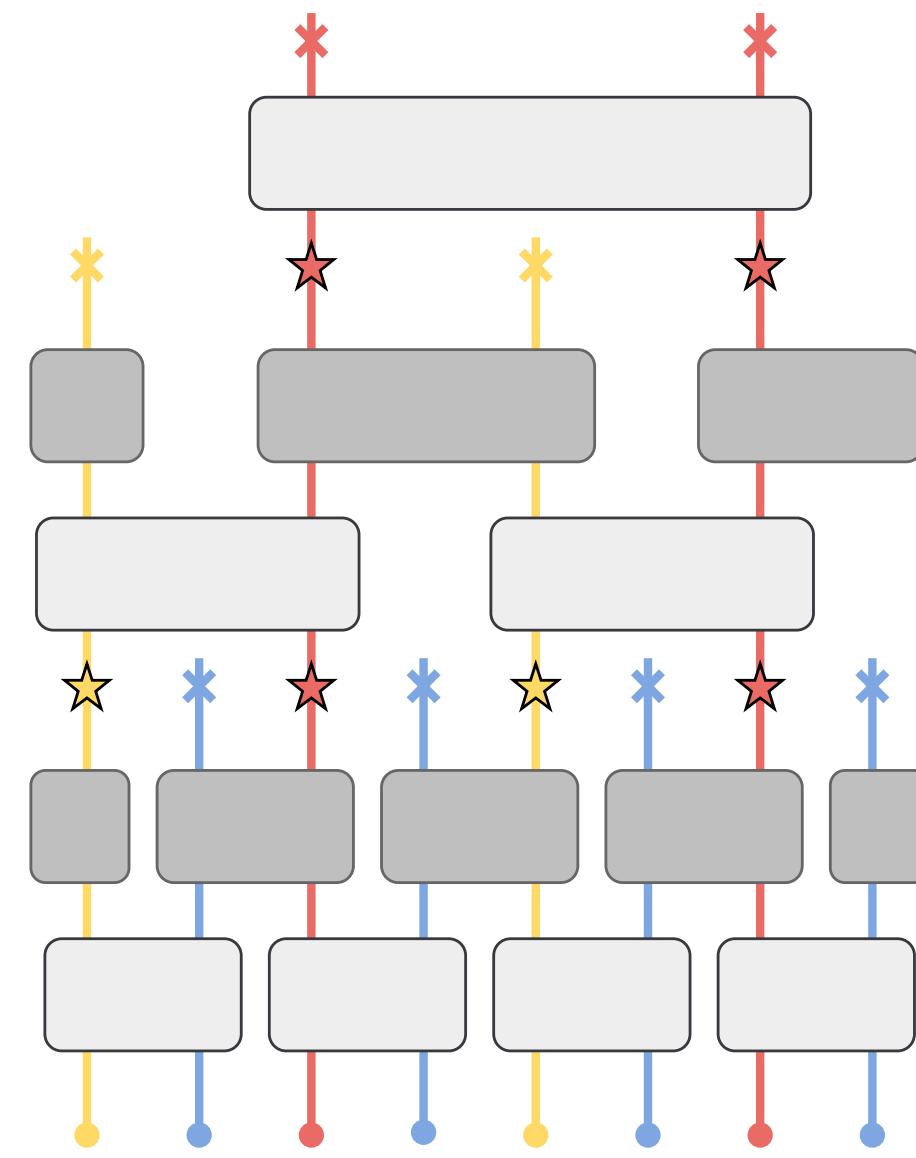
Algorithmic innovation outperforms Moore's law!



Neural Network Renormalization Group

Latent space energy function

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



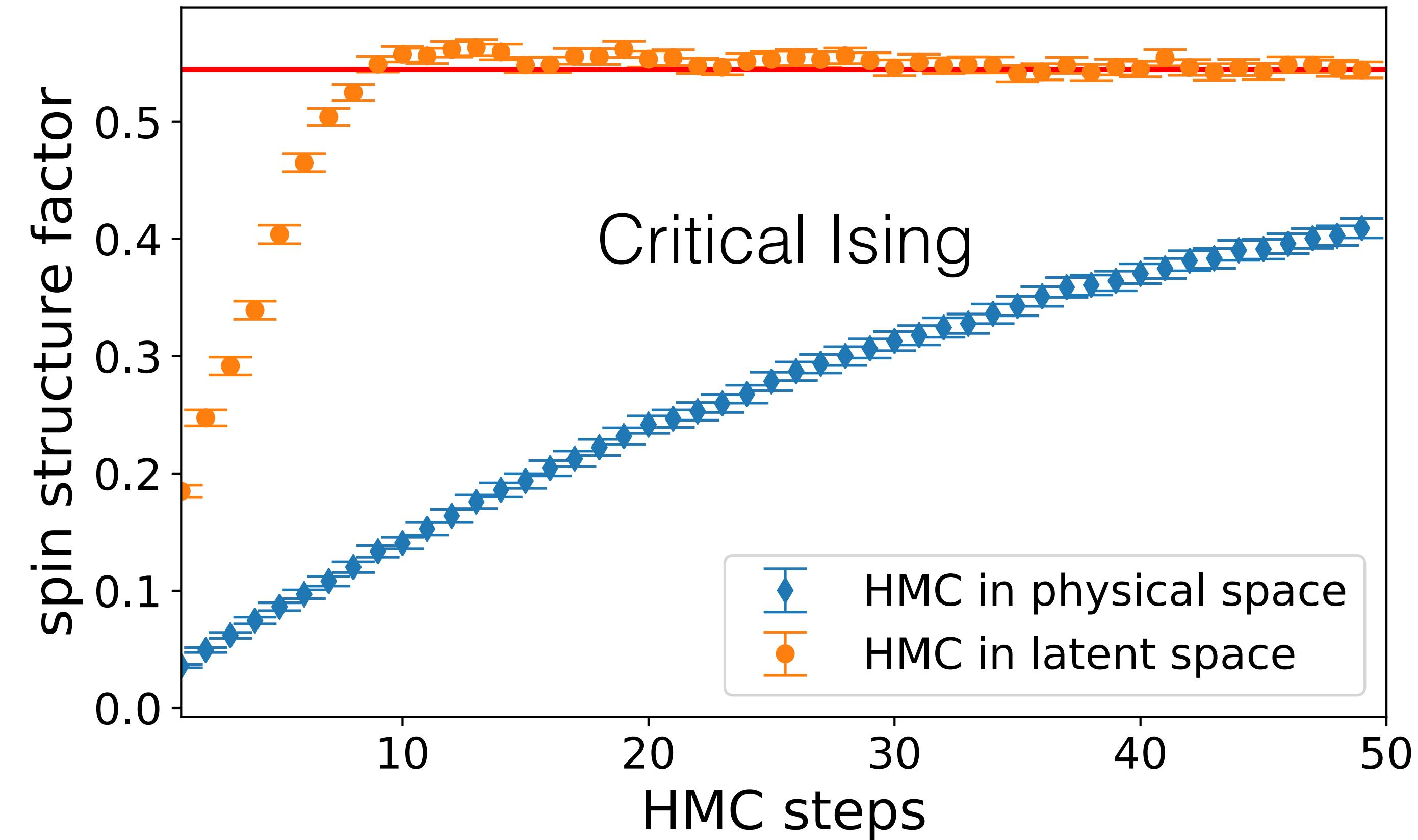
Physical energy function $E(x)$

Sampling thermalizes faster (in the unit of MC steps) in the latent space

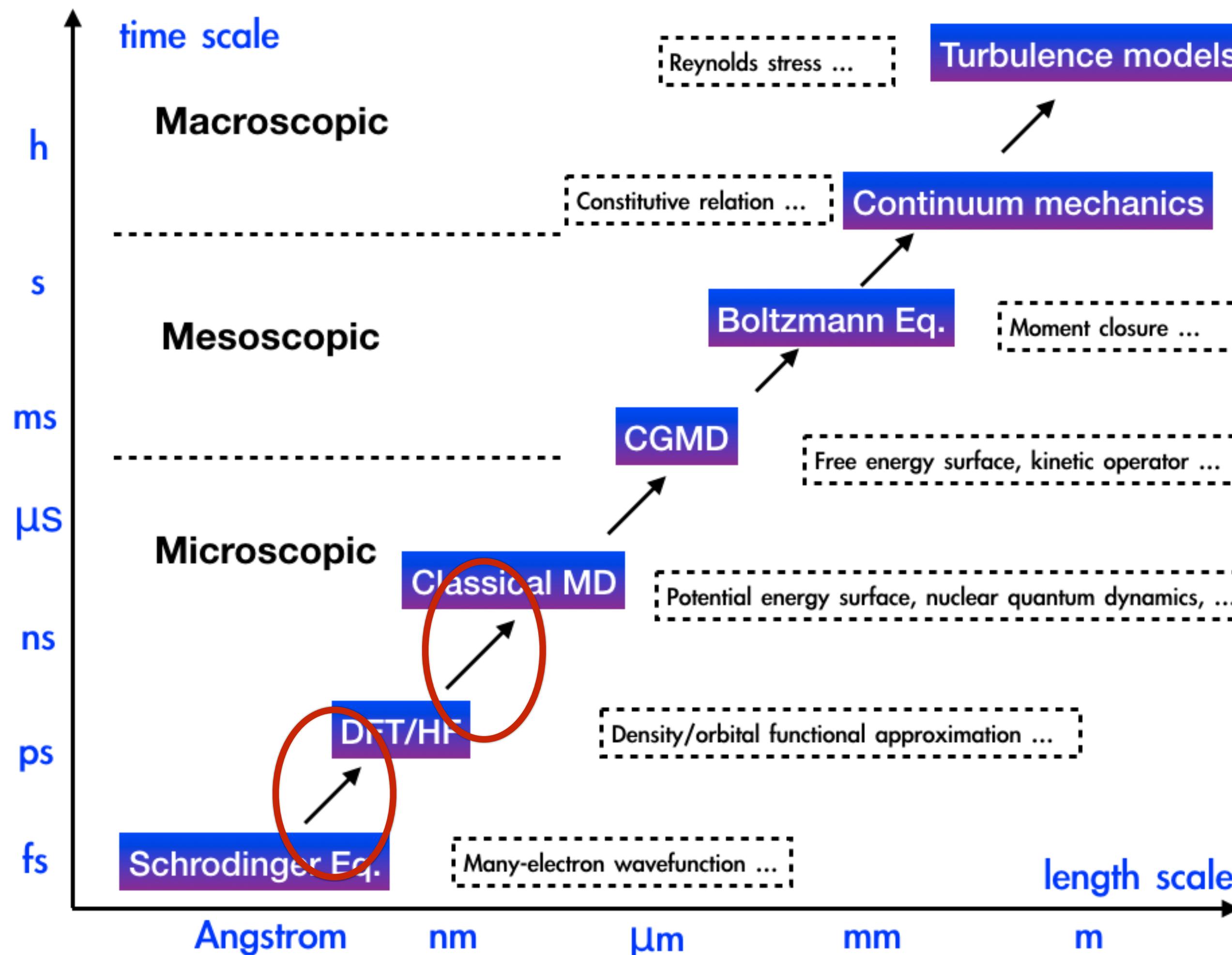
Li, LW, PRL '18



[lio12589/NeuralRG](https://github.com/lio12589/NeuralRG)



Back to deep potential, a bigger picture



Density functional theory in 1 slide

$$\begin{array}{ccc} \Psi(\mathbf{r}_1, \mathbf{r}_2, \dots, \mathbf{r}_N) & \xrightarrow{\hspace{1cm}} & \rho(\mathbf{r}) \\ \mathbb{R}^{3N} \mapsto \mathbb{C} & & \mathbb{R}^3 \mapsto \mathbb{R} \end{array}$$

$$E[\rho] = F[\rho] + \int d\mathbf{r} v_{ext}(\mathbf{r})$$

Hohenberg, Kohn 1964

Universal functional

$$K[\rho] + E_h[\rho] + E_{xc}[\rho]$$

kinetic
energy

mean-field
energy

exchange-correlation
energy

$$\begin{array}{c} \leftarrow \quad \downarrow \\ K[\rho] + E_h[\rho] + E_{xc}[\rho] \quad \frac{\delta E}{\delta \rho} = 0 \\ \text{kinetic energy} \quad \text{mean-field energy} \quad \text{exchange-correlation energy} \\ \rho = \sum_i |\psi_i|^2 \end{array}$$

$$\left(-\frac{\nabla_i^2}{2} + v_{ext} + v_h[\rho] + v_{xc}[\rho] \right) \psi_i = \varepsilon_i \psi_i$$

Kohn, Sham 1965

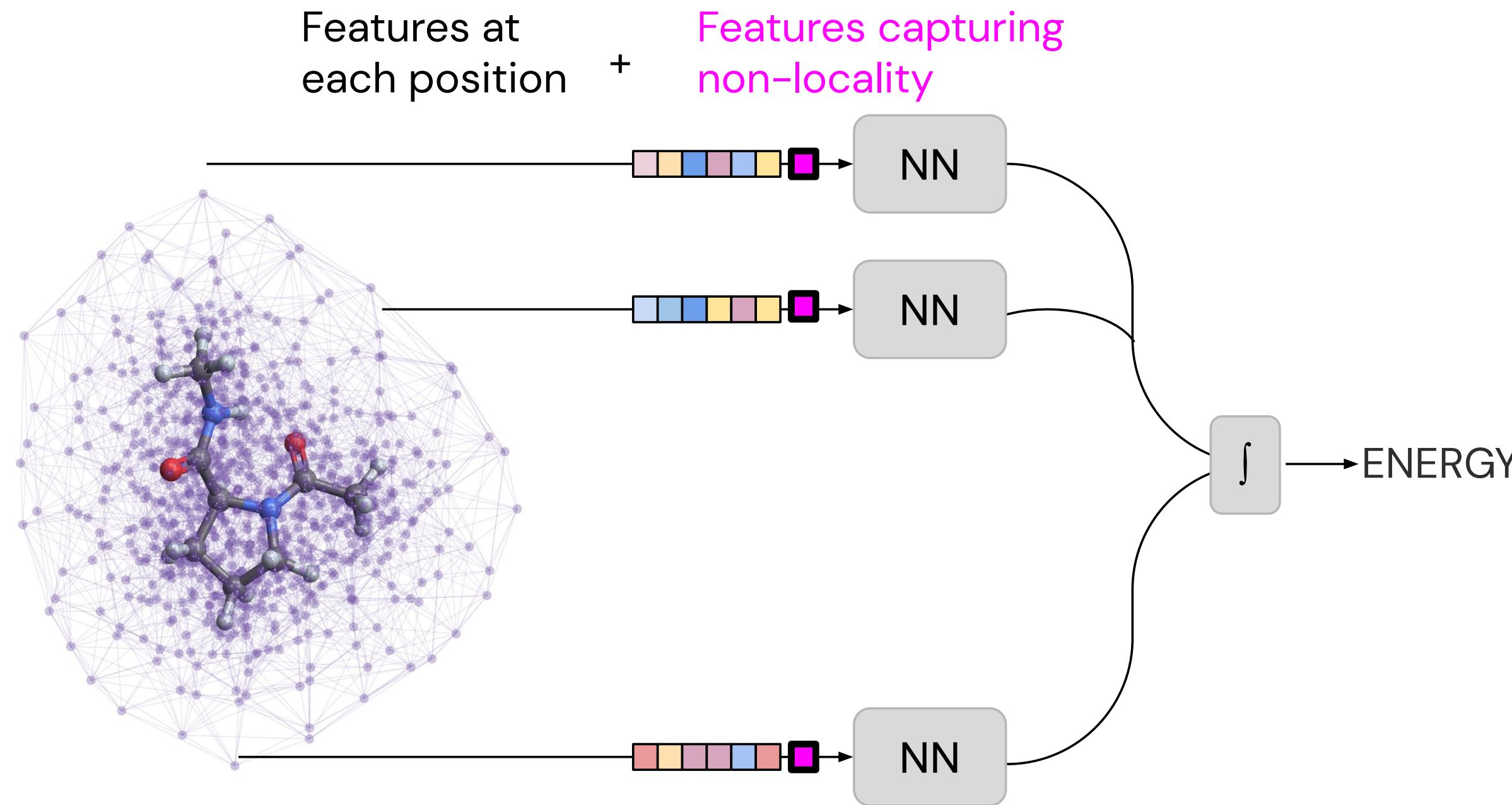
Machine learning density-functionals

Author(s)	Date	Title	Citation Count	Abstract
W. Kohn, L. J. Sham	1965-07-01	An 1000 Citations Through June 2003	1000	The ground-state energy of the independent-particle approximation to the many-body problem in one dimension is calculated by the method of variational principle. The basis functions are plane waves with wave numbers which are constrained to satisfy the periodic boundary condition. The resulting equations are solved by the method of steepest descent. The energy per electron is plotted against the density of electrons. The minimum of this curve is at zero density. At non-zero density the energy per electron is higher than the energy per electron of the free-electron model.
P. Hohenberg, W. Kohn	1964-07-01	An 1000 Citations Through June 2003	1000	The ground-state energy of the independent-particle approximation to the many-body problem in one dimension is calculated by the method of variational principle. The basis functions are plane waves with wave numbers which are constrained to satisfy the periodic boundary condition. The resulting equations are solved by the method of steepest descent. The energy per electron is plotted against the density of electrons. The minimum of this curve is at zero density. At non-zero density the energy per electron is higher than the energy per electron of the free-electron model.
J. P. Perdew, A. Zunger	1981-07-01	An 1000 Citations Through June 2003	1000	A new local-density approximation to the energy functional of the independent-particle theory of the electronic structure of matter is derived. The method is based on the local density approximation to the exchange potential and on the local form of the correlation function. The resulting energy functional is exact for atomic systems and is compared with the exact Hartree-Fock energy for atoms and molecules.
D. M. Ceperley, B. J. Alder	1973-07-01	An 1000 Citations Through June 2003	1000	The ground-state energy of the independent-particle approximation to the many-body problem in one dimension is calculated by the method of variational principle. The basis functions are plane waves with wave numbers which are constrained to satisfy the periodic boundary condition. The resulting equations are solved by the method of steepest descent. The energy per electron is plotted against the density of electrons. The minimum of this curve is at zero density. At non-zero density the energy per electron is higher than the energy per electron of the free-electron model.
J. Bardeen, L. N. Cooper, J. R. Schrieffer	1957-07-01	An 1000 Citations Through June 2003	1000	The ground-state energy of the independent-particle approximation to the many-body problem in one dimension is calculated by the method of variational principle. The basis functions are plane waves with wave numbers which are constrained to satisfy the periodic boundary condition. The resulting equations are solved by the method of steepest descent. The energy per electron is plotted against the density of electrons. The minimum of this curve is at zero density. At non-zero density the energy per electron is higher than the energy per electron of the free-electron model.
S. Weinberg	1964-07-01	An 1000 Citations Through June 2003	1000	The ground-state energy of the independent-particle approximation to the many-body problem in one dimension is calculated by the method of variational principle. The basis functions are plane waves with wave numbers which are constrained to satisfy the periodic boundary condition. The resulting equations are solved by the method of steepest descent. The energy per electron is plotted against the density of electrons. The minimum of this curve is at zero density. At non-zero density the energy per electron is higher than the energy per electron of the free-electron model.
O. K. Anderson	1957-07-01	An 1000 Citations Through June 2003	1000	The ground-state energy of the independent-particle approximation to the many-body problem in one dimension is calculated by the method of variational principle. The basis functions are plane waves with wave numbers which are constrained to satisfy the periodic boundary condition. The resulting equations are solved by the method of steepest descent. The energy per electron is plotted against the density of electrons. The minimum of this curve is at zero density. At non-zero density the energy per electron is higher than the energy per electron of the free-electron model.
U. Fano	1964-07-01	An 1000 Citations Through June 2003	1000	The ground-state energy of the independent-particle approximation to the many-body problem in one dimension is calculated by the method of variational principle. The basis functions are plane waves with wave numbers which are constrained to satisfy the periodic boundary condition. The resulting equations are solved by the method of steepest descent. The energy per electron is plotted against the density of electrons. The minimum of this curve is at zero density. At non-zero density the energy per electron is higher than the energy per electron of the free-electron model.
T. A. Lee, T. V. Ramakrishnan	1985-07-01	An 1000 Citations Through June 2003	1000	The ground-state energy of the independent-particle approximation to the many-body problem in one dimension is calculated by the method of variational principle. The basis functions are plane waves with wave numbers which are constrained to satisfy the periodic boundary condition. The resulting equations are solved by the method of steepest descent. The energy per electron is plotted against the density of electrons. The minimum of this curve is at zero density. At non-zero density the energy per electron is higher than the energy per electron of the free-electron model.
H. J. Monkhorst, J. D. Pack	1957-07-01	An 1000 Citations Through June 2003	1000	The ground-state energy of the independent-particle approximation to the many-body problem in one dimension is calculated by the method of variational principle. The basis functions are plane waves with wave numbers which are constrained to satisfy the periodic boundary condition. The resulting equations are solved by the method of steepest descent. The energy per electron is plotted against the density of electrons. The minimum of this curve is at zero density. At non-zero density the energy per electron is higher than the energy per electron of the free-electron model.
view Letters; RMP, Reviews of Modern Physics.	1957-07-01	An 1000 Citations Through June 2003	1000	The ground-state energy of the independent-particle approximation to the many-body problem in one dimension is calculated by the method of variational principle. The basis functions are plane waves with wave numbers which are constrained to satisfy the periodic boundary condition. The resulting equations are solved by the method of steepest descent. The energy per electron is plotted against the density of electrons. The minimum of this curve is at zero density. At non-zero density the energy per electron is higher than the energy per electron of the free-electron model.
Redner, Phys. Today 2005	2005-07-01	Authors are all DFT-foundational papers!	1000	The ground-state energy of the independent-particle approximation to the many-body problem in one dimension is calculated by the method of variational principle. The basis functions are plane waves with wave numbers which are constrained to satisfy the periodic boundary condition. The resulting equations are solved by the method of steepest descent. The energy per electron is plotted against the density of electrons. The minimum of this curve is at zero density. At non-zero density the energy per electron is higher than the energy per electron of the free-electron model.
TABLE III. All the necessary information to construct our MLA, trained from 100 densities with $N = 1$ on a grid of 500 points, with $\lambda = 12 \times 10^{14}$ and $\sigma = 43$. For purposes of saving space, we do not list these densities. They may be reconstructed from these potentials via Numerov's method.	2012-07-01	machine learned one dimensional kinetic-energy functional with $O(10^3)$ parameters	100	The ground-state energy of the independent-particle approximation to the many-body problem in one dimension is calculated by the method of variational principle. The basis functions are plane waves with wave numbers which are constrained to satisfy the periodic boundary condition. The resulting equations are solved by the method of steepest descent. The energy per electron is plotted against the density of electrons. The minimum of this curve is at zero density. At non-zero density the energy per electron is higher than the energy per electron of the free-electron model.

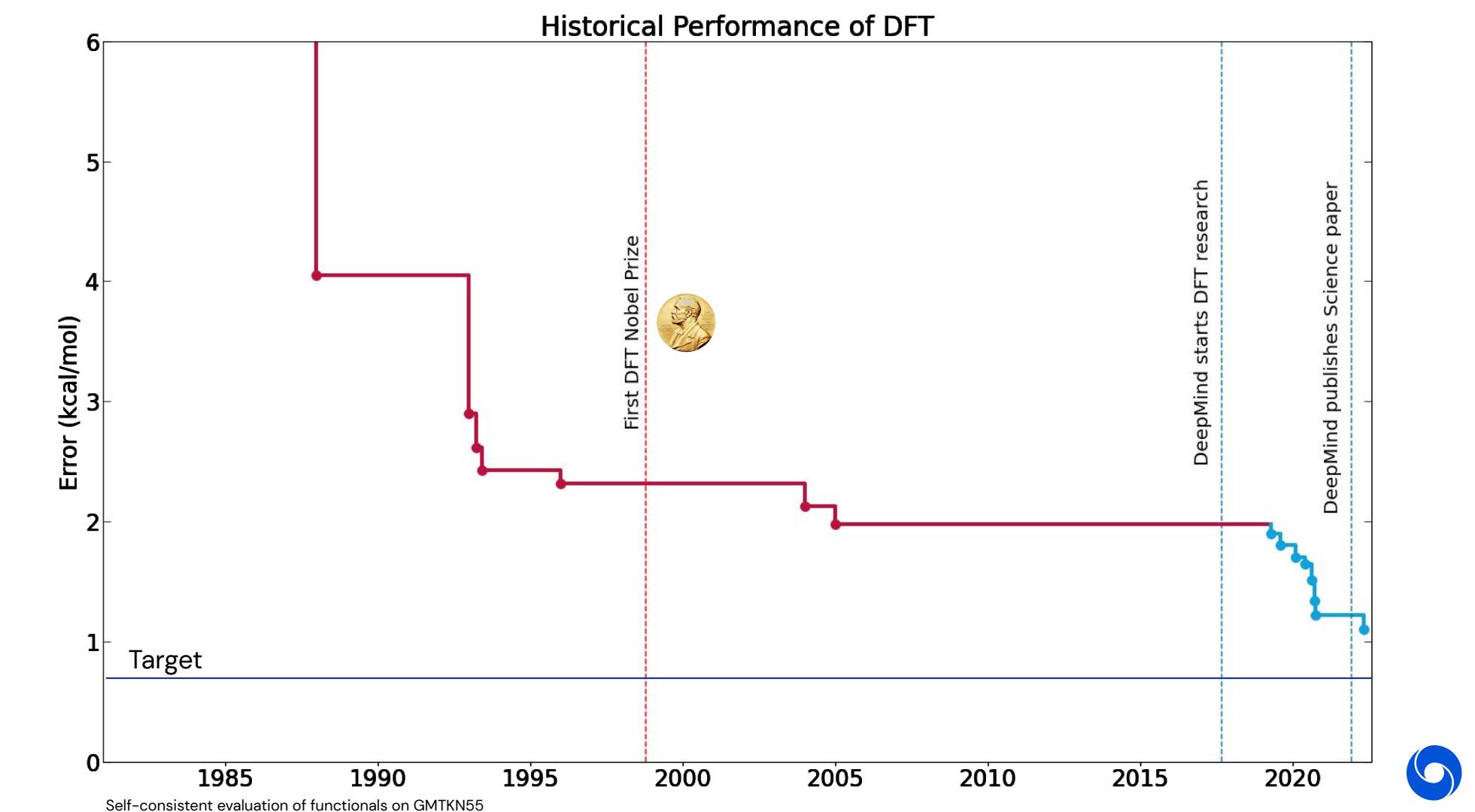
Snyder et al, PRL 108, 253002 (2012)

The DM-21 functional

Kirkpatrick et al, Science 2021



https://drive.google.com/file/d/1yJ6ocWs_aicfp2gLQu5XuJ_Kdcjryx6P/view?usp=share_link

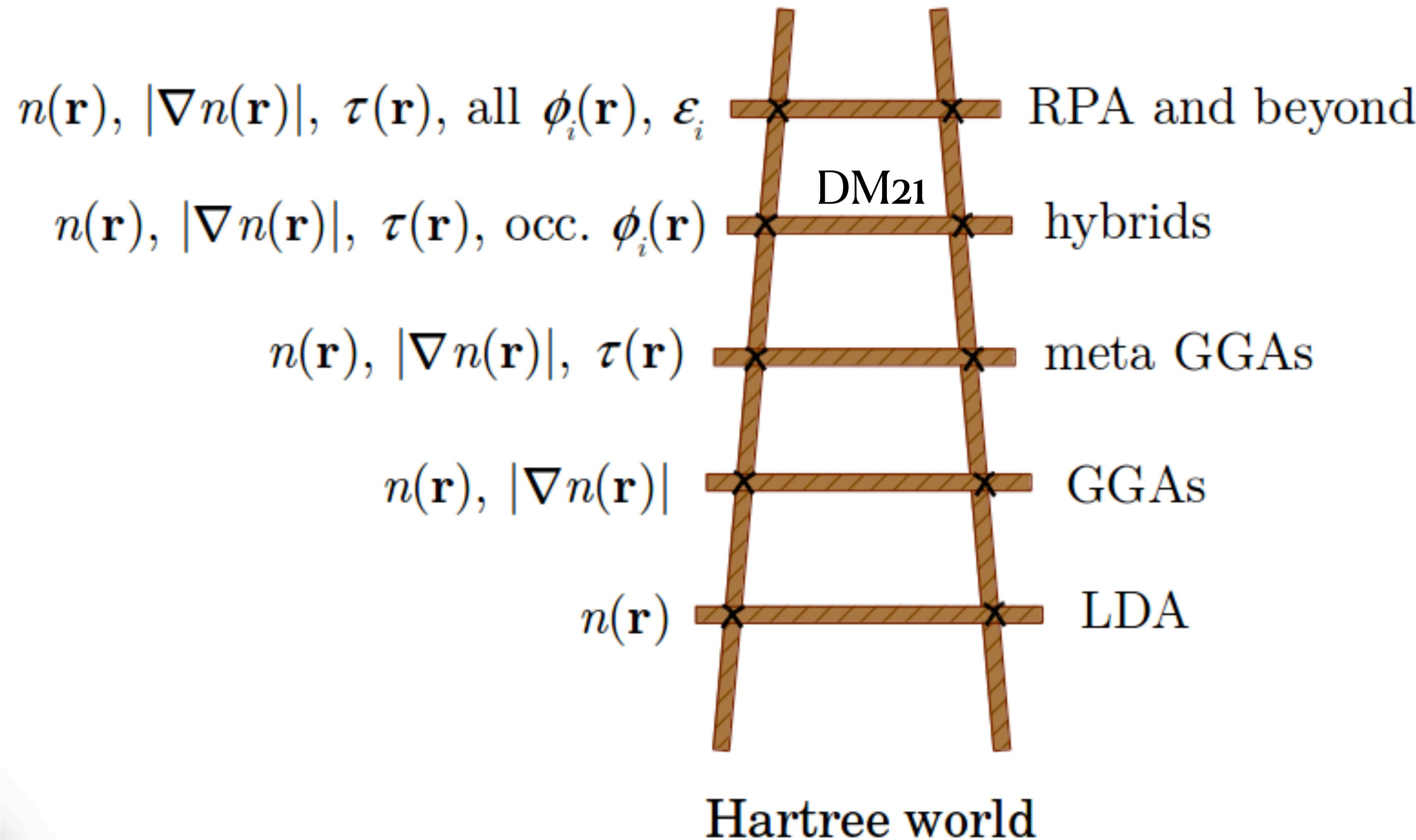


“Training a neural network
on molecular data and
on fictitious systems with
fractional charge and spin”

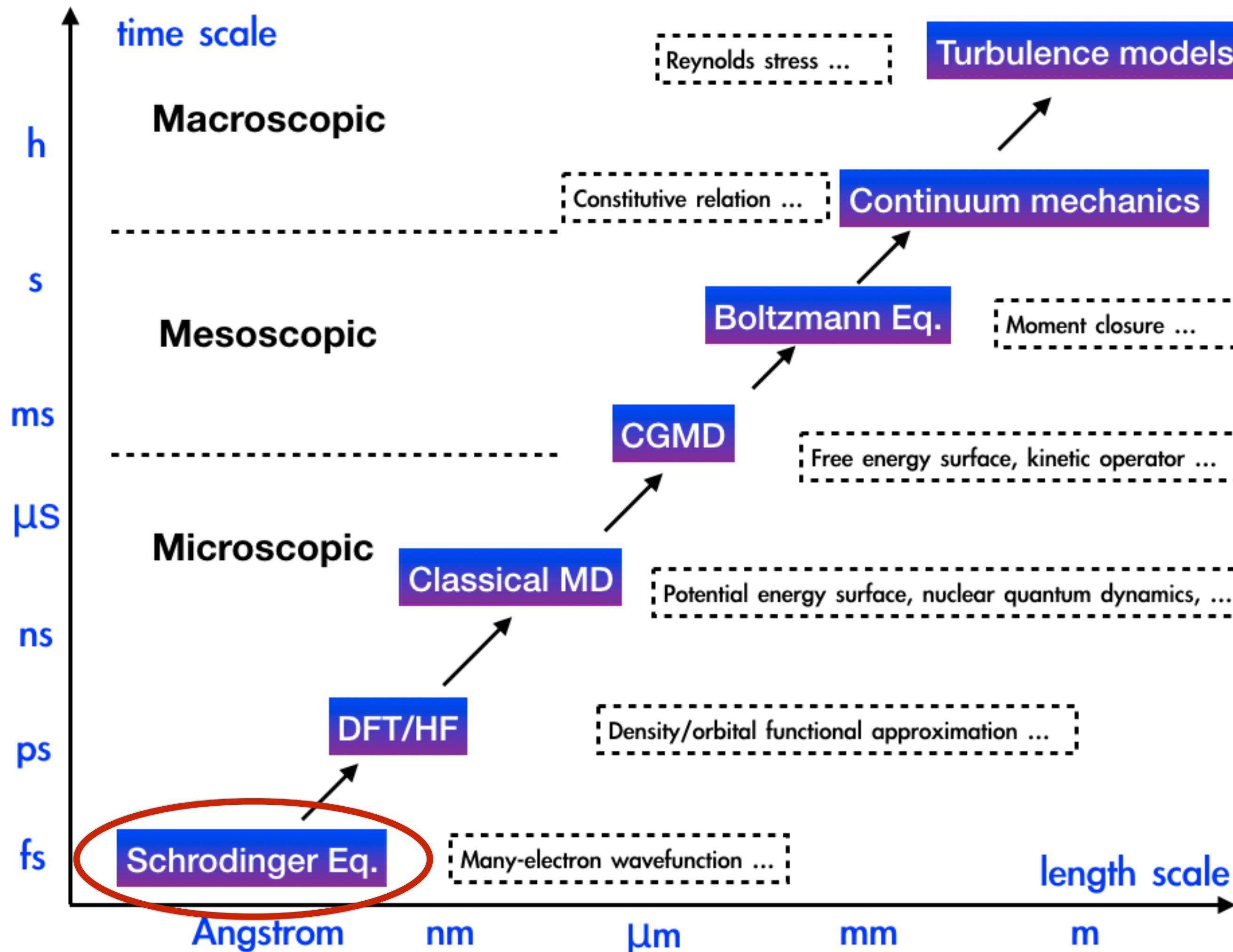
the manuscript, with help from L.R.C. **Competing interests:** There is a pending patent application, US Provisional Application 63/135,223, on this work. The remaining authors declare no competing

Way to the ultimate functional $F[\rho]$?

Heaven of chemical accuracy



What if there are no data ?

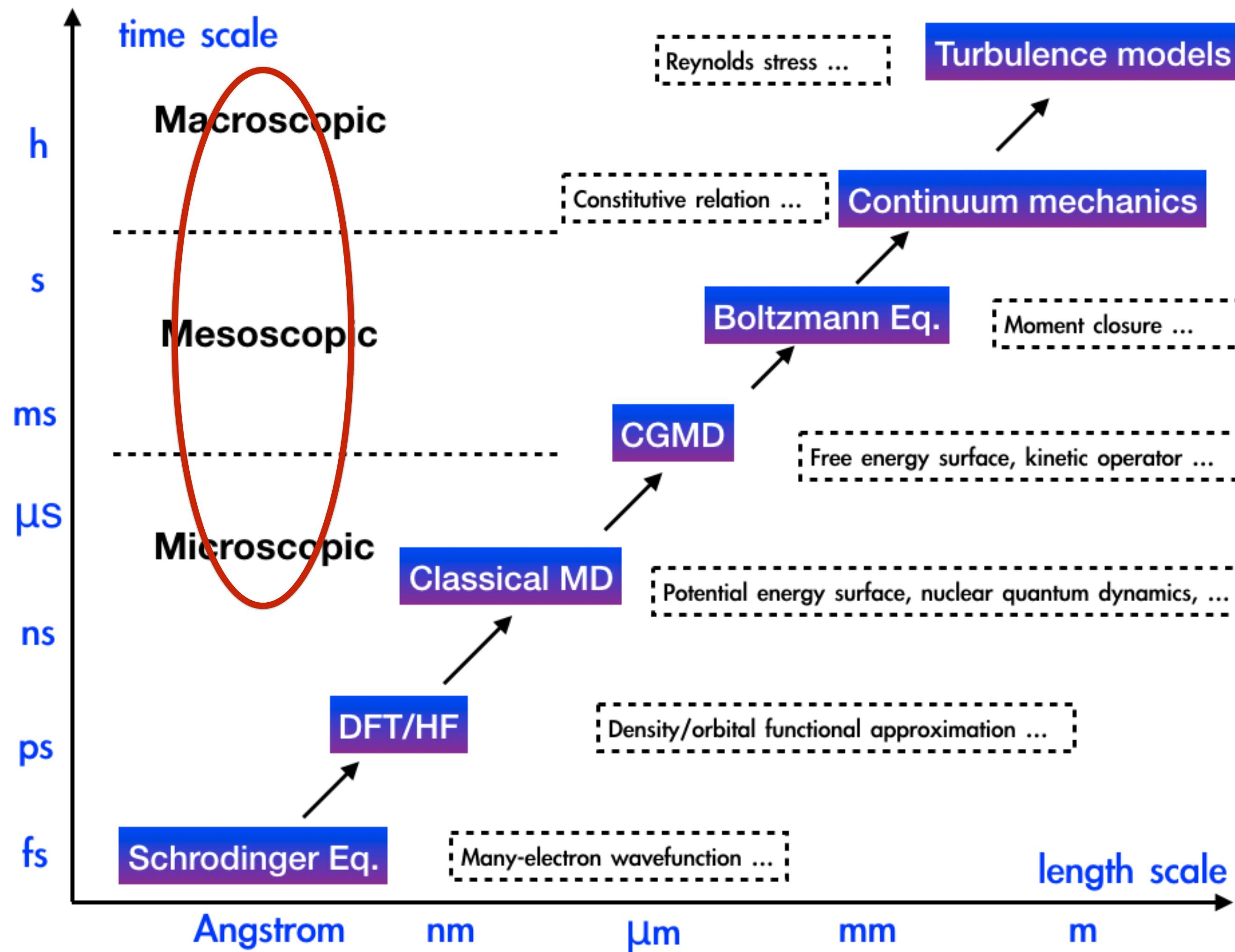


$$H|\Psi\rangle = E|\Psi\rangle$$

Theory of everything (for your everyday life)

$$x \sim \frac{e^{-\beta E(x)}}{Z}$$

Intractable!



The principle of least action

Machine learning:
objective function

$$S = \int d\tau \sqrt{g} [\frac{1}{2} R$$

$$+ \frac{1}{2} F^2 + \nabla \cdot P \Psi$$

$$+ (\partial_\mu \Psi)^2 + V(\Psi)$$

$$+ \bar{\Psi} \Psi]$$

Physics:
THE action

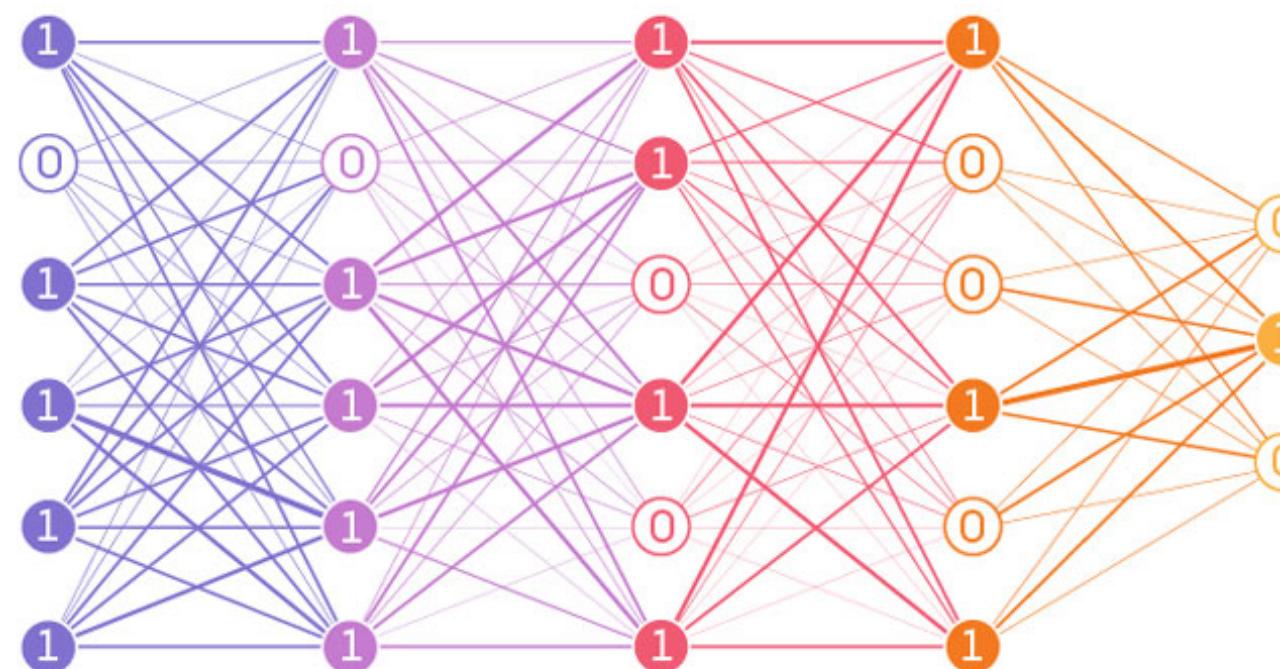
Both quantum and statistical physics can be turned into
optimization problems (for wavefunction and probability respectively)

Solving Schrödinger equation with ML

$$H|\Psi\rangle = E|\Psi\rangle$$

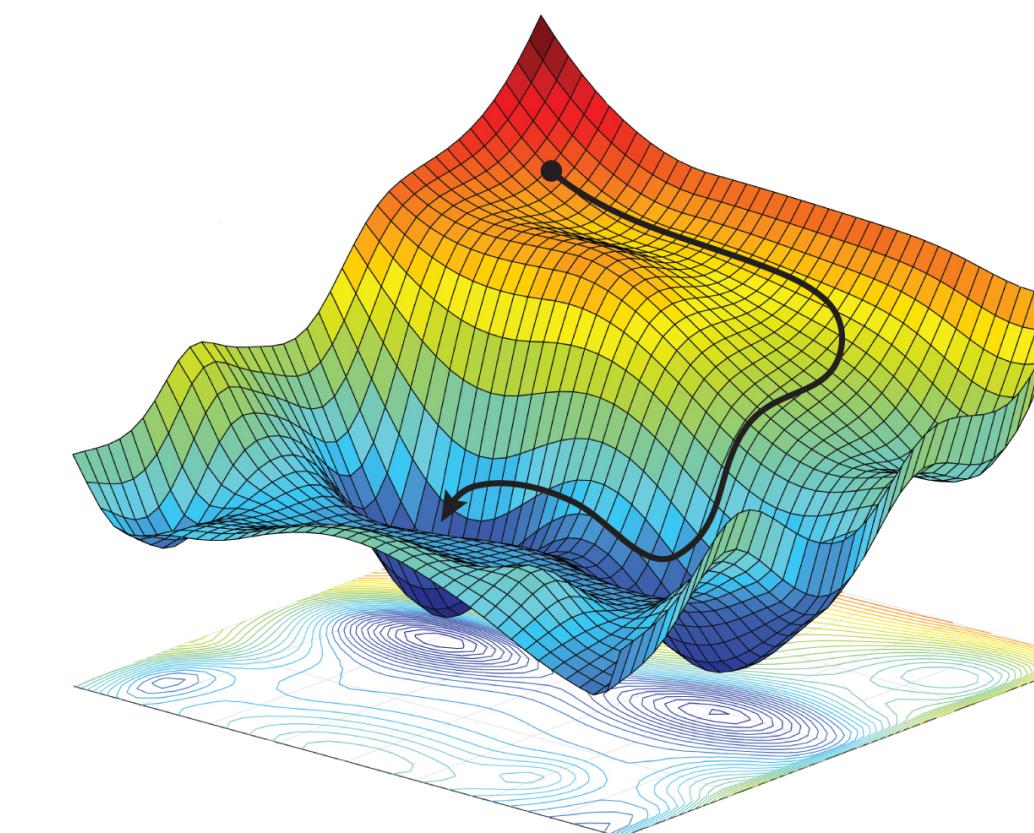
Variational ground state energy $\mathcal{L} = \langle \Psi | H | \Psi \rangle \geq E_g$

Representation



Neural
Networks

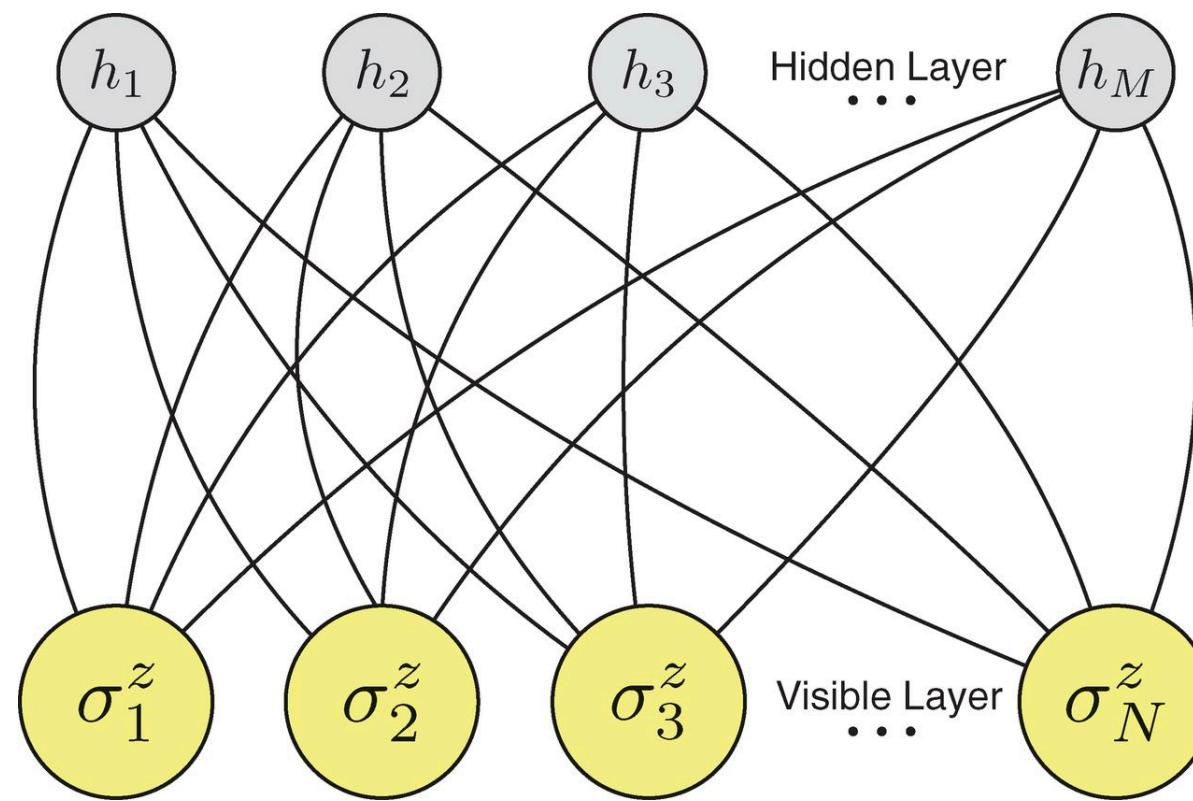
Optimization



Differentiable
Programming

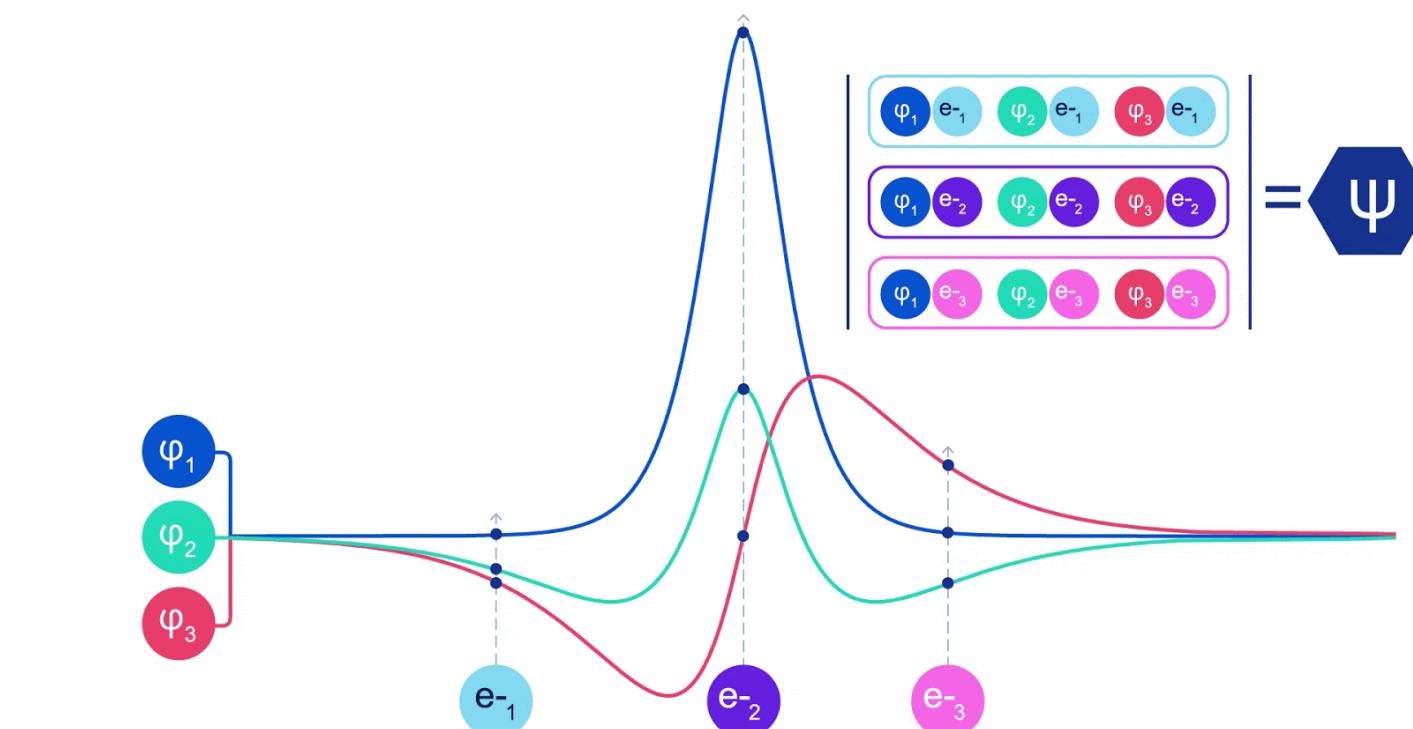
Neural network quantum states

Effective model for quantum magnetism



Carleo and Troyer, Science '17

Ab initio quantum chemistry

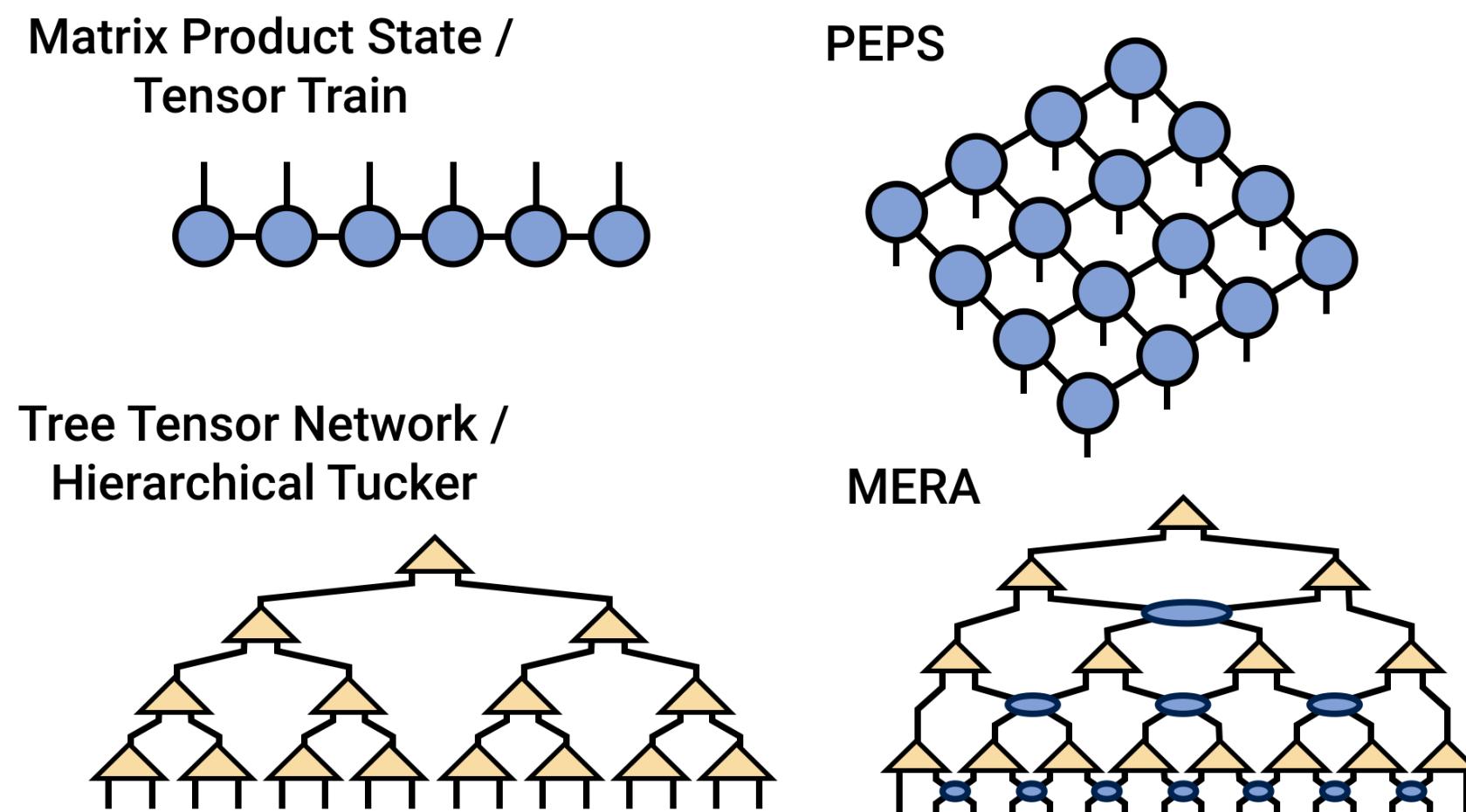


Pfau et al, PRResearch '20, Cassella, PRL '23
<https://deepmind.com/blog/article/FermiNet>

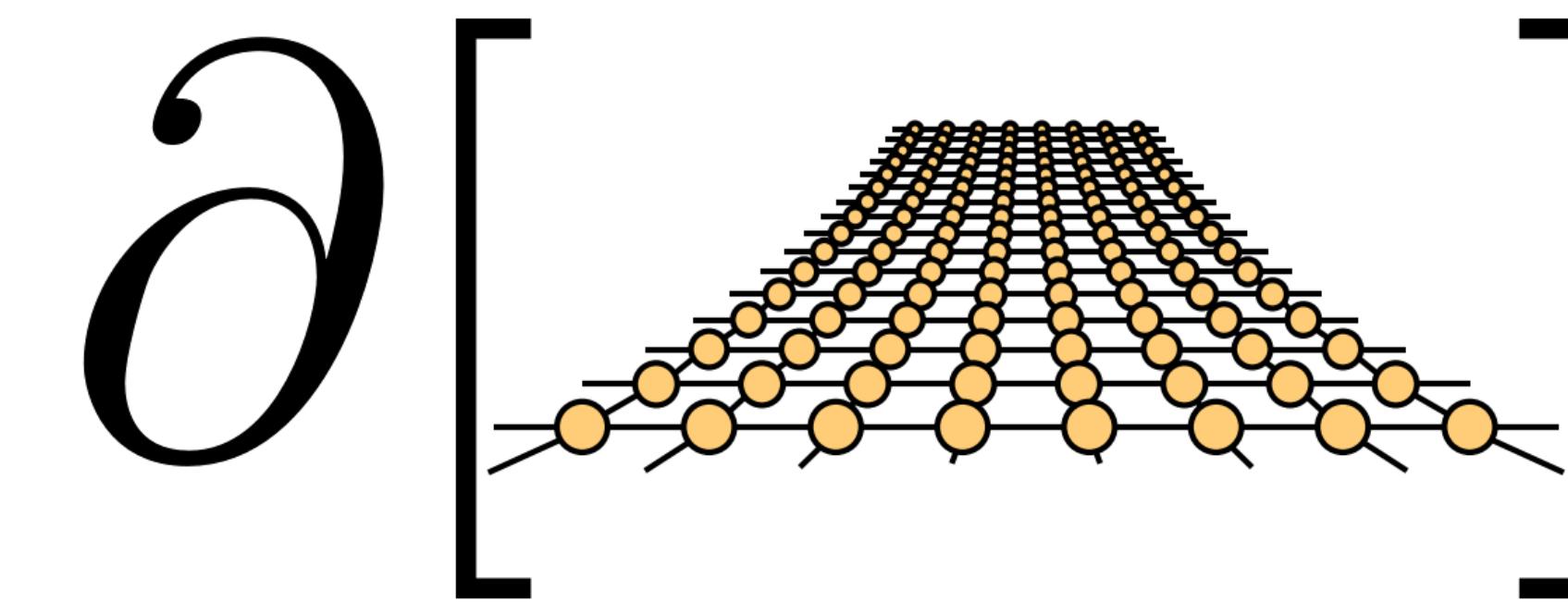
Optimization has become the bottleneck

Tensor network quantum states

Representation



Optimization



Large family of network architecture
with physical inductive biases

Differentiable programming
tensor networks

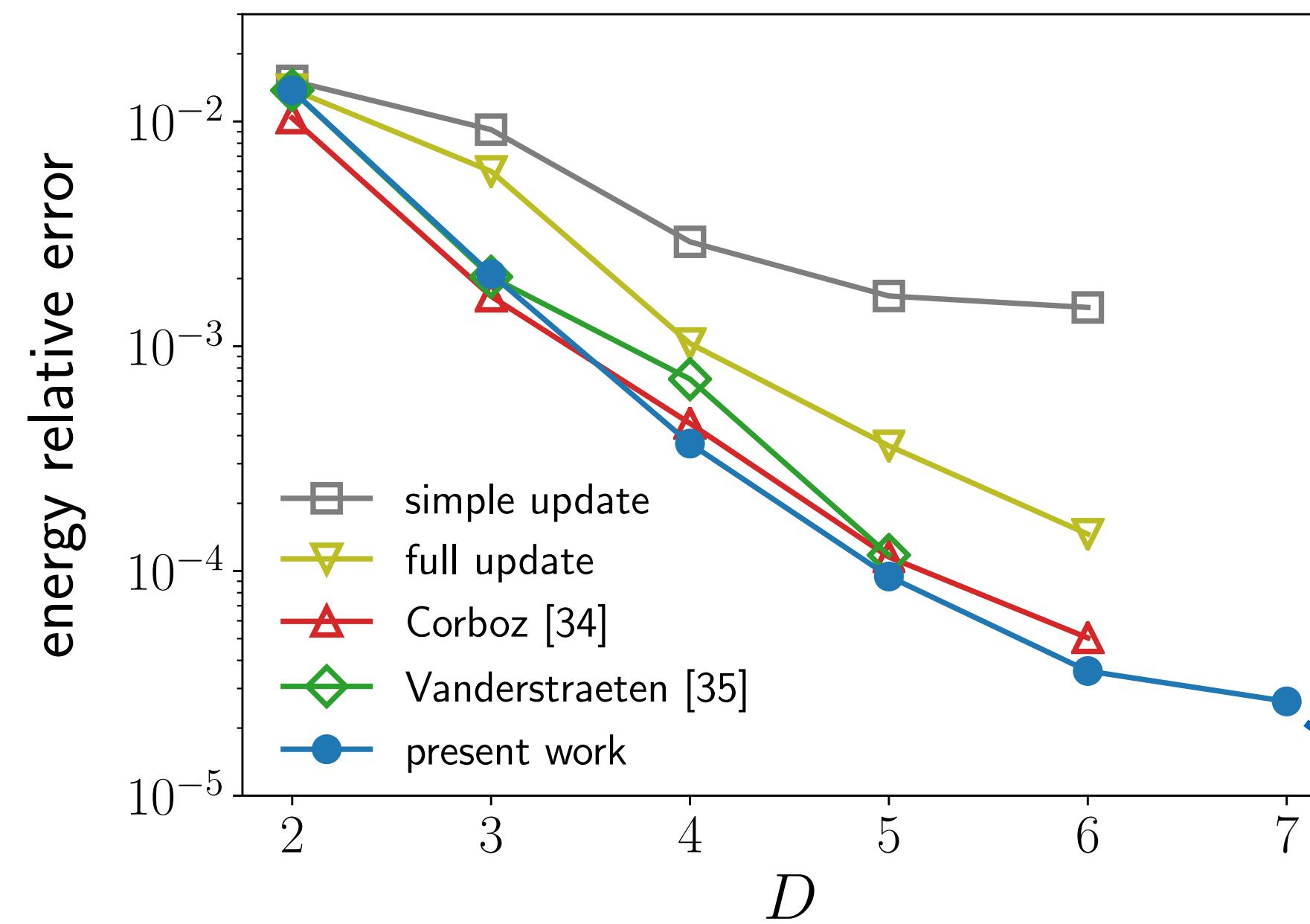
Liao, Liu, LW, Xiang, PRX '19

<https://github.com/wangleiphy/tensorgrad>

Differentiable tensor network optimization

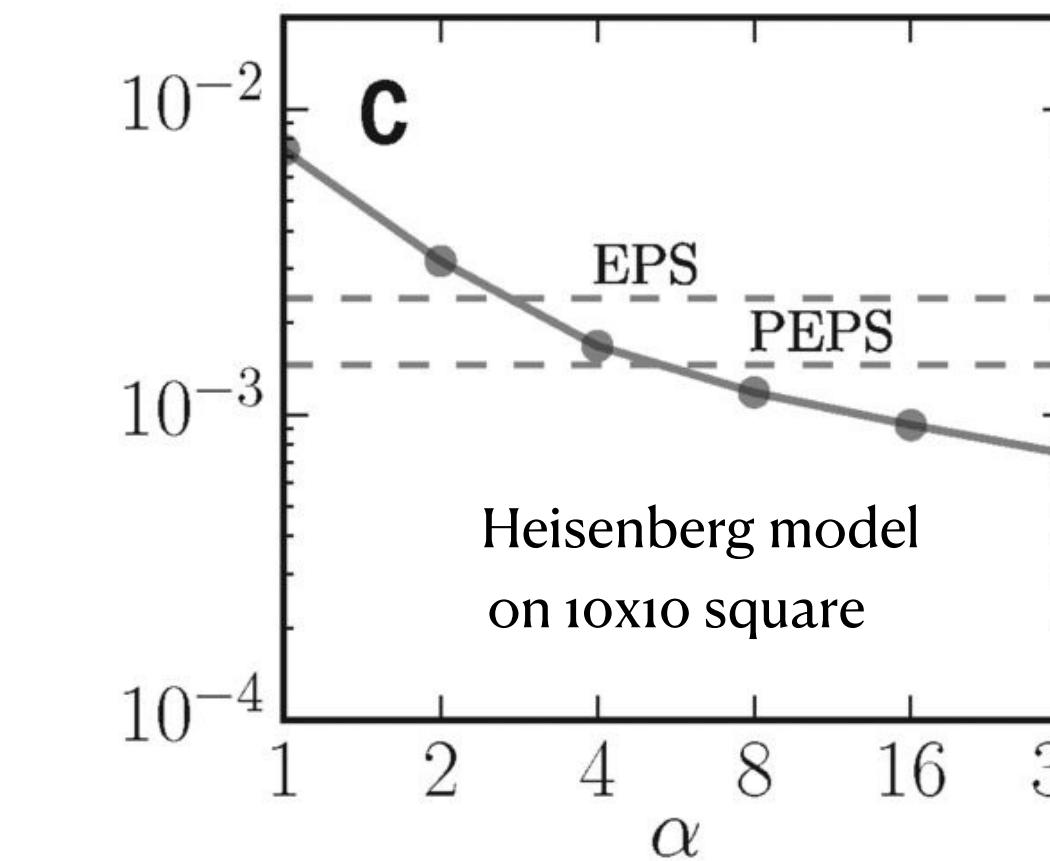
Infinite size tensor network

Liao, Liu, LW, Xiang, PRX '19



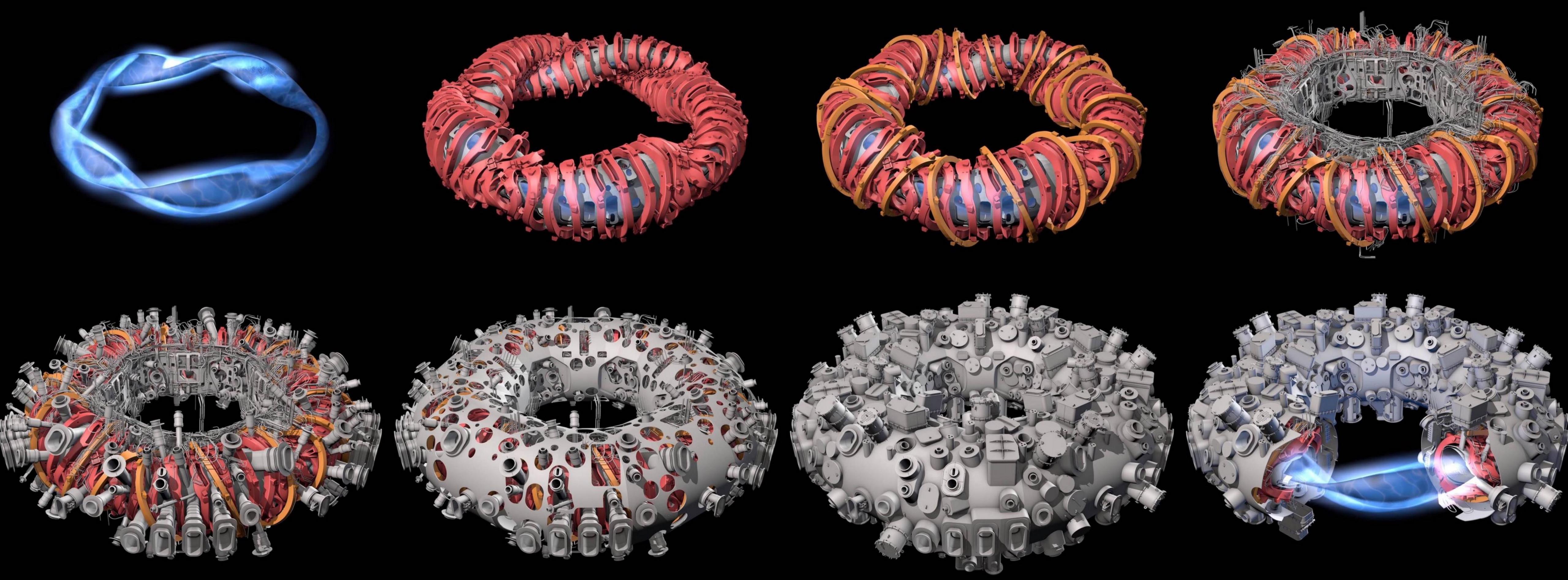
Finite size neural network

Carleo & Troyer, Science '17



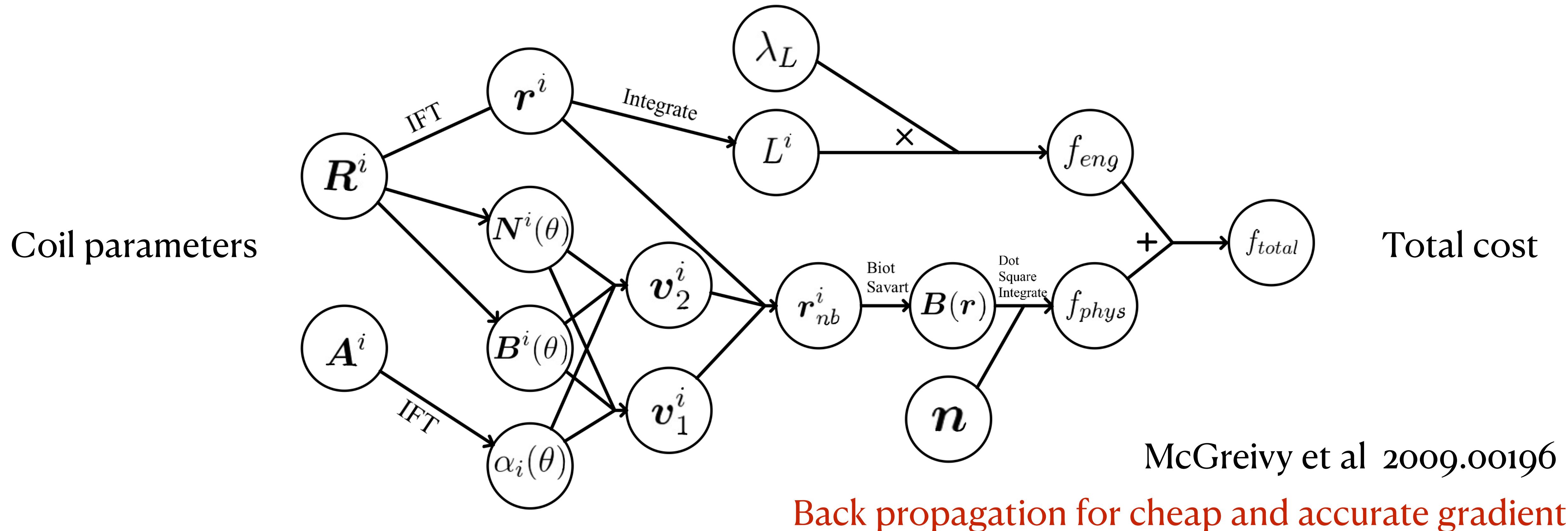
lowest variational energy
with 1 GPU week

Opens many opportunities to other challenging problems:
finite temperature, excited states, tensor network machine learning...



Coil design in fusion reactors (stellarator)

Differentiable stellarator design



Differentiable programming is broader than training neural networks

微分万物:深度学习的启示*

王磊^{1,2,†} 刘金国³

(1) 中国科学院物理研究所 北京 100190

(2) 松山湖材料实验室 东莞 523808

(3) 哈佛大学物理系 剑桥 02138)

《物理》

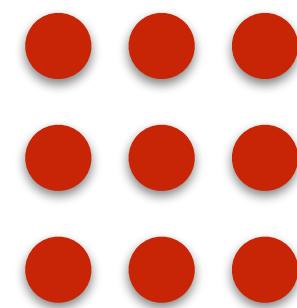
2021年2月

Solving statistical mechanics with ML

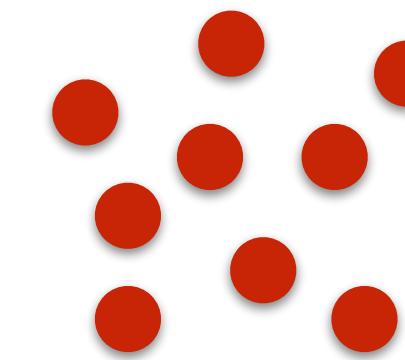
Nature tries to minimize free energy

$$F = E - TS$$

energy



entropy

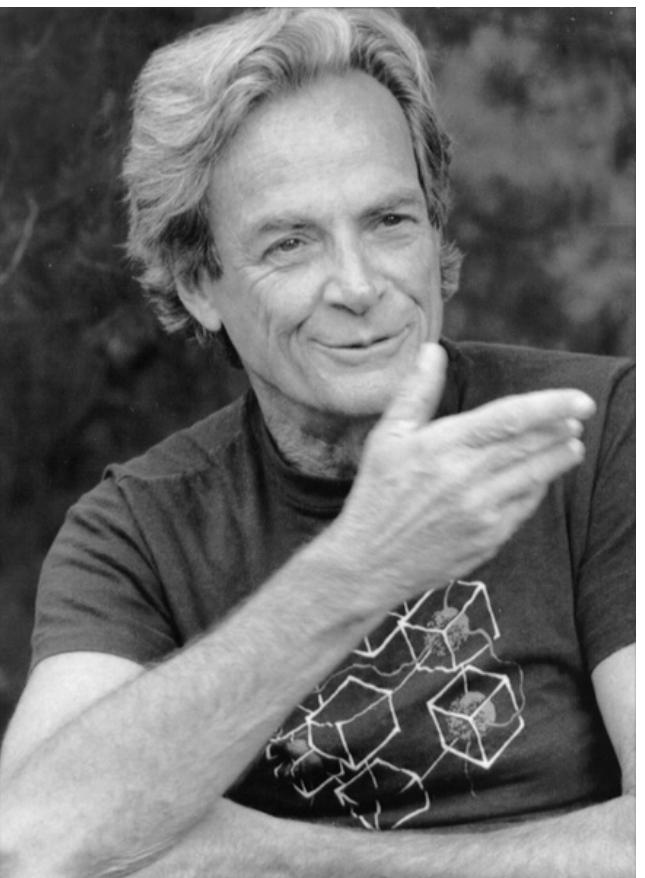


F is a **cost function** given by Nature

The variational free energy principle

$$F[p] = \int dx p(x) [H(x) + k_B T \ln p(x)] \geq F$$

↓ ↓ ↓
variational density energy entropy 😱



**Difficulties in Applying the Variational
Principle to Quantum Field Theories¹**

Richard P. Feynman

Generative
models!

¹transcript of his talk in 1987

Deep variational free energy approach

Use deep generative models as the variational density

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

↑
energy ↓
entropy 

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

with normalizing flow &
autoregressive models

- ✓ Tractable entropy
- ✓ Direct sampling
- ✓ Turning sampling/equation solving into optimization
better leverages the deep learning engine:



These points are even more true for quantum systems

Xie, Zhang, LW, JML '22

Discriminative learning



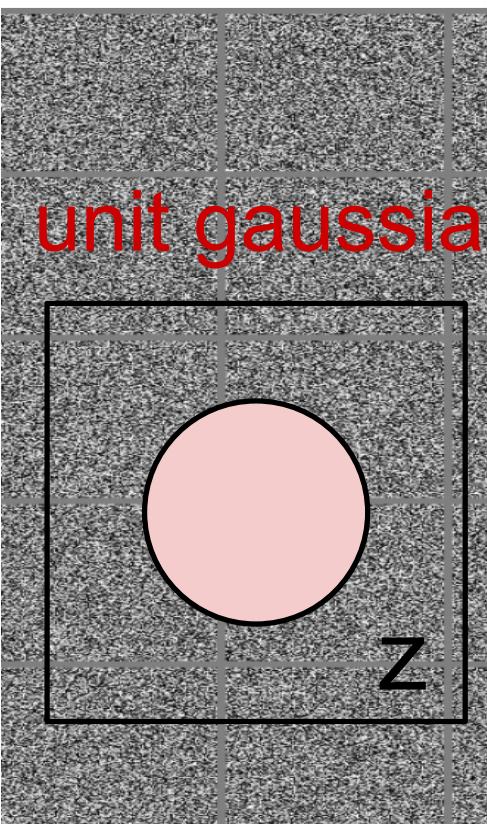
Generative learning



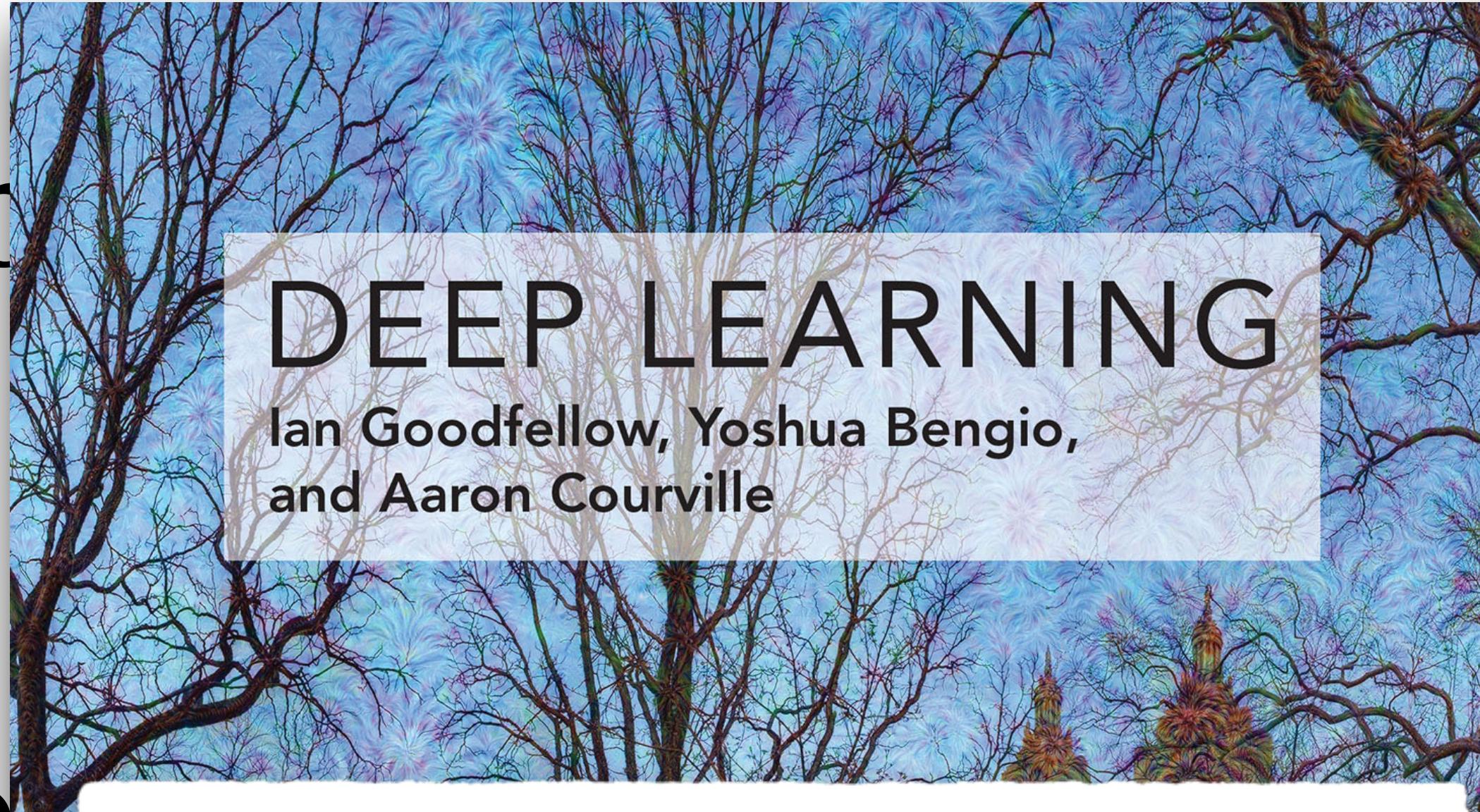
“~~What~~ $f(x)$ can not create, I do not understand”
or $p(y | x)$

Probability modeling

How to
high-dim



“random”



Page 159

“... the images encountered in AI applications occupy a negligible proportion of the volume of image space.”



distribution

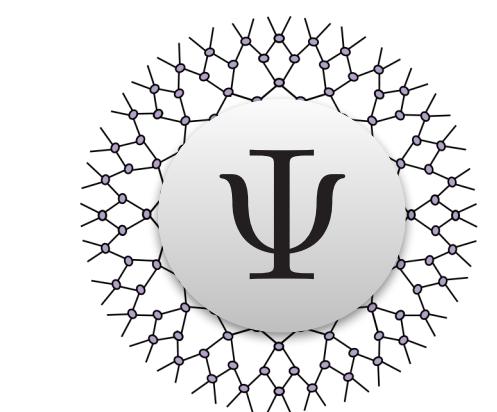


image space

[neural-models/](#)

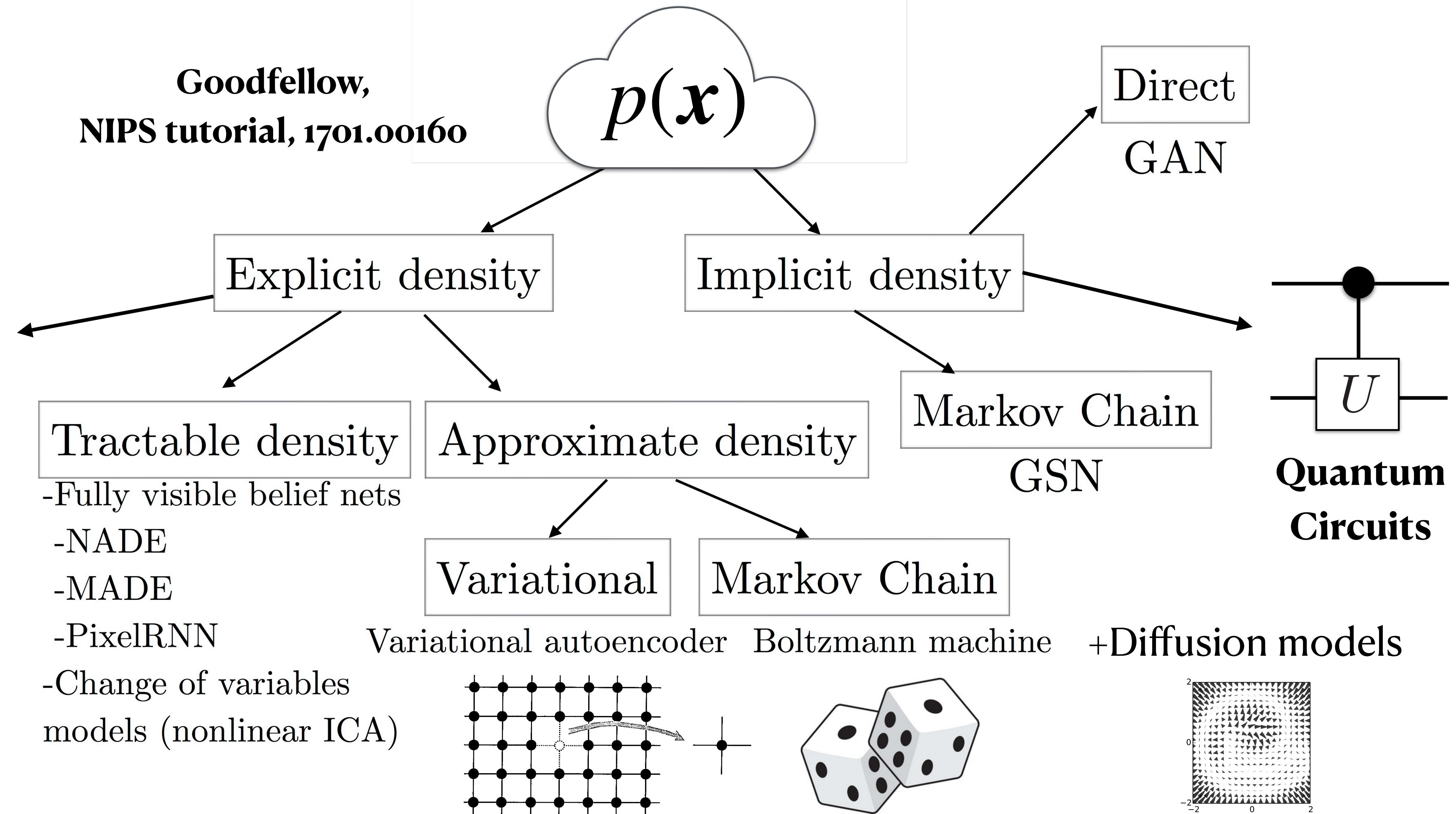
from a
ition ?

Generative models and their physics genes



**Tensor
Networks**

**Goodfellow,
NIPS tutorial, 1701.00160**



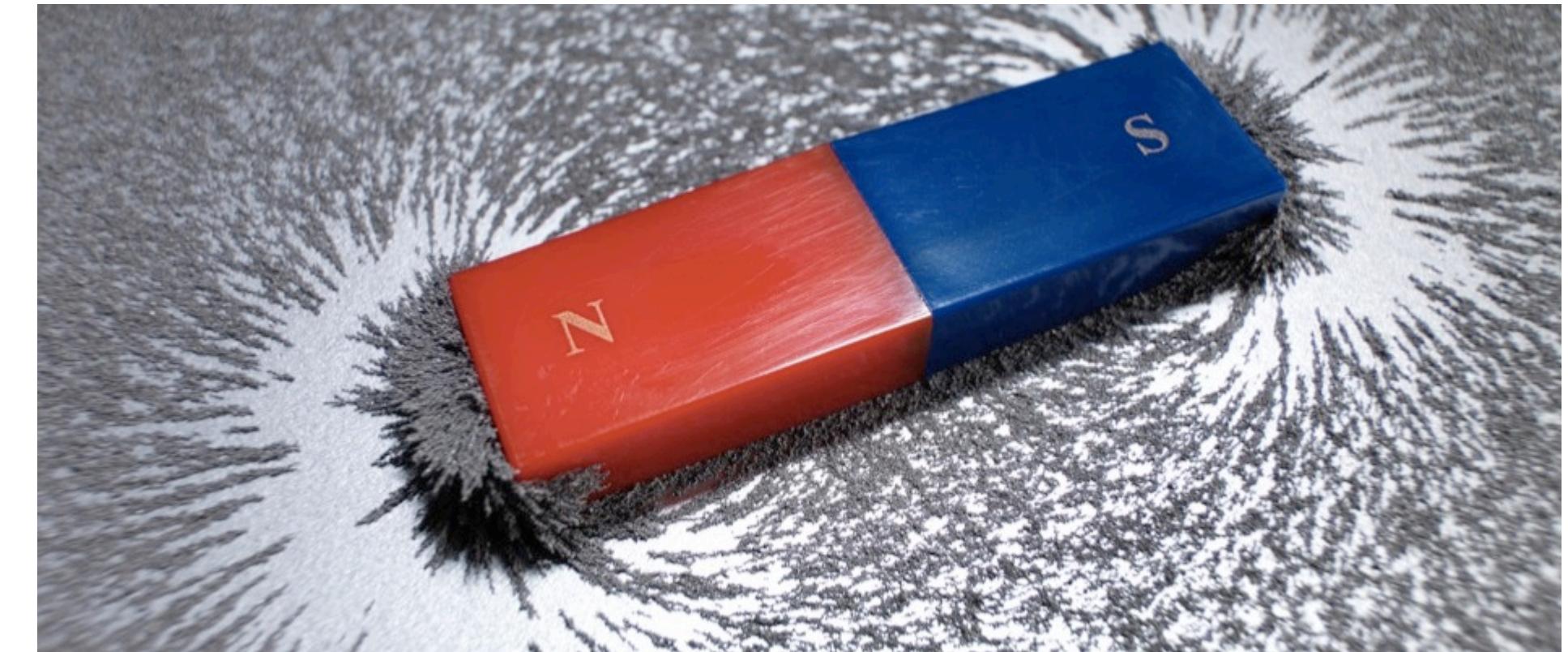
Generative modeling



Known: samples

Unknown: generating distribution

Statistical physics

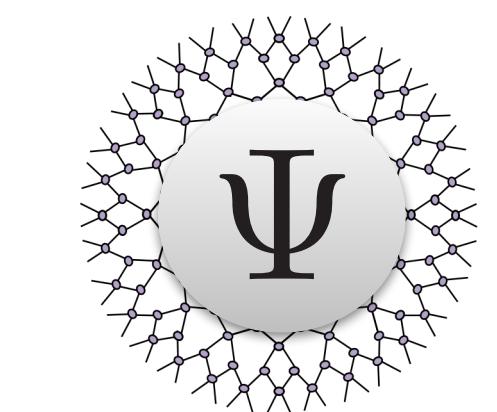


Known: energy function

Unknown: samples, partition function

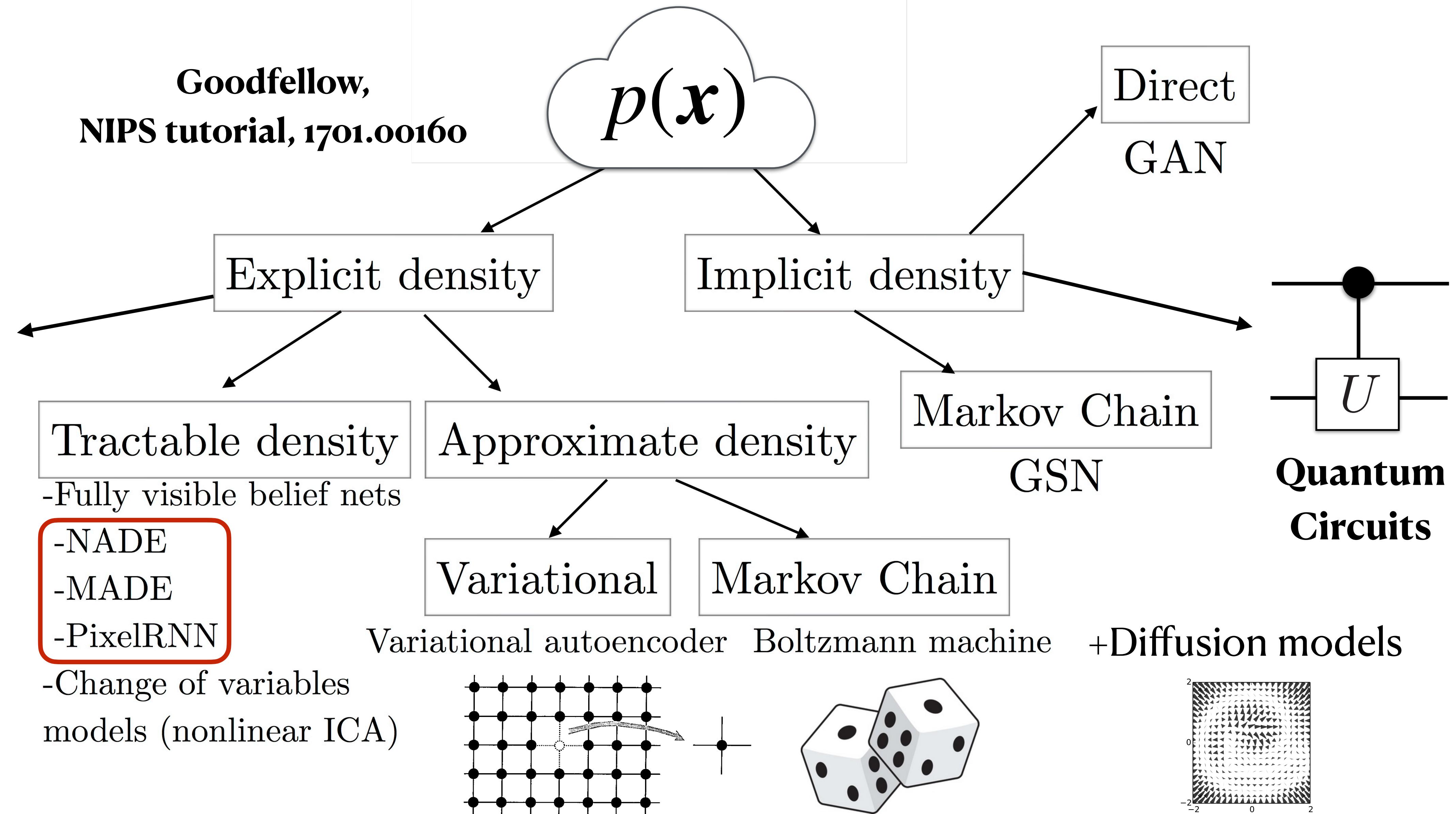
Generative models for physics
Physics of generative modeling

Generative models and their physics genes



**Tensor
Networks**

**Goodfellow,
NIPS tutorial, 1701.00160**



Autoregressive models

$$p(\mathbf{x}) = p(x_1)p(x_2 \mid x_1)p(x_3 \mid x_1, x_2)\cdots$$

Language: GPT 2005.14165



Speech: WaveNet 1609.03499

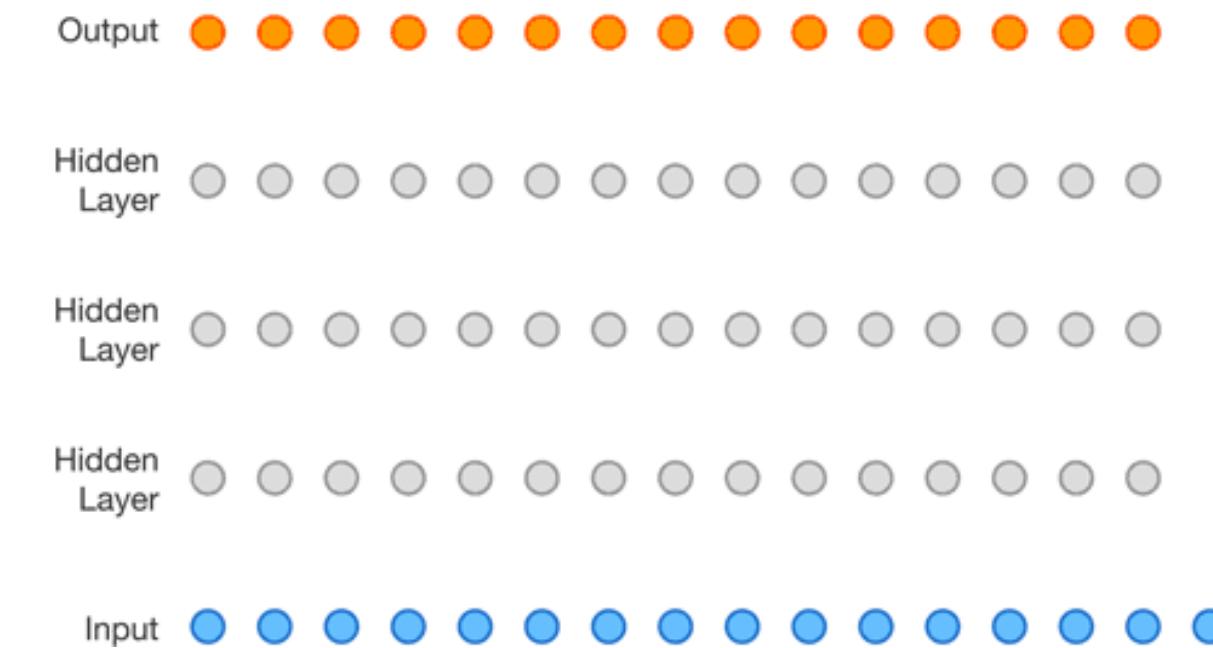
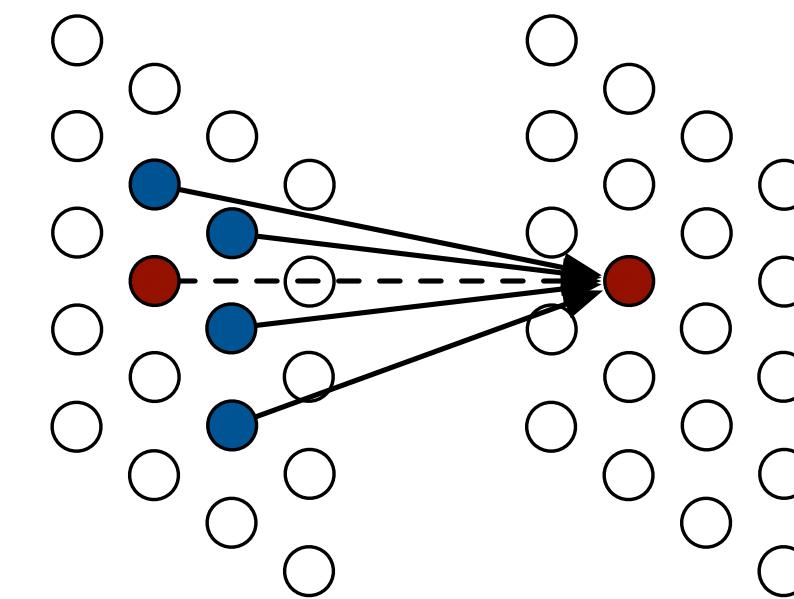
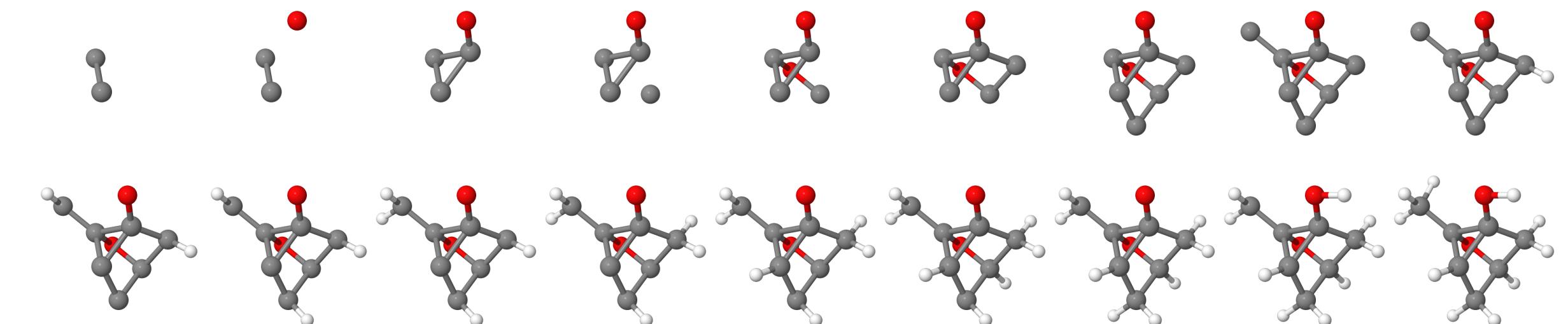


Image: PixelCNN 1601.06759

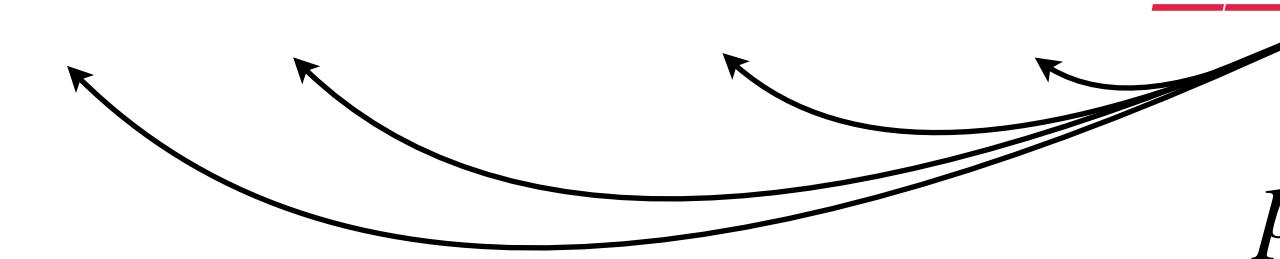


Molecular graph: 1810.11347



$$p(\mathbf{x}) = p(x_1)p(x_2 | x_1)p(x_3 | x_1, x_2)\cdots$$

“... *the murderer is* ”



$p(\underline{} | \dots)$



Jim (Linxi) Fan
@DrJimFan

...

It's not at all obvious why simply predicting the next word can give us such abilities. One intuitive explanation is to imagine a detective story. Suppose the model needs to fill in the last blank: “the murderer is ”, then it has to do deep reasoning to answer correctly. 2/

L

阅读下面的材料，根据要求作文。(60分)

有一位登山者，在途中遇上暴风雪。他深知不尽快找到避风处，非冻死不可。他走啊走啊，腿已经迈不开了。就在这时，脚碰到一个硬邦邦的东西，扒开雪一看，竟然是个快冻僵的人。

登山者犯难了：是继续向前，还是停下来援救这个陌生人？心灵深处翻江倒海之后，他毅然作出决定，脱下手套，给那人做按摩。经过一番按摩，陌生人可以活动了，而登山者也因此暖和了自己的身心。最后，两个人互相搀扶着走出了困境。

也许不是人人都会碰上这种生死的抉择，但是每个人却常常遇到、见到、听到一些触动心灵需要作出选择的事情。那时，我们大家是怎样选择的呢？又应该如何选择呢？请以“心灵的选择”为话题写一篇作文，所写内容必须在这个话题范围之内。

[注意]立意自定；文体自选；题目自拟；不少于800字；不得抄袭。

习



心灵的选择

凸 ⊖

生活中，我们常常需要做出一些重大的选择。这些选择可能关系到我们的人生道路，也可能关系到我们的道德品质。每一个人都有自己的选择，但是，我们该如何选择呢？

在这个世界上，有很多不同的选择方式。有的人会选择那些他们认为正确的事情，因为他们认为这样做是对的。这样的选择方式在一定程度上是正确的，因为我们都应该做正确的事情。但是，在某些情况下，这种选择方式并不一定是最好的。

另外一种选择方式是根据自己的感觉做出选择。比如说，当我们面对一个抉择的时候，我们会根据自己的感觉来做出选择。如果我们觉得某件事情是对的，我们就会去做。这样的选择方
式有它的劣势，因为我们

Try again

prompt

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

<https://chat.openai.com/chat>

L 警察来到一个家庭凶杀案现场，发现五人出现过凶杀现场。

A说：D是凶手
B：不是我干的
C：不是E干的
D：A说我干的，纯属谎言
E：B说的是实话

警察确定，有三个人的陈述是事实
问：谁是凶手？

根据题目中给出的信息，我们可以推断出：

- 由于有三个人的陈述是事实，所以有且只有一个人是在说谎。
- 由于E说B说的是实话，所以E是在说实话。
- 由于D说A说的是谎话，所以A是在说谎。
- 由于A说D是凶手，所以D不是凶手。
- 由于D说A说的是谎话，所以A不是在说实话。

因此，A、B、C、D、E五个人中只有A和D是在说谎。而由于D说A说的是谎话，所以A不是在说实话。也就是说，A说的是谎话，即D不是凶手。因此，凶手不是D。

我们可以继续分析，如果凶手是B，那么B说的就是谎话，而E说的也是谎话。但是根据题目中给出的信息，E说的是实话，因此B不是凶手。

如果凶手是C，那么C说的就是实话，而E说的也是实话。但是根据题目中给出的信息，E说的是实话，因此C不是凶手。

如果凶手是E，那么E说的就是谎话，而B说的也是谎话。但是根据题目中给出的信息，B说是实话，因此E不是凶手。

综上所述，A、B、C、D、E五个人中只有A和D是在说谎，B、C、E三个人的陈述是事实。
因

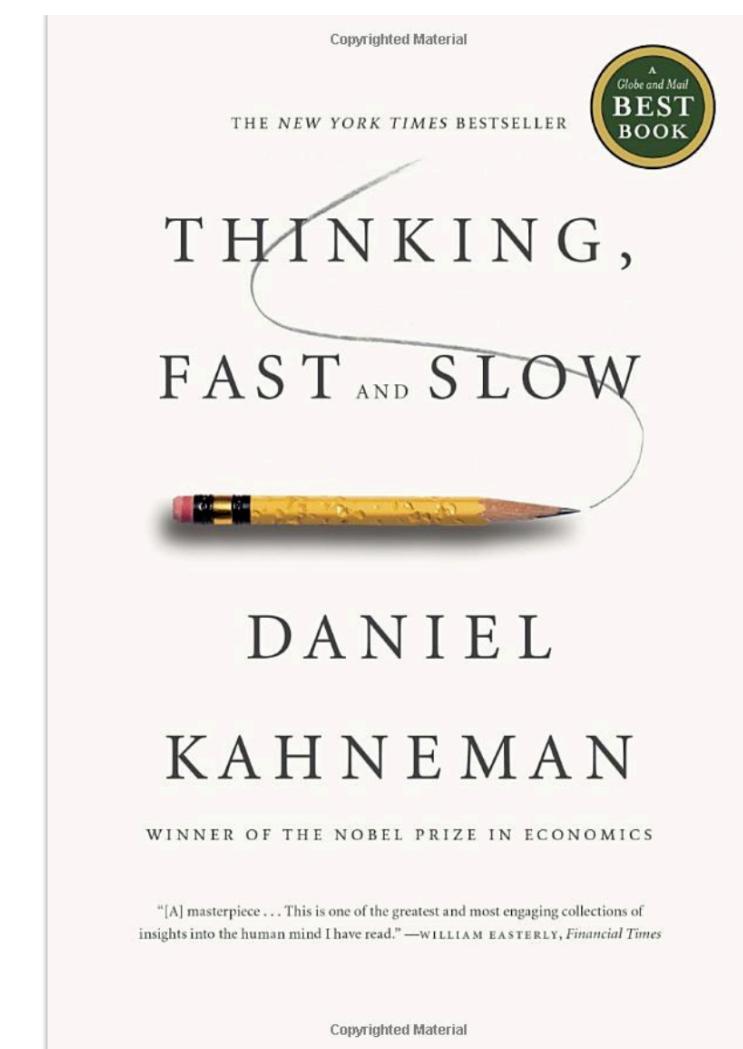
L 继续

此，凶手是A、B、C、D、E五个人中的一个人，且不是D。根据题目中给出的条件，我们无法确定凶手到底是谁。因此，我们无法确定谁是凶手。

BTW, a failed example of chatGPT

Is it possible to make it work with prompt engineering ?

How to integrate symbolic logic into large language models ?





Chat GPT在2022高考全国卷客观题的表现

	得分率	准确率
语文	-	-
英语	93.0 / 95	56/60
地理	24/36	6/9
政治	44/48	11/12
历史	24/40	6/10
数学 (理)	30/50	6/10
数学 (文)	35/55	7/11
物理	0/12	0/2
化学	6/42	1/7
生物	18/30	3/5
合计	274/408=0.67	96/126=0.76

客观题：在排除了坏样本（带图题,听力题等）之后，在全部126个样本数可以达到76%的准确率和67%的得分率，其客观题能力与两名500分左右的高考生（文科和理科各一名）相当。

主观题：对Chat GPT结果的人工打分，在文科综合（历史，地理，政治）上取得了不错的成绩（得分率78%），在生物可以达到50%的得分率，而在数学，物理，化学，历史上则表现不佳（得分率低于30%）。

高 端 人 工 智 能 学 院 学 术 前 沿 讲 座

大型语言模型的能力分析与应用

报告时间：2023年2月17日（周五）10:00-12:00
会议地点：立德楼18层多功能厅

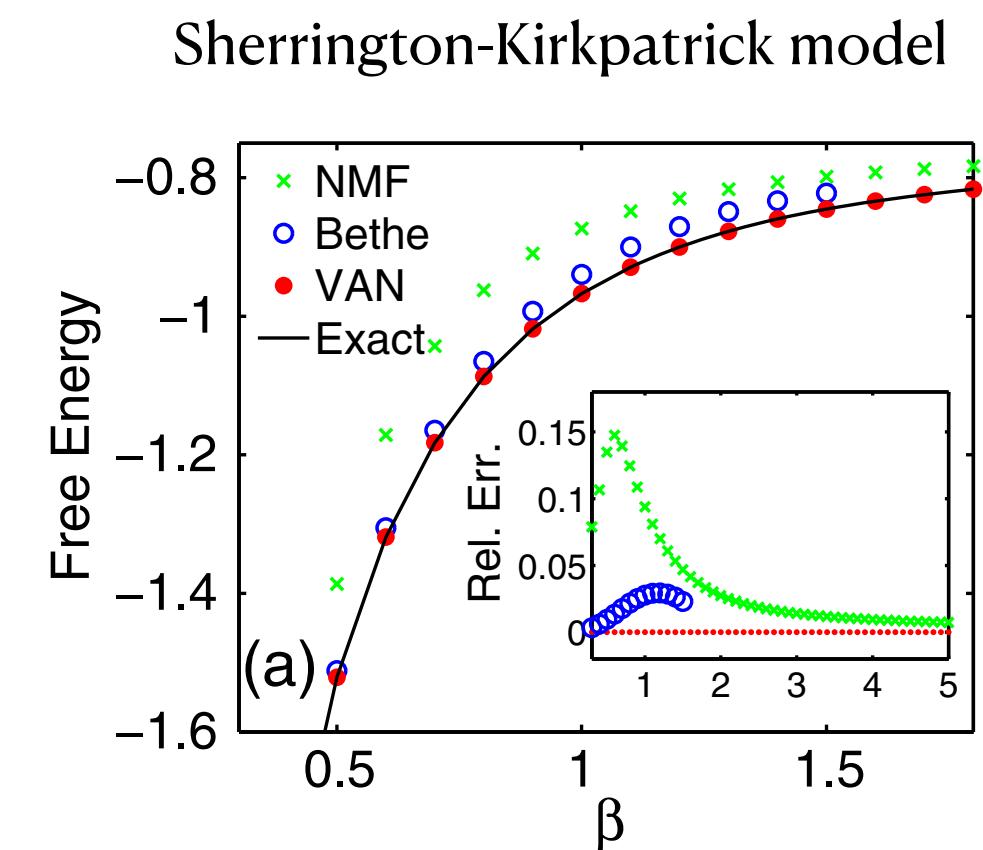
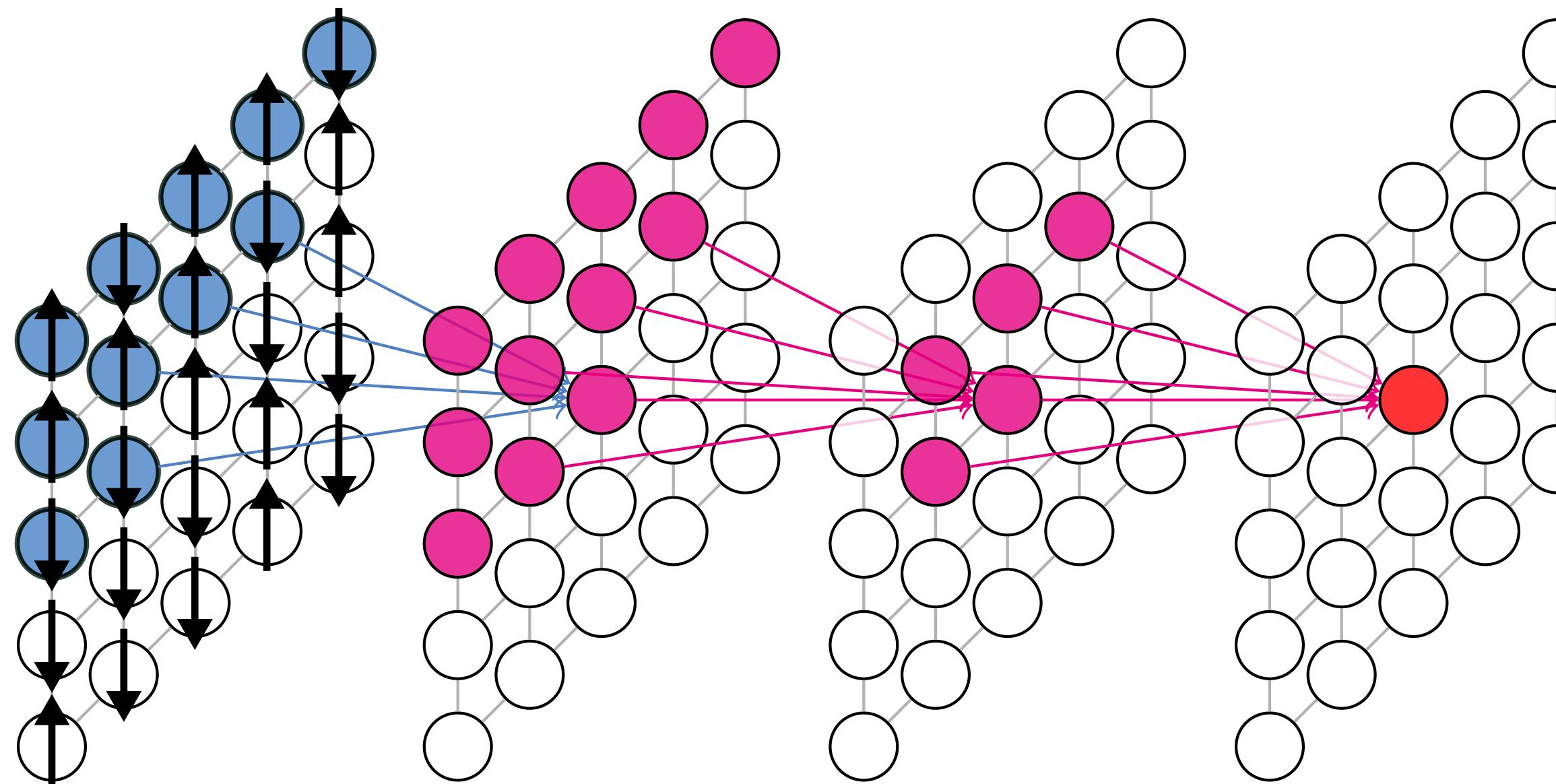
主讲嘉宾：邱锡鹏 复旦大学计算机学院 教授

担任中国中文信息学会理事、上海市计算机学会自然语言处理专委会主任等，主要研究方向为自然语言处理基础技术和基础模型，发表CCFA/B类论文80余篇，被引用1万余次，获得ACL 2017杰出论文奖（CCFA类）、CCL 2019最佳论文奖、《中国科学：技术科学》2021年度高影响力论文奖，有5篇论文入选ACL/EMNLP等会议的最有影响力论文，主持开发了开源框架FudanNLP和FastNLP，已被国内外数百家单位使用，发布了CPT、BART-Chinese、ElasticBERT等中文预训练模型。在中文模型中下载量排名前列。曾获中国科协青年人才托举工程项目、国家优青项目、科技创新2030“新一代人工智能”重大项目课题等，2020年获第四届上海高校青年教师教学竞赛一等奖，2022年获首届上海市计算机学会教学成果奖一等奖（第一完成人），2022年获钱伟长中文信息处理科学技术奖一等奖（第一完成人）。培养学生多人次获得国家一级学会优博。

报告摘要：

2022年底，美国OpenAI公司发布了ChatGPT，一个可以与人类对话交互的千亿规模参数的大型语言模型。它可以根据用户输入的指令完成各种语言相关的任务，例如写文章、写代码、回答问题、日常聊天等等，能够极大地提高人们的生产力，这归功于它所表现出的强大意图理解能力、流畅的对话能力和丰富的世界知识。ChatGPT展示了研发通用人工智能助手广阔的研究和应用前景。目前，ChatGPT的技术细节及模型参数均未开源。本报告主要介绍大型语言模型的能力评测、能力演化路线分析以及如何在下游任务中更高效利用大模型的能力。

Variational autoregressive networks for solving statistical mechanics problems



Wu, LW, Zhang, PRL '19 <https://github.com/wdphy16/stat-mech-van>

Direct sampling, tractable density, while still being expressive

Quantum states: Sharir, Levine, Wies, Carleo, Shashua, PRL '20
Hibat-Allah, Ganahl, Hayward, Melko, Carrasquilla, PRResearch '20

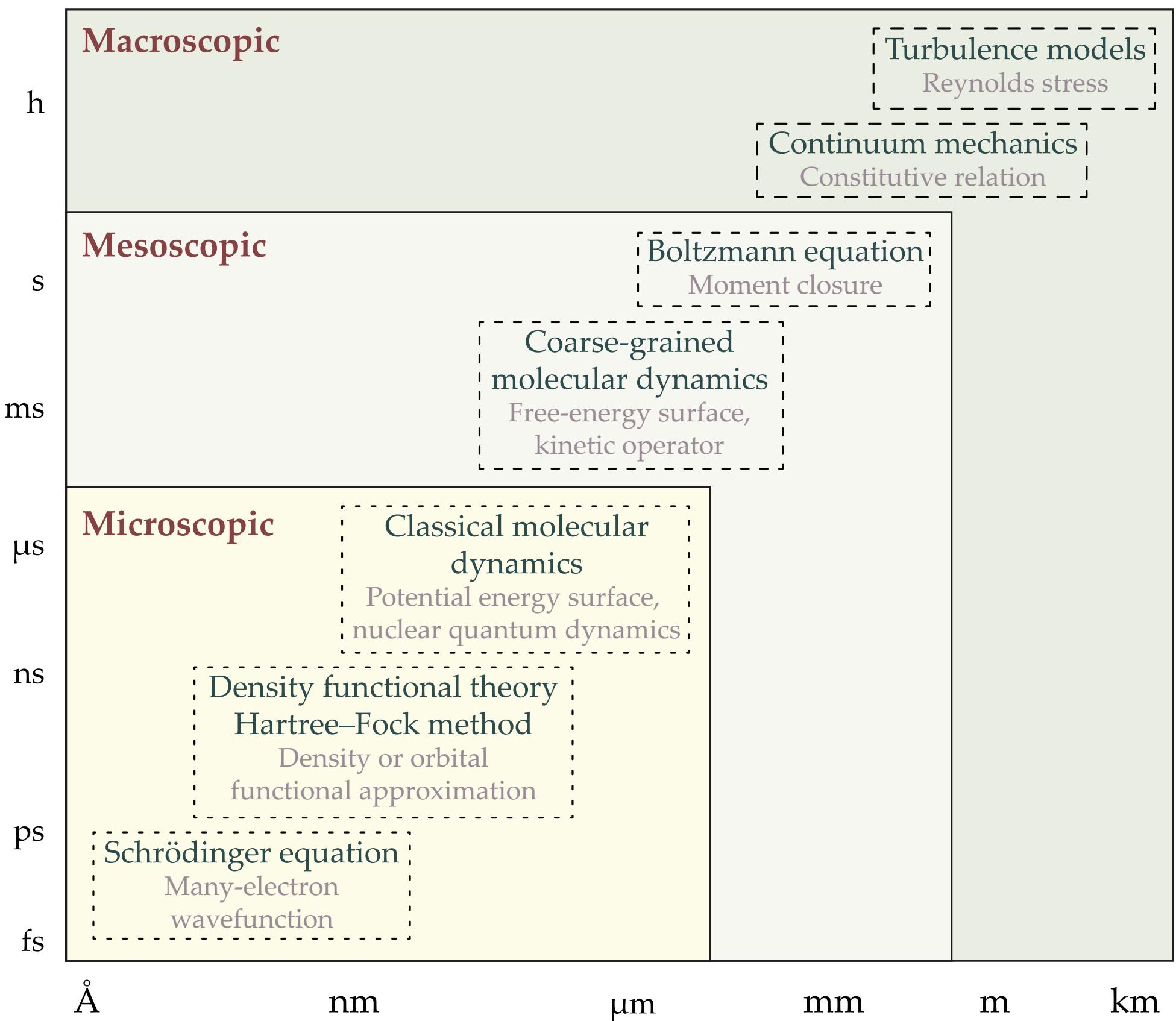
Hybrid quantum algorithm: Liu, Mao, Zhang, LW, MLST '21

Monte Carlo update: Wu, Rossi, Carleo, 2105.05650

Material application: Damewood et al , 2107.05109

Deep variational free-energy in the context

E, Han,Zhang, Physics Today 2020



Objective	Model	Data	Task
MD potential energy surface	$3N$ -dim function	DFT energy/force	Generalization
DFT xc energy functional	3-dim functional	QMC/CCSD/...	
Variational free-energy	$3N$ -dim functional	No	Optimization

Scientific machine learning with data

Scientific
computation

Machine learning
aided simulations

Machine learning
scientific data



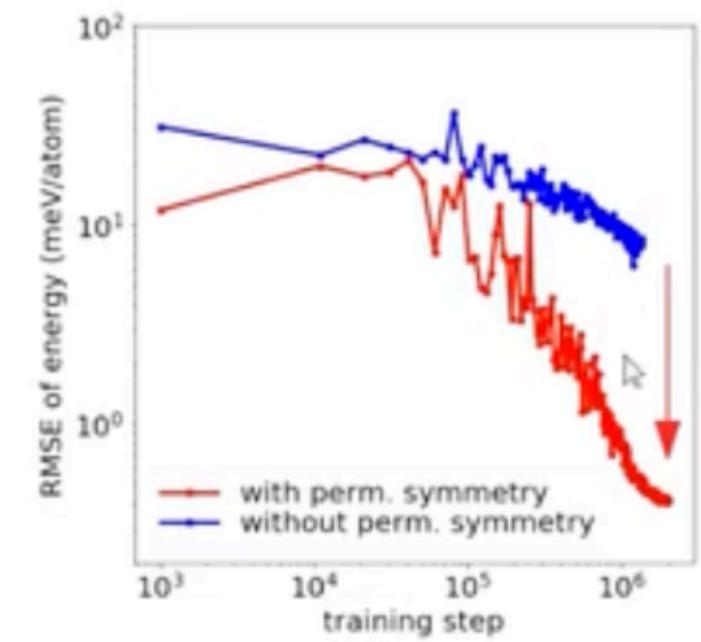
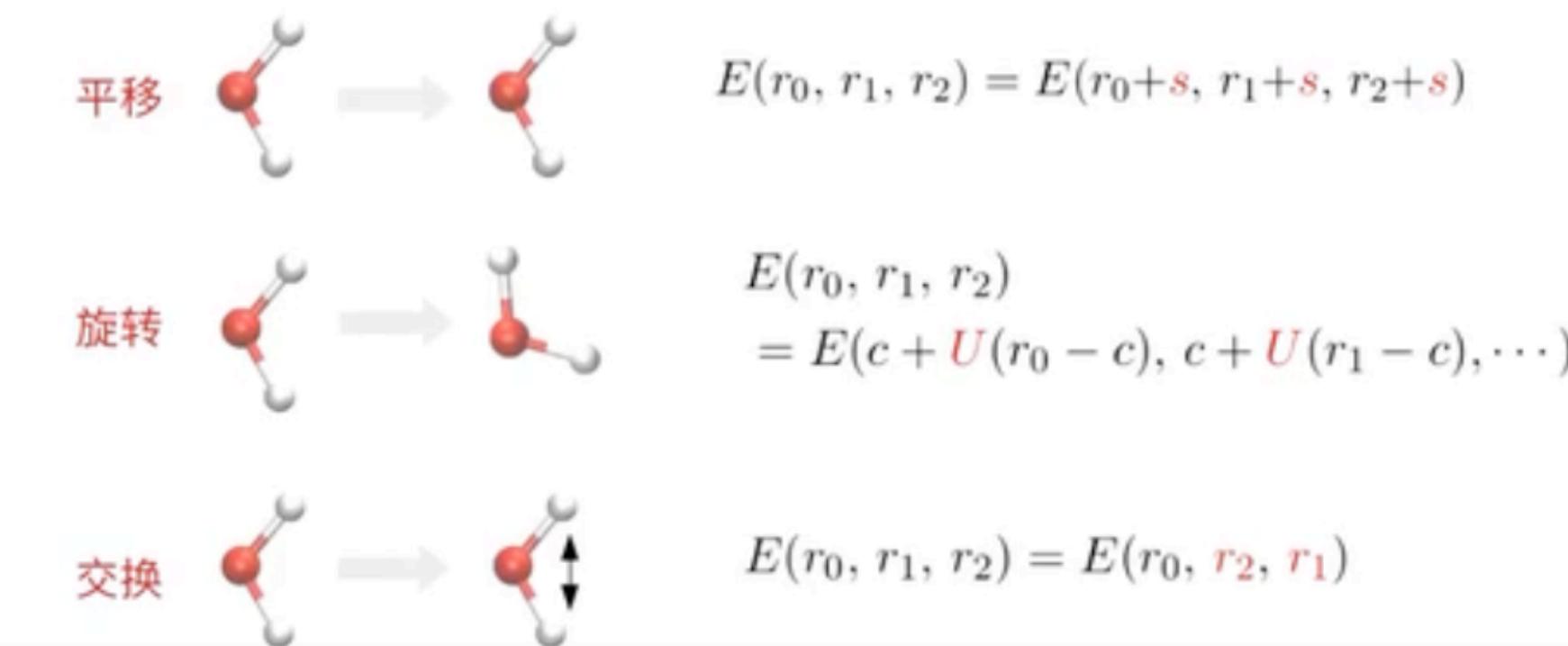
**Use your data wisely;
Physical inductive bias: locality, symmetry, compositability...**

Symmetry as a key inductive bias

Linfeng Zhang et al, NIPS 2018

End-to-end Symmetry Preserving Inter-atomic Potential Energy Model for Finite and Extended Systems

Invariant features



王涵 基于深度学习的分子动力学模拟

https://www.bilibili.com/video/BV14L411E7nf/?spm_id_from=333.999.o.o

Batzner et al, Nature Comm. 2022

E(3)-equivariant graph neural networks for data-efficient and accurate interatomic potentials
Equivariant transformations

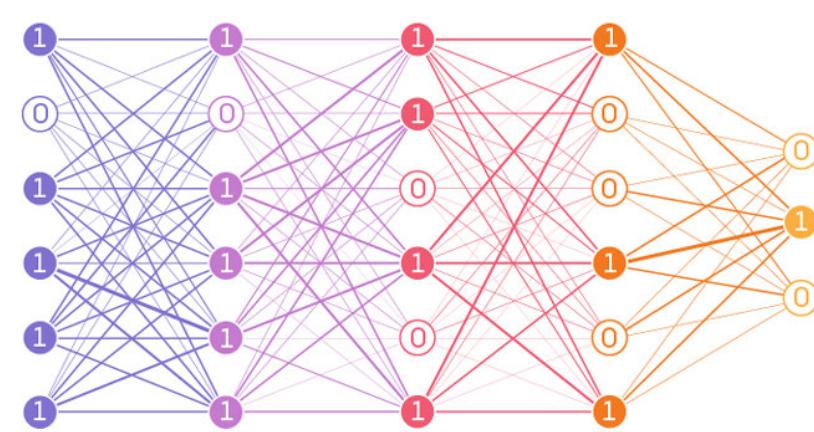
tories that were used in the earlier work¹⁵. Table 3 compares the energy and force errors of NequIP trained on the 133 structures vs DeepMD trained on 133,500 structures. We find that with 1000x fewer training data NequIP significantly outperforms DeepMD on all four parts of the data set in the error on the force components. We note that there are $3N$ force components for

Scientific machine learning without data

Variational Principle

$$S = \int d^4x \sqrt{g} \left[\frac{1}{2} R \right.$$
$$\left. + \frac{1}{2} F^2 + \nabla \bar{\psi} \not{D} \psi \right.$$
$$\left. + (\not{D}_\mu \varphi)^2 + V(\varphi) \right]$$
$$\left. + \bar{\psi} \psi \right]$$

Representation



Optimization

