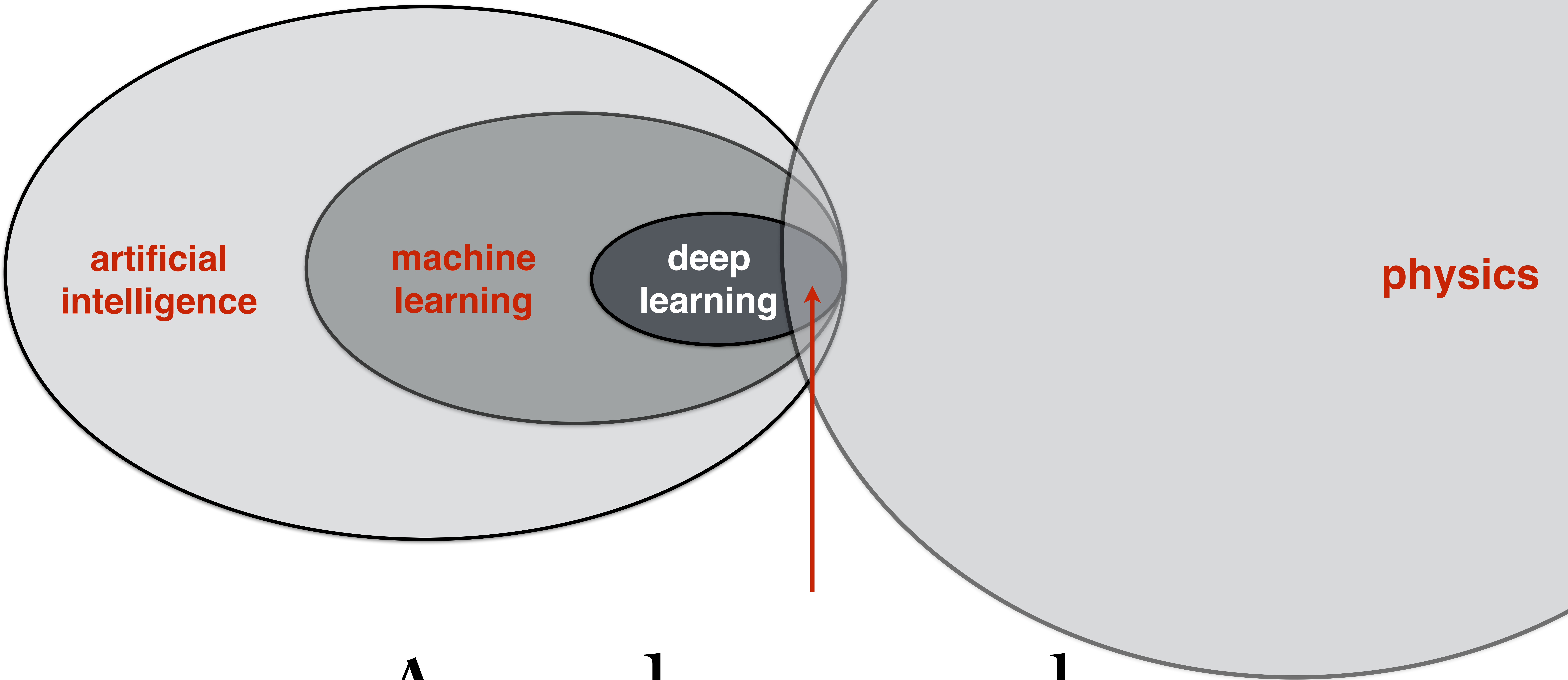
深度学习的数学与物理

王磊
中科院物理研究所 T03 组
https://wangleiphy.github.io



A random sample, and some thoughts

A research story

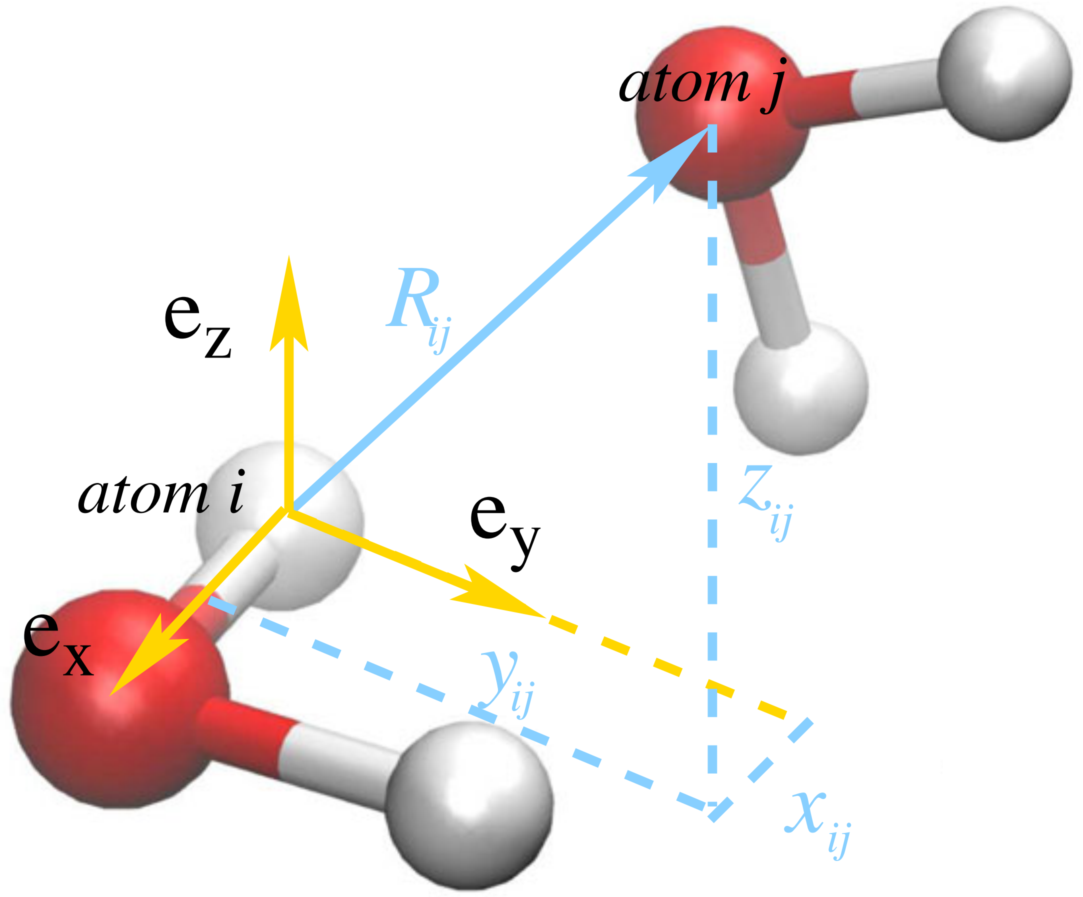
Monge-Ampère Flow for Generative Modeling

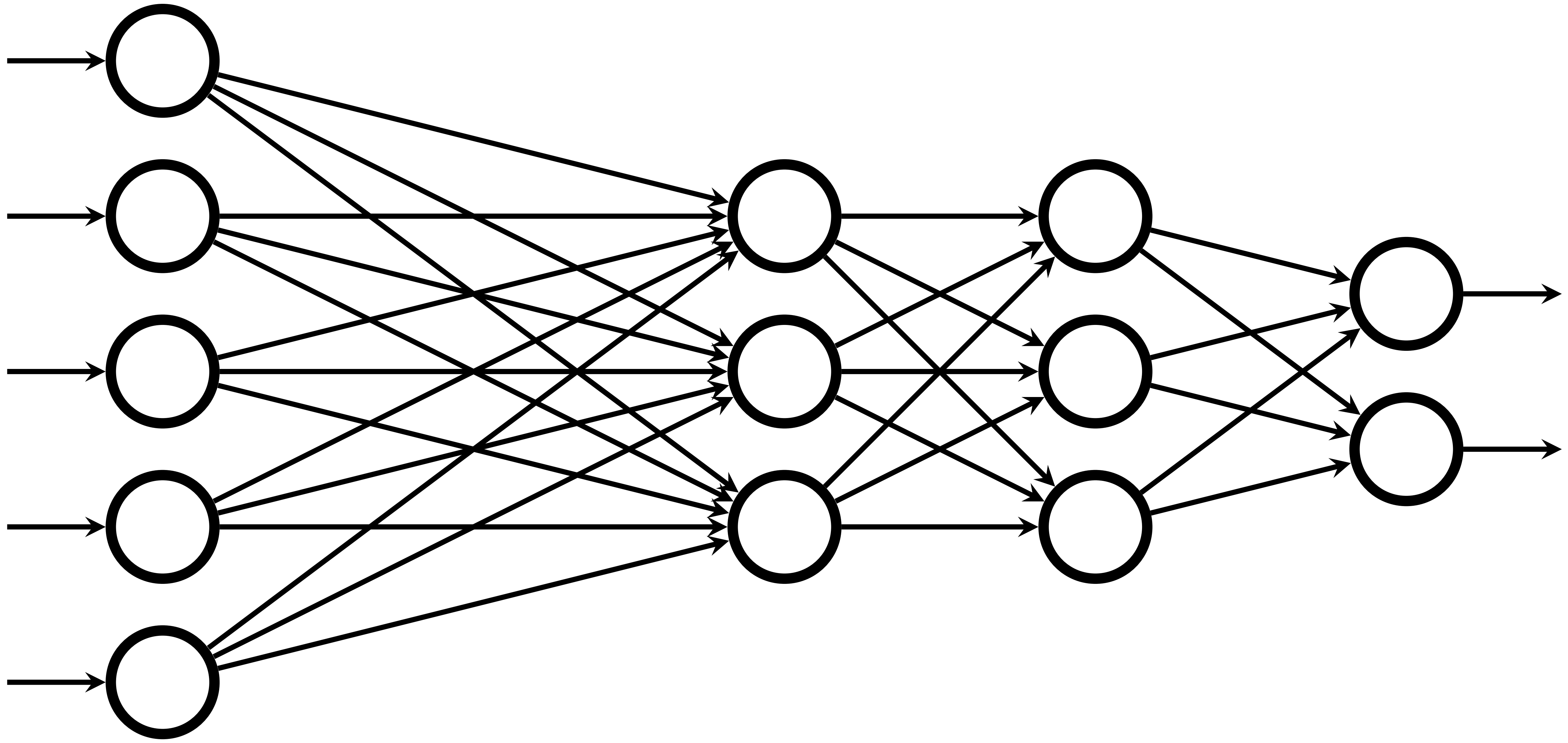
张林峰

鄂维南

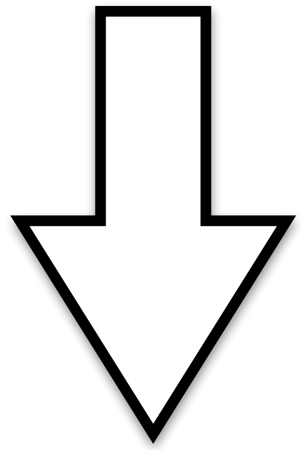
arXiv:1809.10188 https://github.com/wangleiphy/MongeAmpereFlow

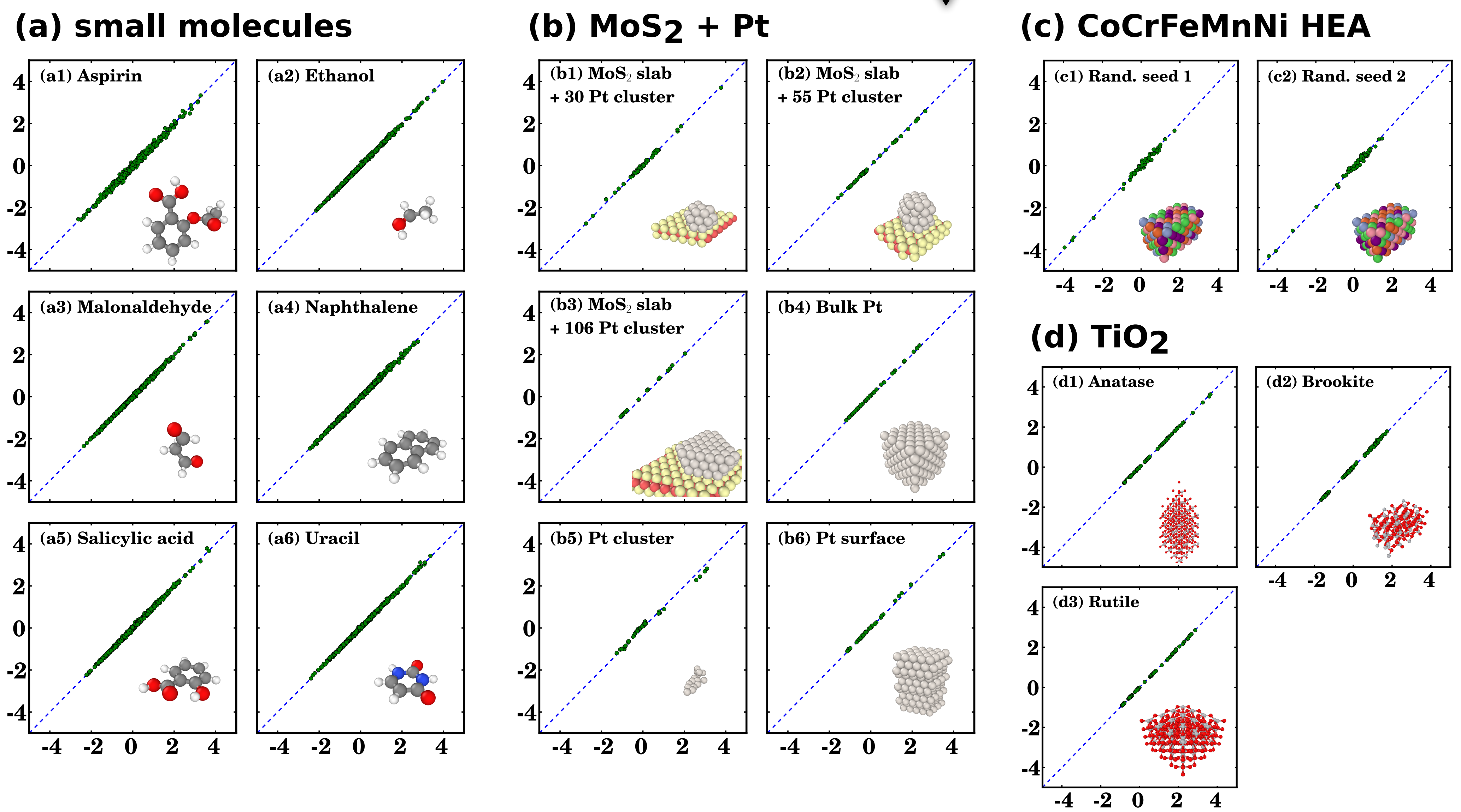
林峰的主业：机器学习分子势能





总能量、力…

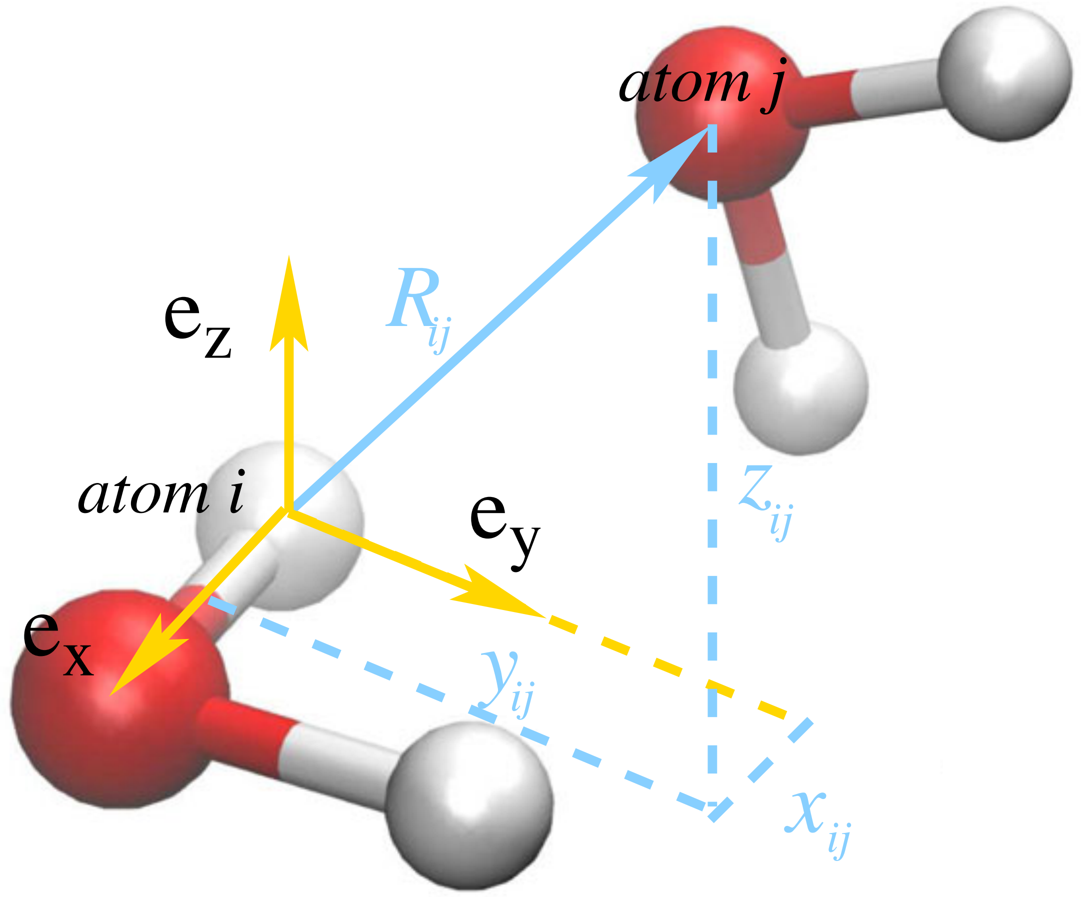


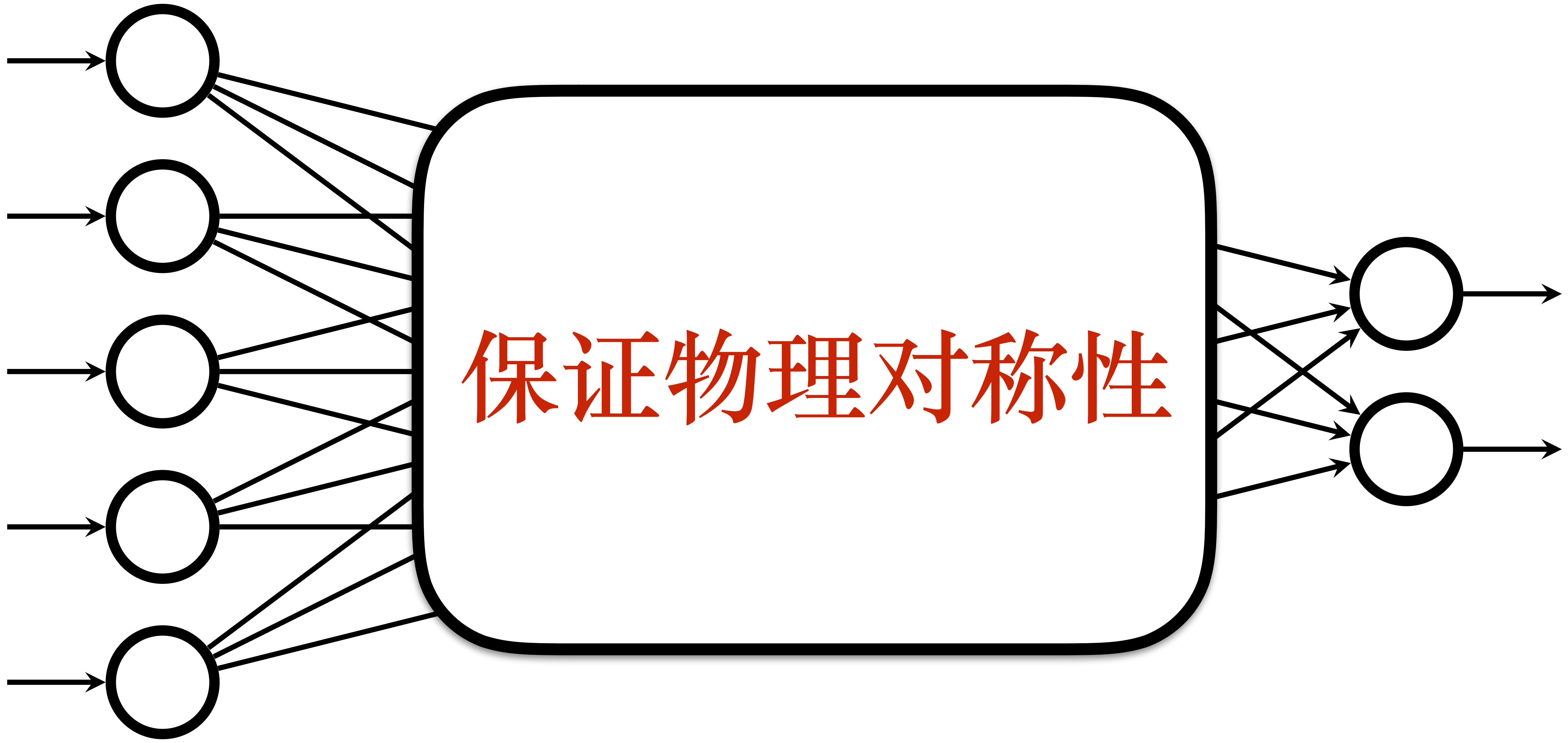


原子种类、坐标…

Zhang, Han, Wang, Car, E, PRL ‘18
Zhang, Han, Wang, Saidi, Car, E, NIPS ’18

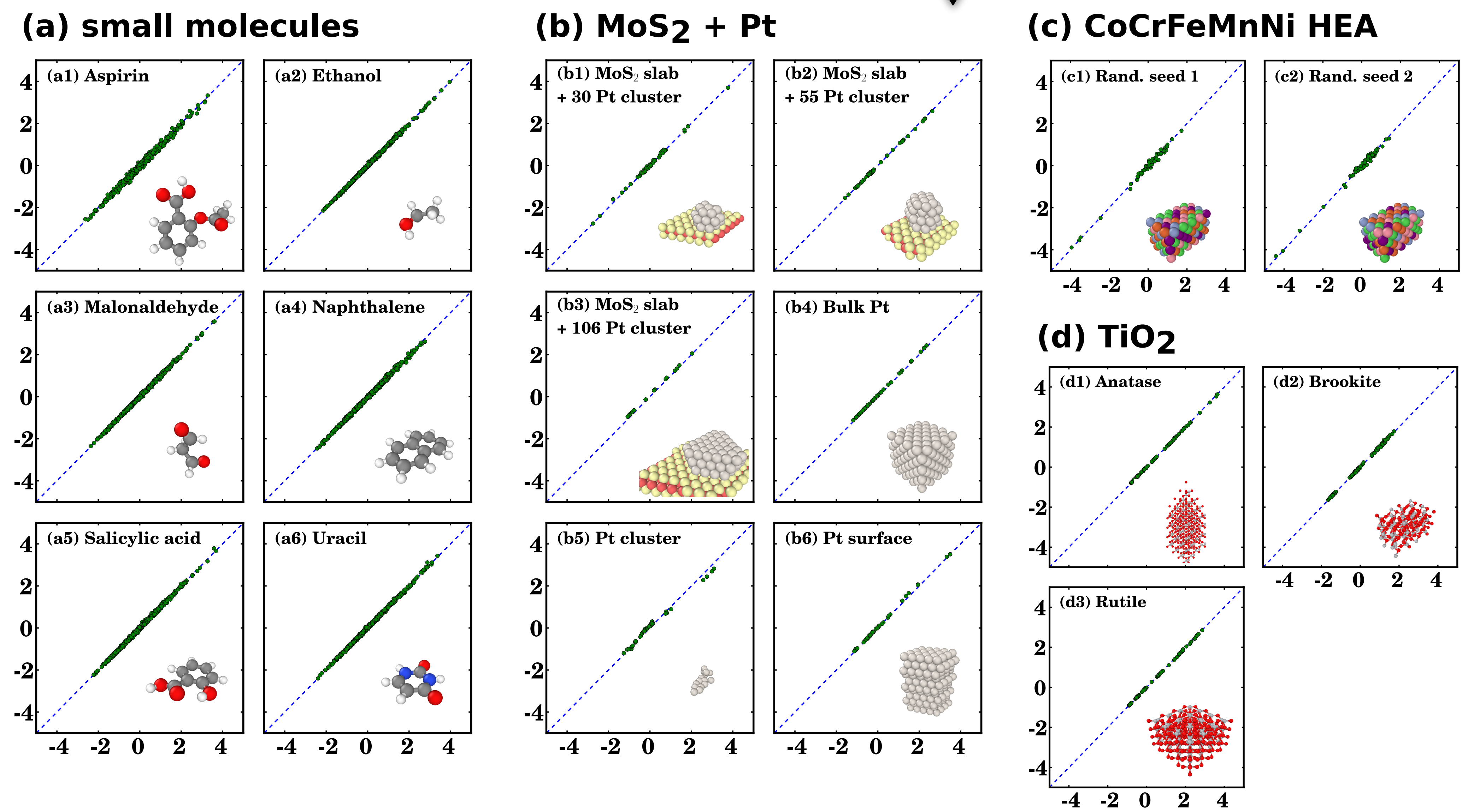
林峰的主业：机器学习分子势能





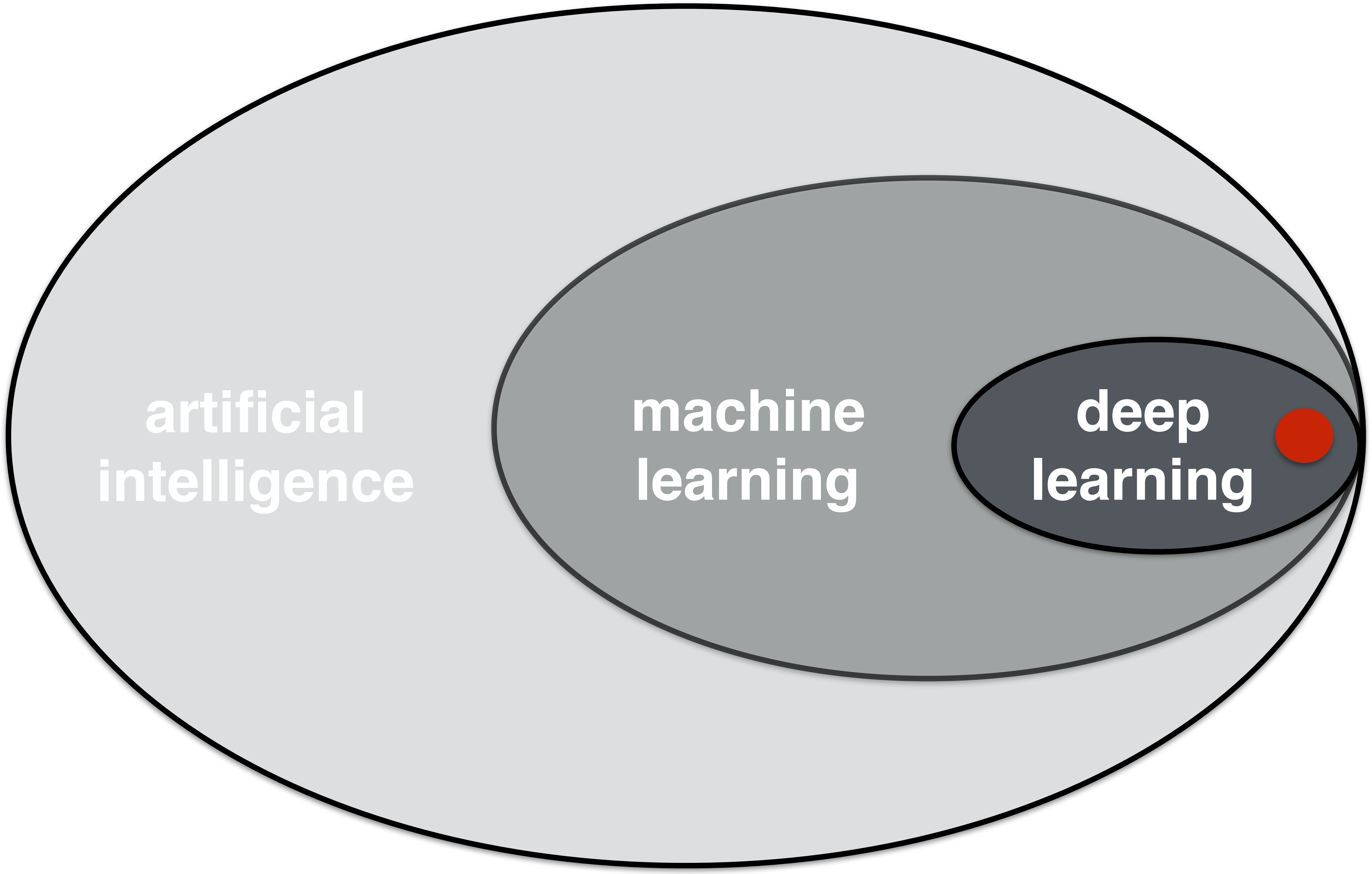
总能量、力…

原子种类、坐标…



Zhang, Han, Wang, Car, E, PRL ‘18

Zhang, Han, Wang, Saidi, Car, E, NIPS ’18



Hi， 林峰！

如何确保**生成型网络**的对称性？

深度学习不仅是函数拟合

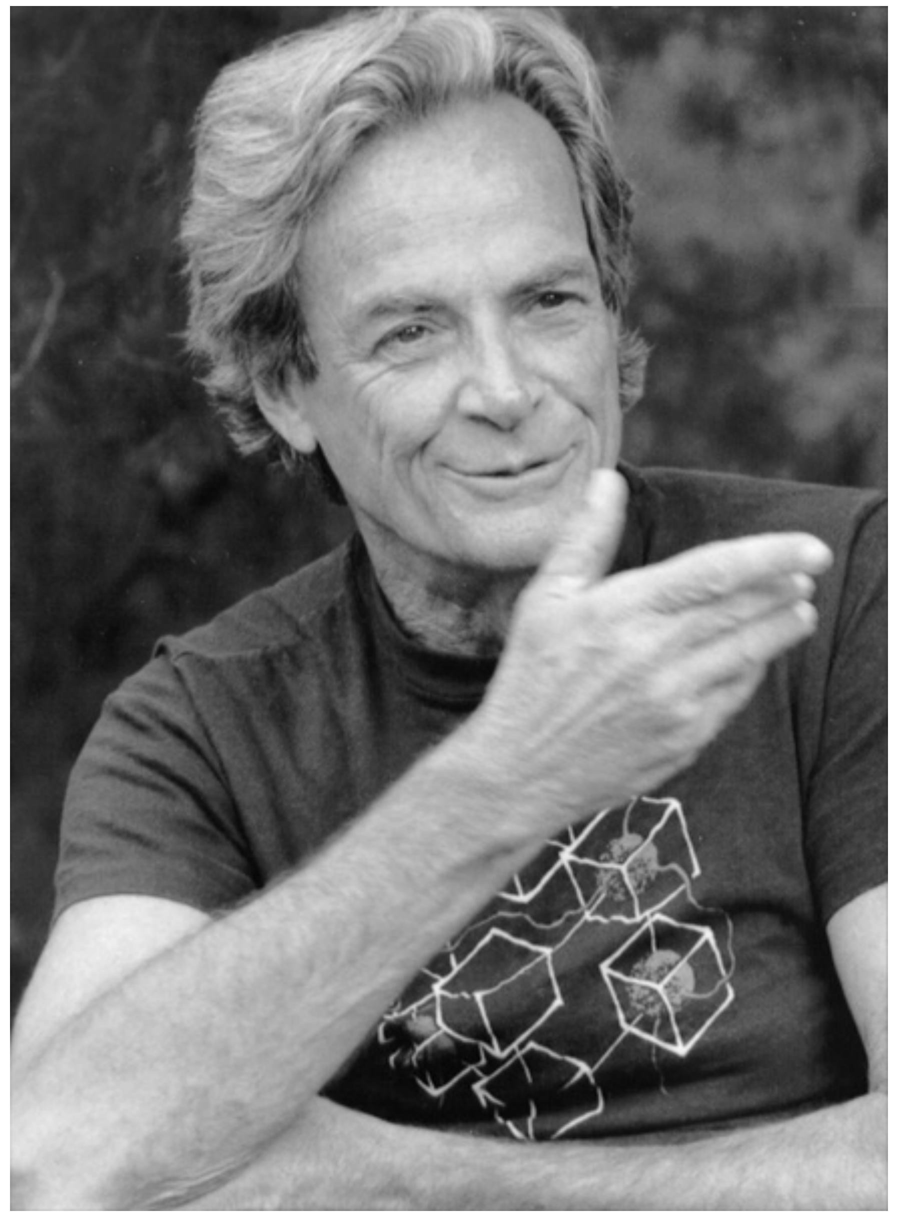
“判别型”学习

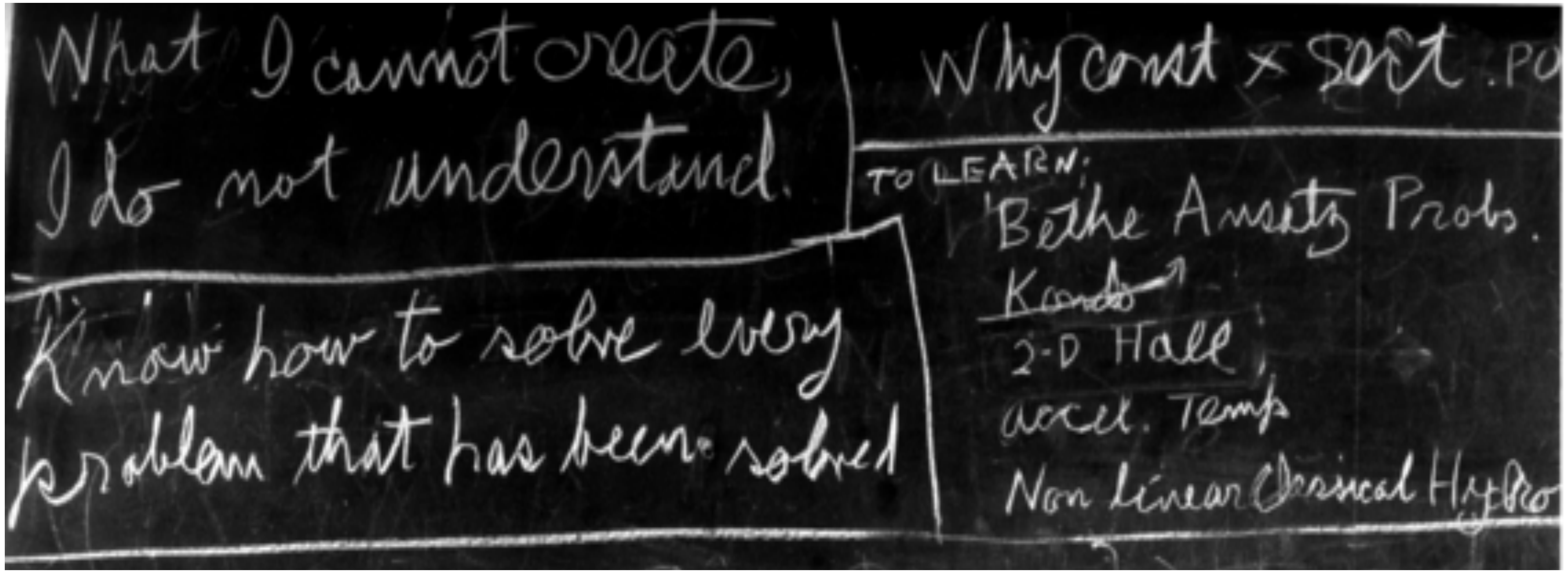
$y = f(\boldsymbol{x})$ or $p(y|\boldsymbol{x})$

“生成型”学习

$p(\boldsymbol{x}, y)$

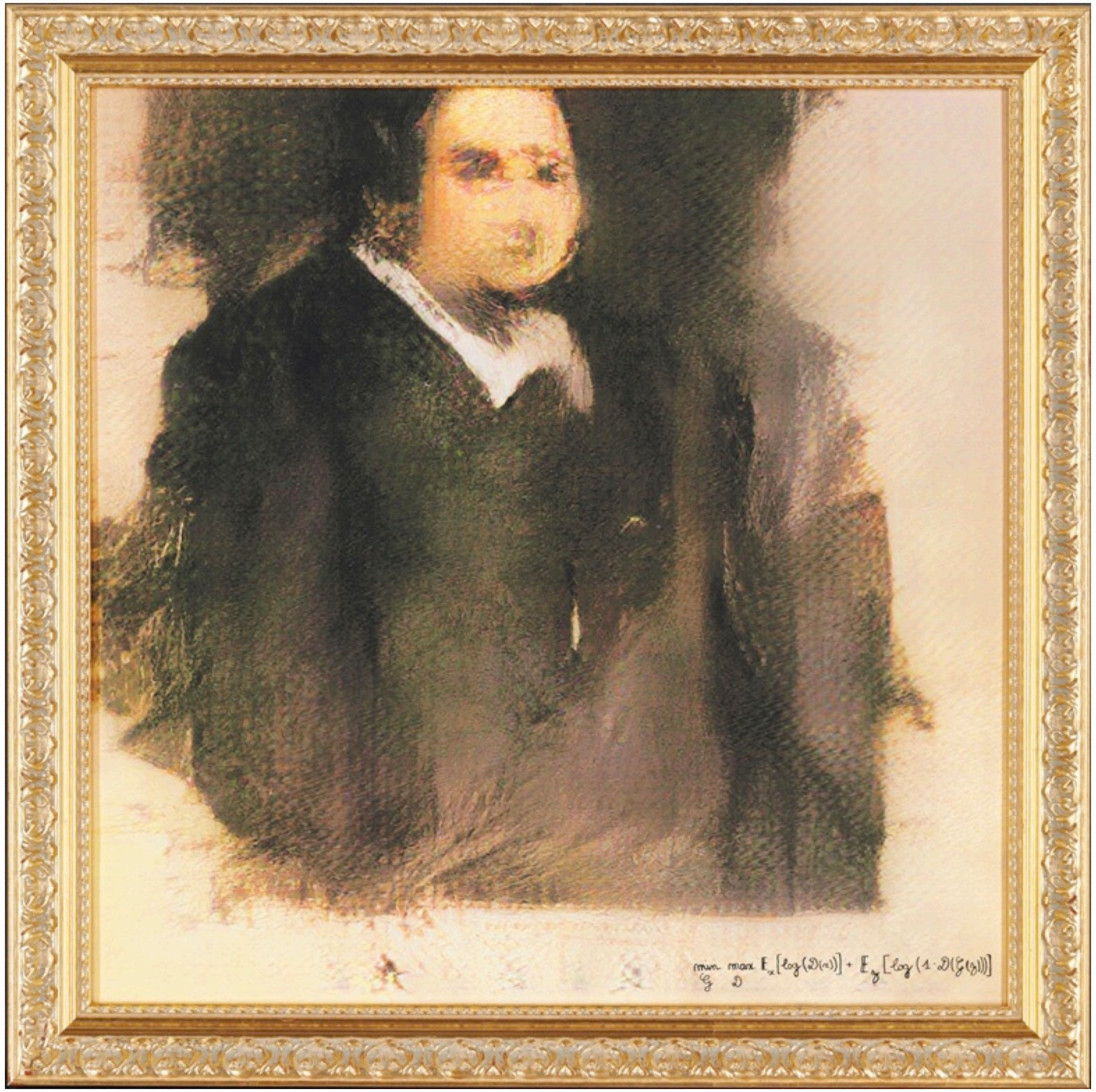
深度学习不仅是函数拟合





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

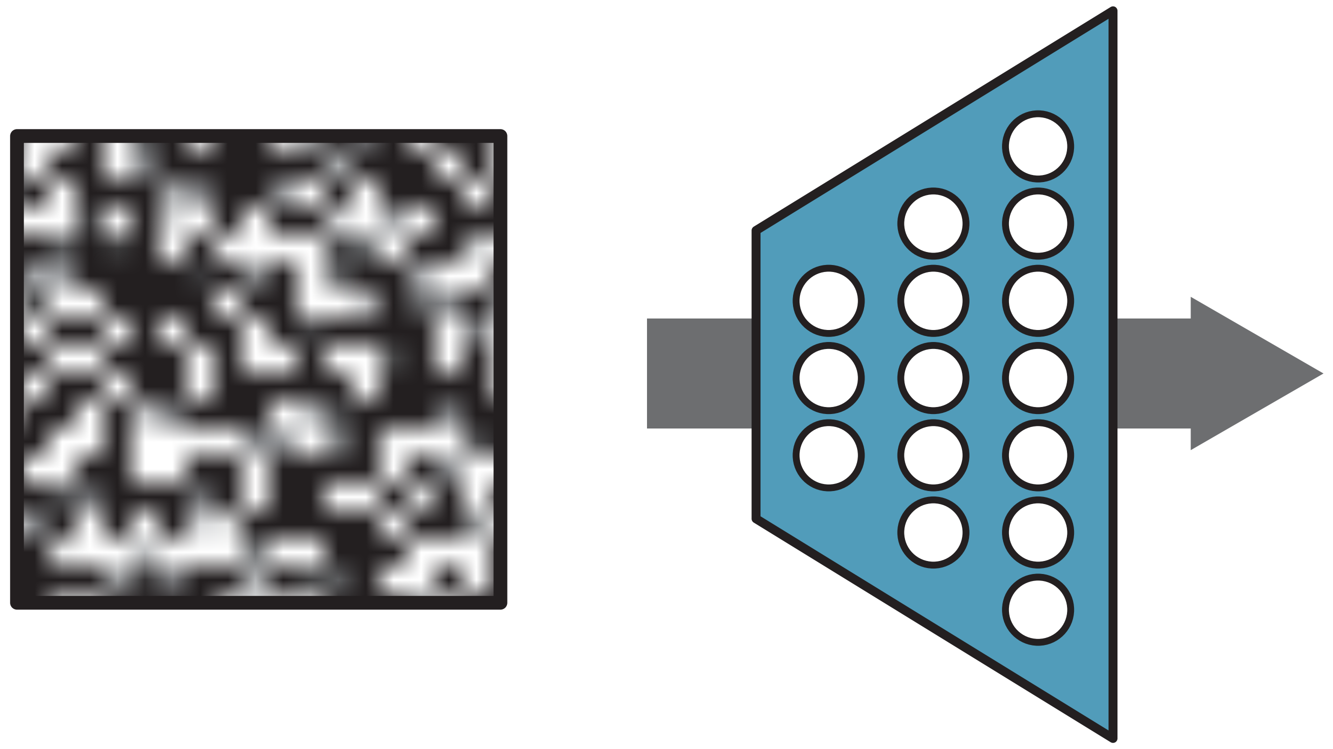
生成“艺术品”



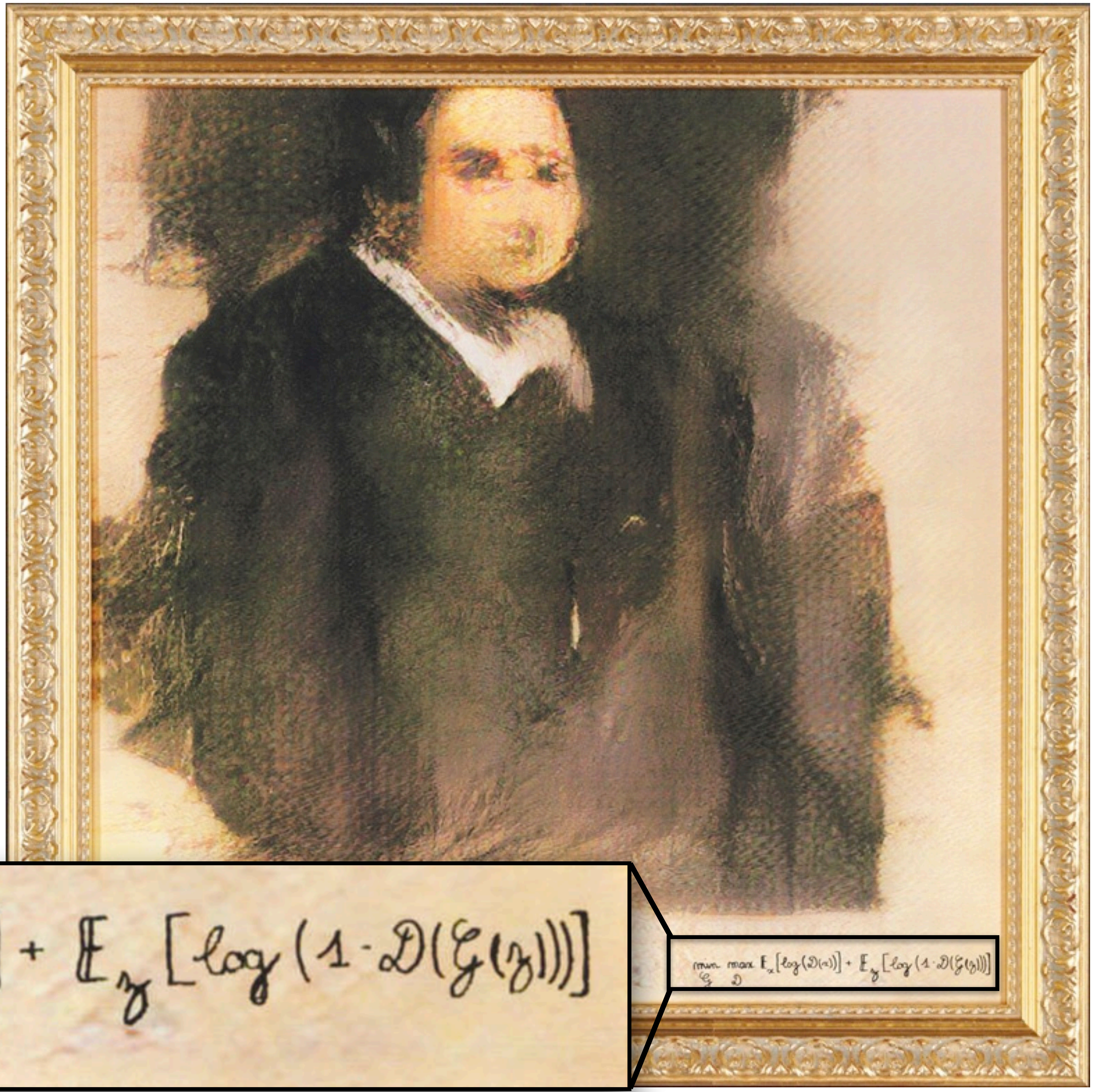
$432,500
佳士得 纽约
2018.10.25

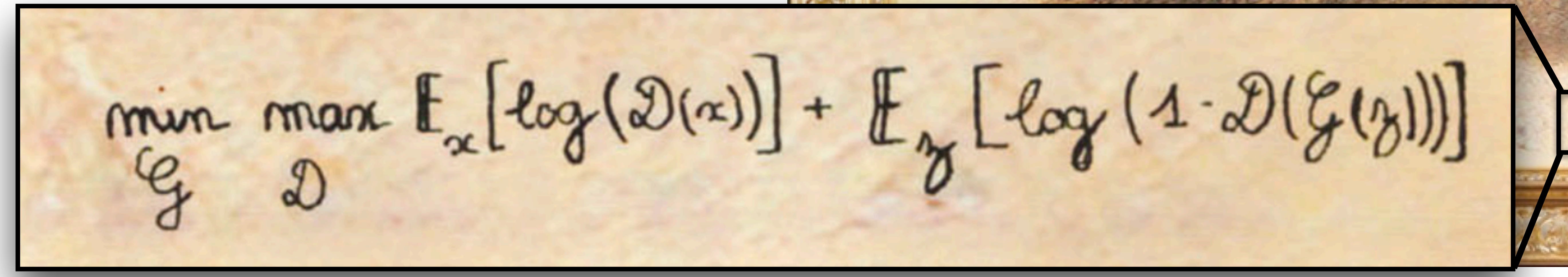
https://www.christies.com/Features/A-collaboration-between-two-artists-one-human-one-a-machine-9332-1.aspx

生成“艺术品”



高斯噪音　　生成网络





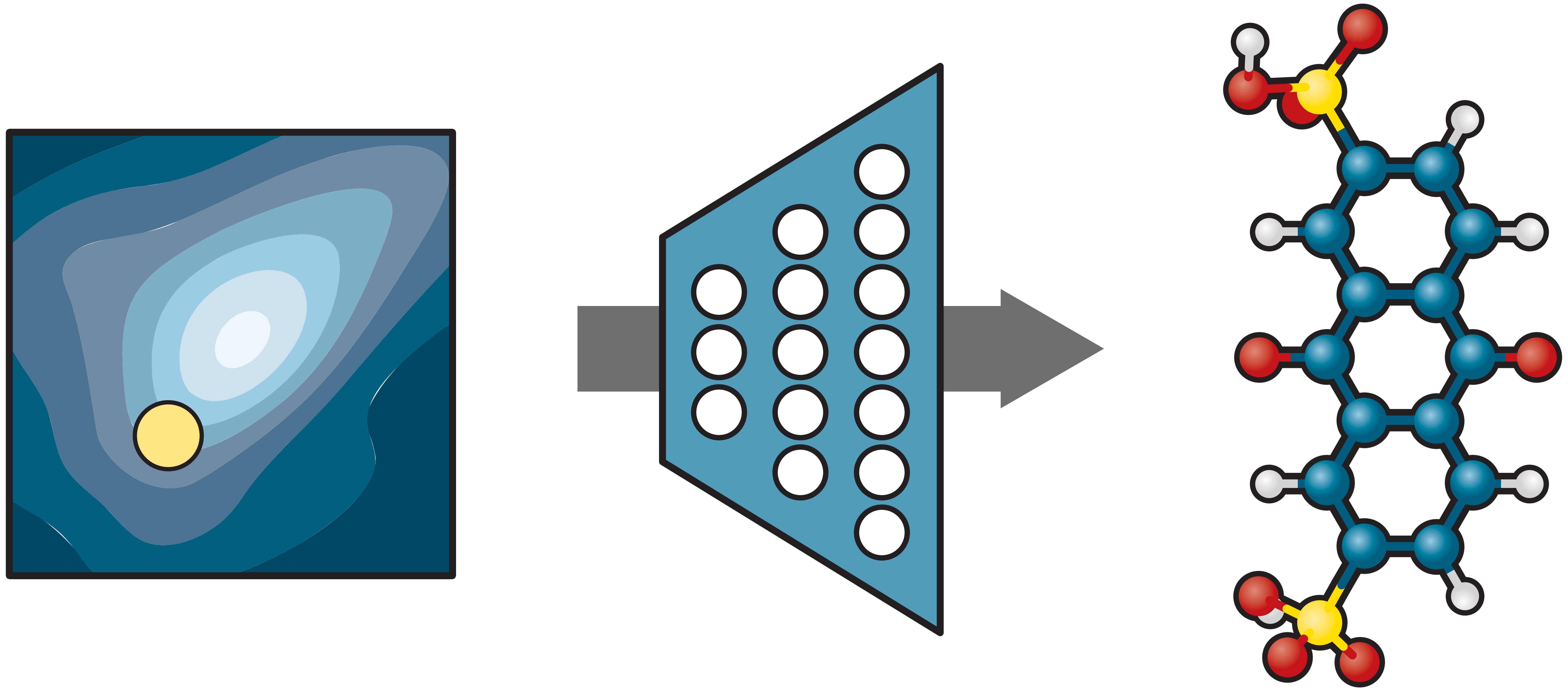
$$\min_{G}\max_{D} \mathbb{E}_x[\log(D(x))] + \mathbb{E}_z[\log(1-D(G(z)))]$$

$432,500
佳士得 纽约
2018.10.25

https://www.christies.com/Features/A-collaboration-between-two-artists-one-human-one-a-machine-9332-1.aspx

生成化学分子

隐变量
（属性）



物理变量
（微观构型）

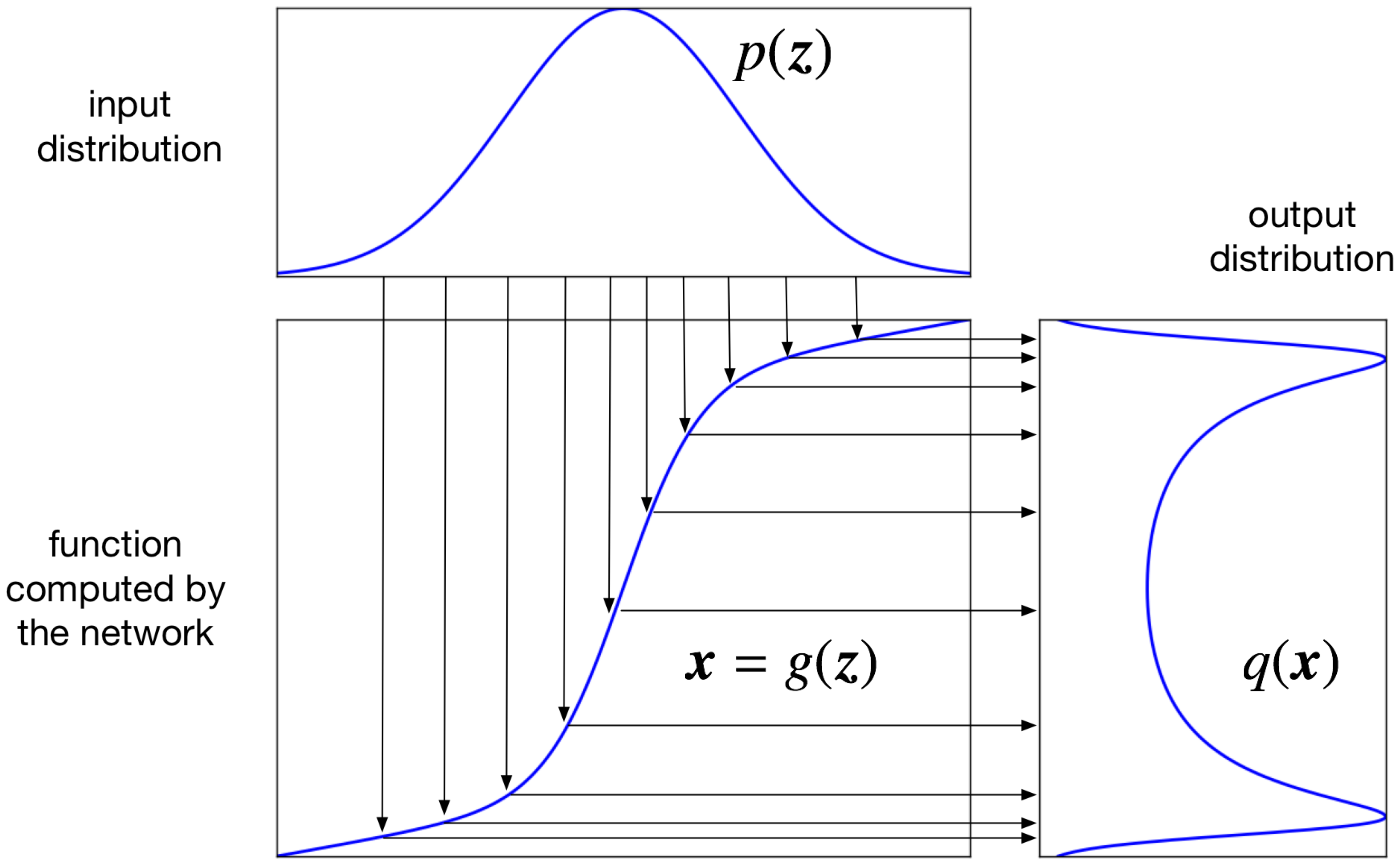
生成模型
背后的数学
“概率变换”

简单分布 ⇄ 复杂分布

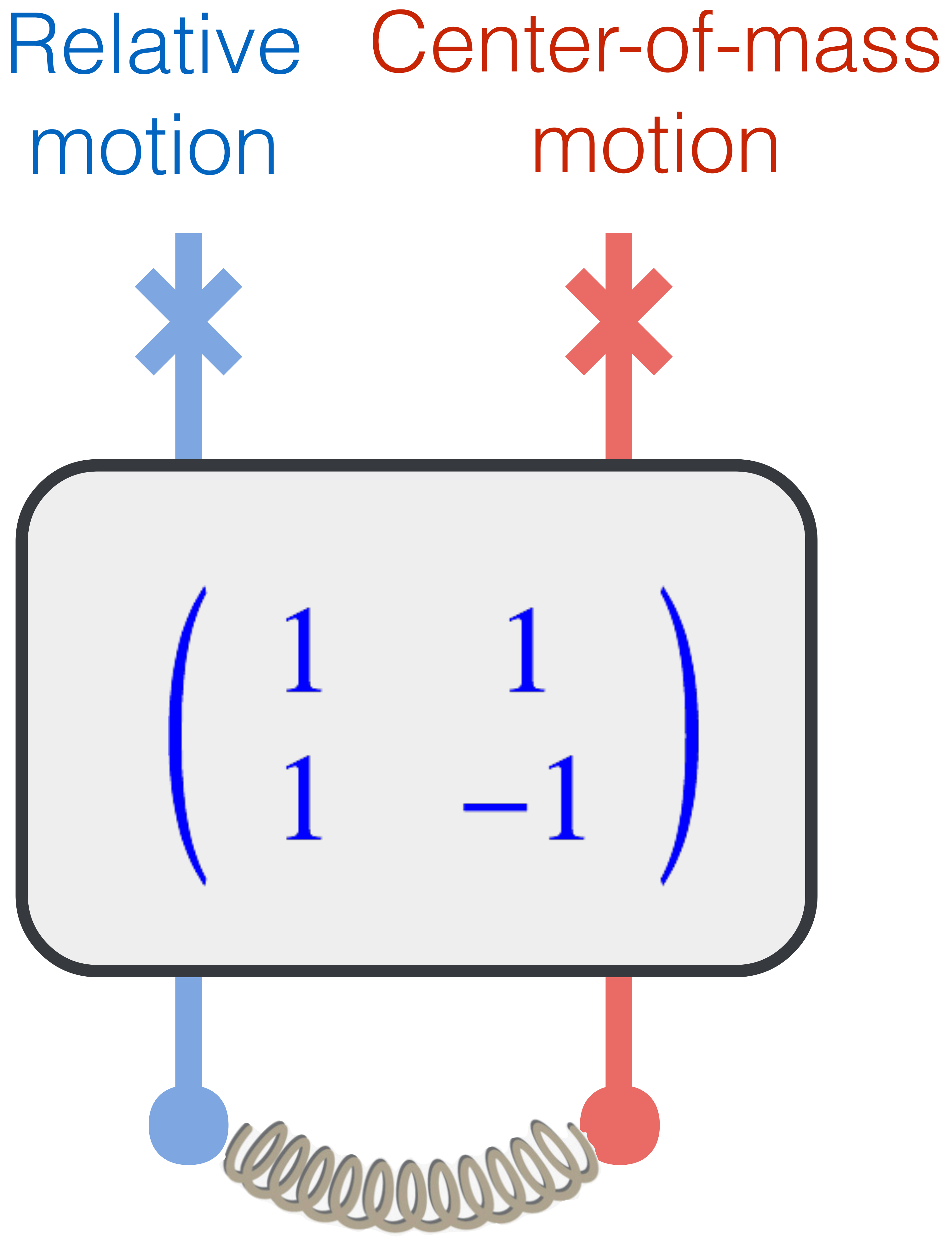
生成（→）
推断（←）

Sanchez-Lengeling & Aspuru-Guzik, Science 2018

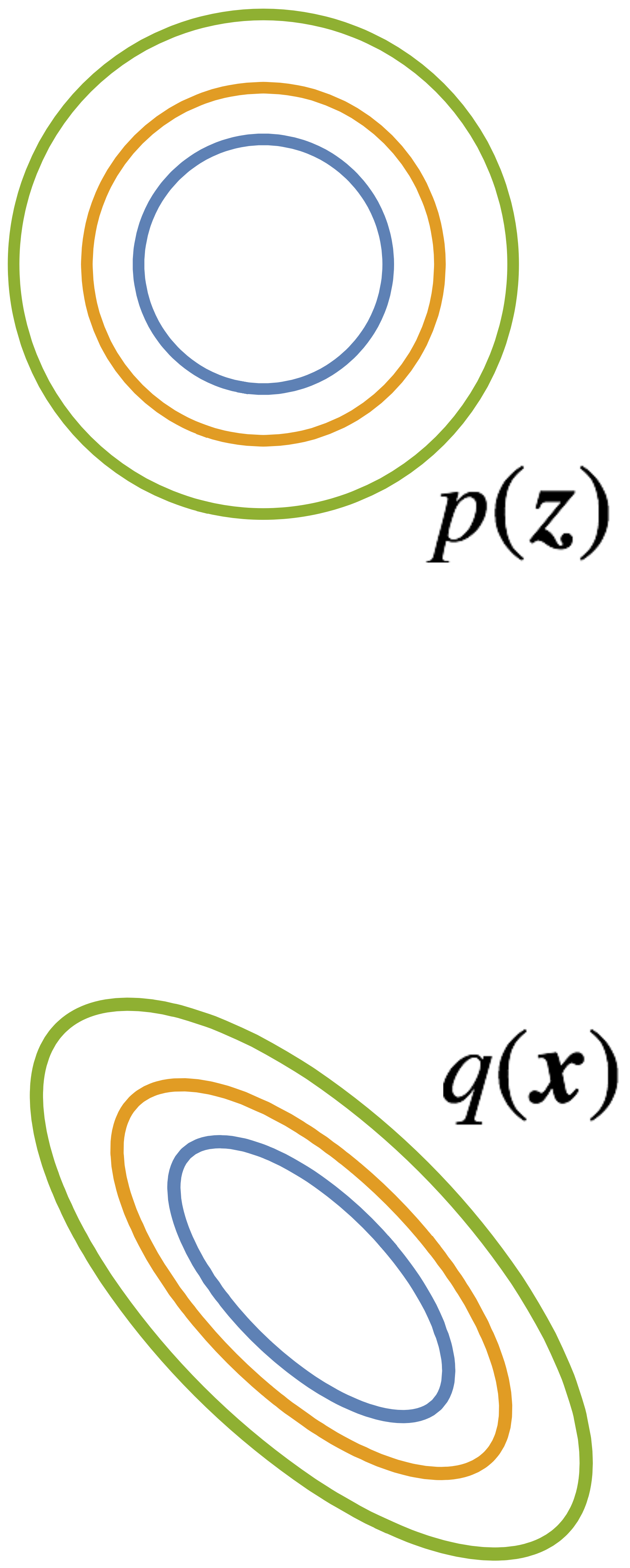
Probability transformation in picture



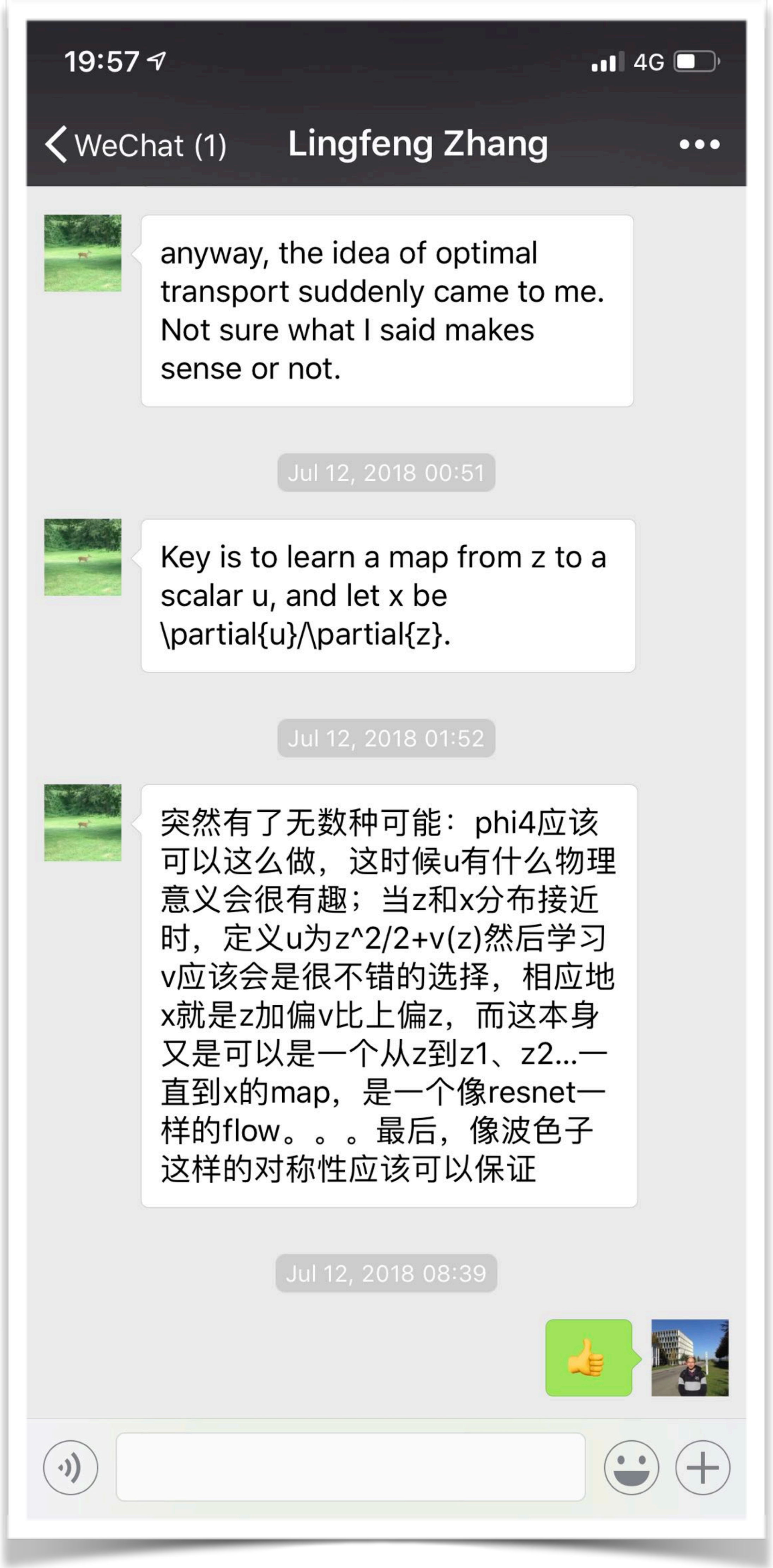
A Toy problem: Harmonic oscillator



Coupled harmonic oscillator

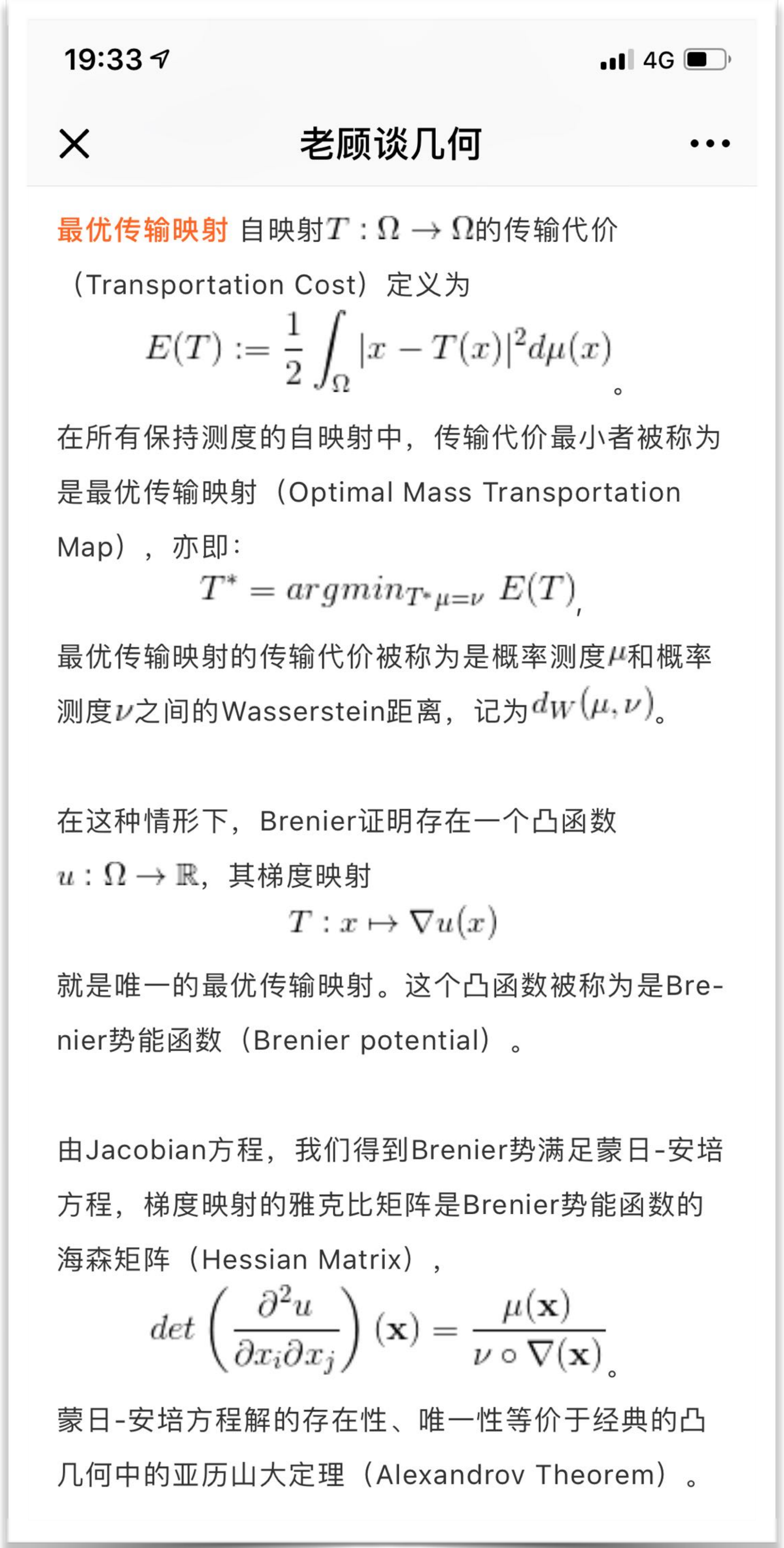


四个月后…



19:57
4G
WeChat (1)
Lingfeng Zhang
anyway, the idea of optimal transport suddenly came to me. Not sure what I said makes sense or not.
Jul 12, 2018 00:51
Key is to learn a map from z to a scalar u, and let x be \partial{u}/\partial{z}.
Jul 12, 2018 01:52
突然有了无数种可能：phi4应该可以这么做，这时候u有什么物理意义会很有趣；当z和x分布接近时，定义u为z^2/2+v(z)然后学习v应该会是很不错的选择，相应地x就是z加偏v比上偏z，而这本身又是可以是一个从z到z1、z2...一直到x的map，是一个像resnet一样的flow。。。最后，像波色子这样的对称性应该可以保证
Jul 12, 2018 08:39

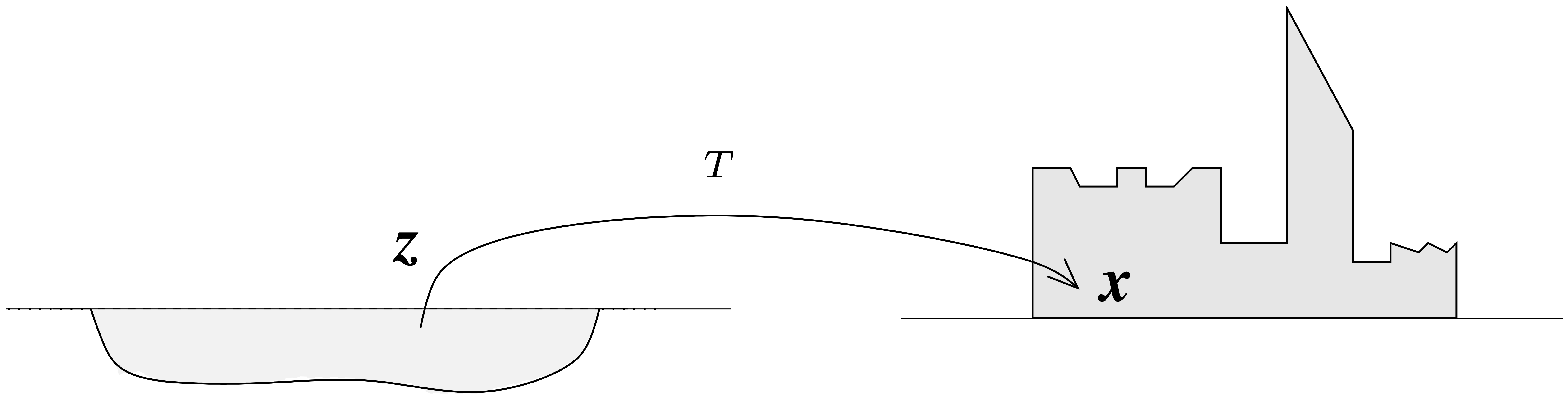
四个月后…



Lei, Su, Cui, Yau, Gu 1710.05488

Optimal Transport Theory

Monge problem (1781): How to transport earth with optimal cost ?



Monge Kantorovich Koopmans Dantzig Brenier Otto McCann Villani Figalli

Nobel Prize in Economics ’75

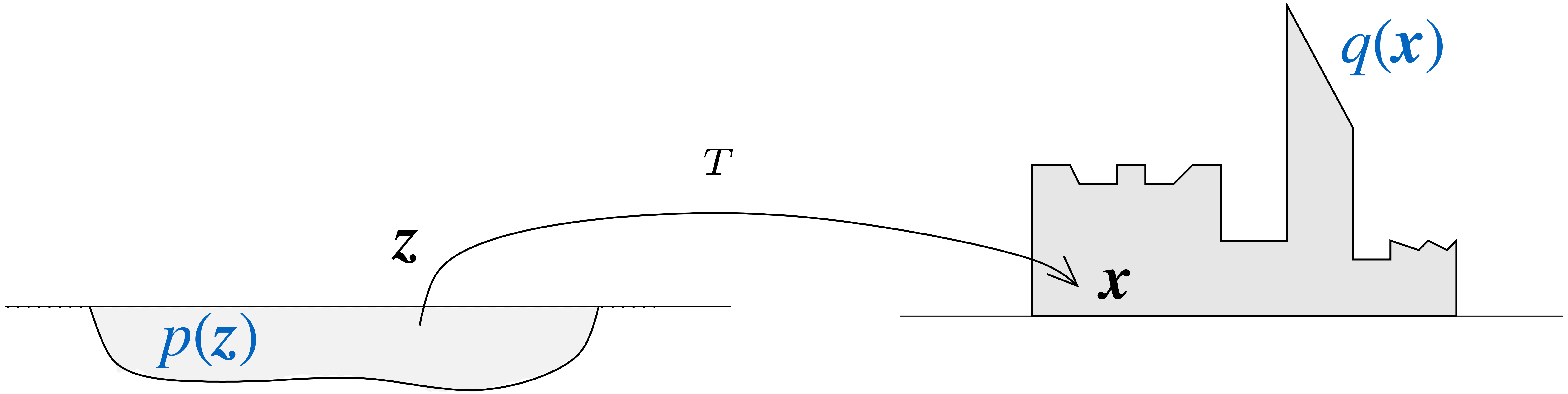
Fields Metal ’10

Fields Metal ’18

from Cuturi, Solomon NISP 2017 tutorial

Optimal Transport Theory

Monge problem (1781): How to transport earth with optimal cost ?




Monge


Kantorovich


Koopmans


Dantzig


Brenier


Otto


McCann


Villani


Figalli

Nobel Prize in Economics '75

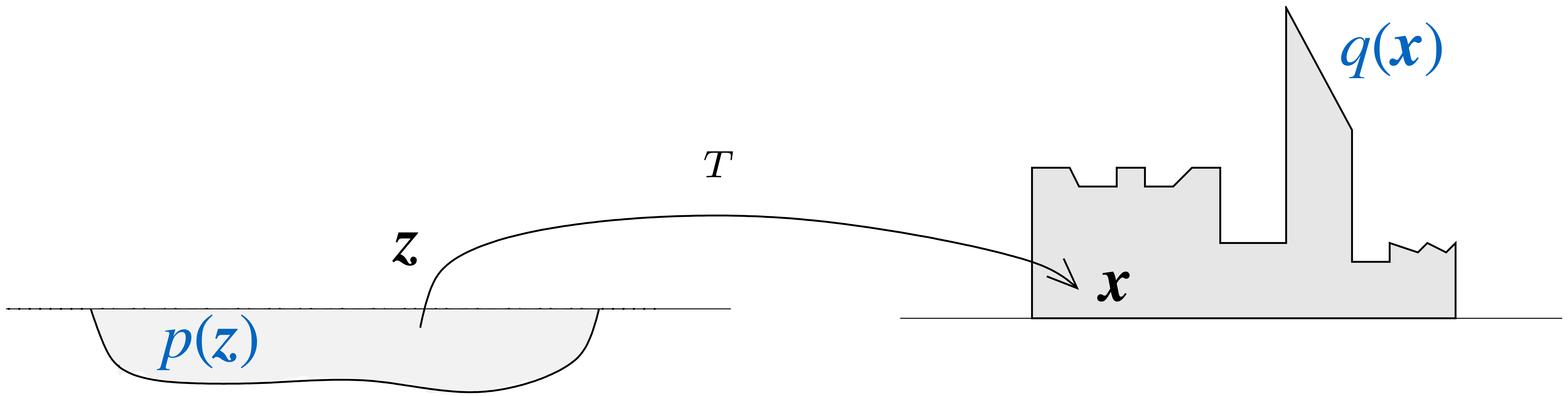
Fields Metal '10

Fields Metal '18

from Cuturi, Solomon NISP 2017 tutorial

Optimal Transport Theory

Monge problem (1781): How to transport earth with optimal cost ?



★ Brenier theorem (1991)

Under reasonable conditions the optimal map is

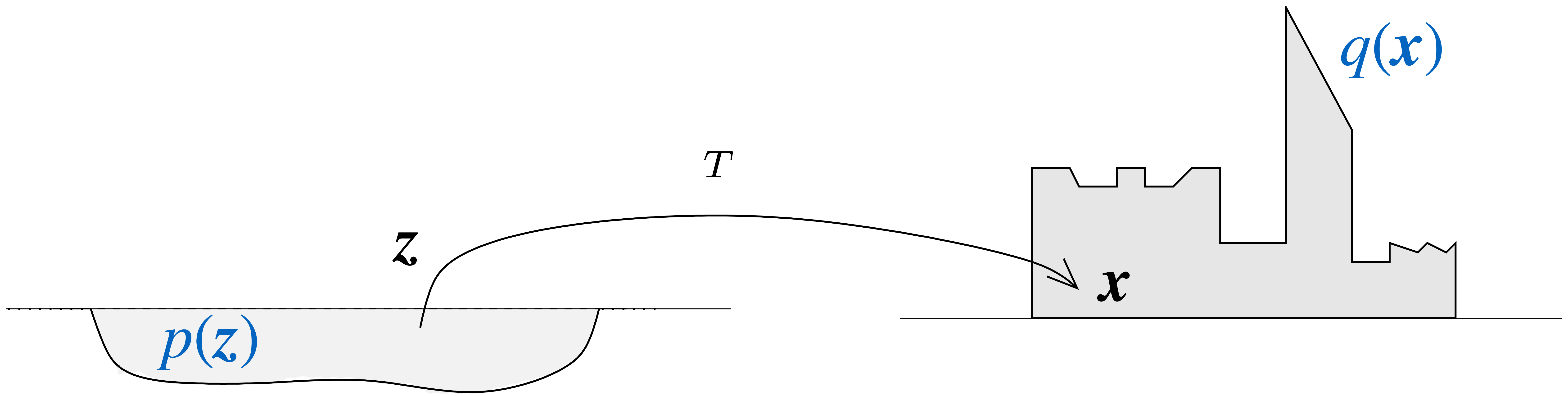
$$\boldsymbol{z} \mapsto \boldsymbol{x} = \nabla u(\boldsymbol{z})$$

Simply impose symmtry in the scalar generating potential

—林峰的主业!

Optimal Transport Theory

Monge problem (1781): How to transport earth with optimal cost ?



★ Brenier theorem (1991)

Under reasonable conditions
the optimal map is

$$z \mapsto \boldsymbol{x} = \nabla u(z)$$

Monge-Ampère Equation

$$\frac{p(z)}{q(\nabla u(z))} = \det\left(\frac{\partial^2 u}{\partial z_i \partial z_j}\right)$$

Shing-Tung Yau

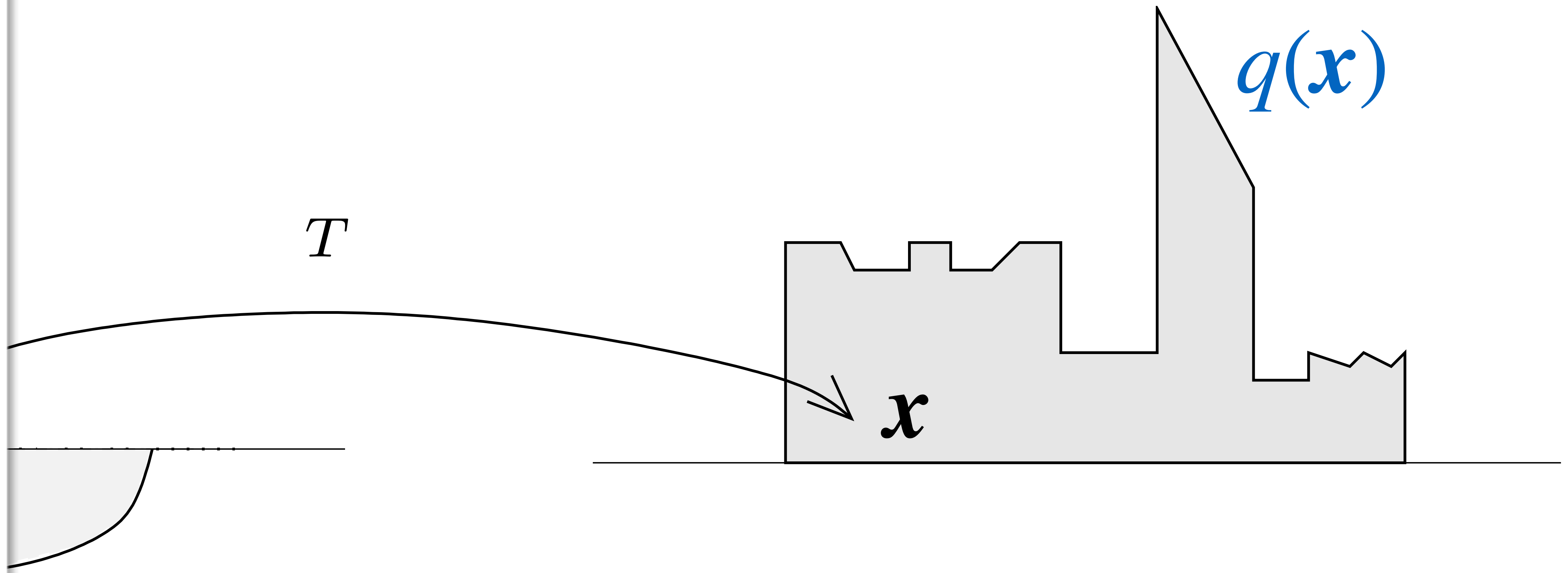
丘成桐

Fields Metal '82

Made contributions in differential equations, also to the Calabi conjecture in algebraic geometry, to the positive mass conjecture of general relativity theory, and to real and complex Monge-Ampère equation

Transport Theory

How to transport earth with optimal cost ?



★ Brenier theorem (1991)

Under reasonable conditions the optimal map is

$$z \mapsto \boldsymbol{x} = \nabla u(z)$$

Monge-Ampère Equation

$$\frac{p(z)}{q(\nabla u(z))} = \det\left(\frac{\partial^2 u}{\partial z_i \partial z_j}\right)$$

Shing-Tung Yau

丘成桐

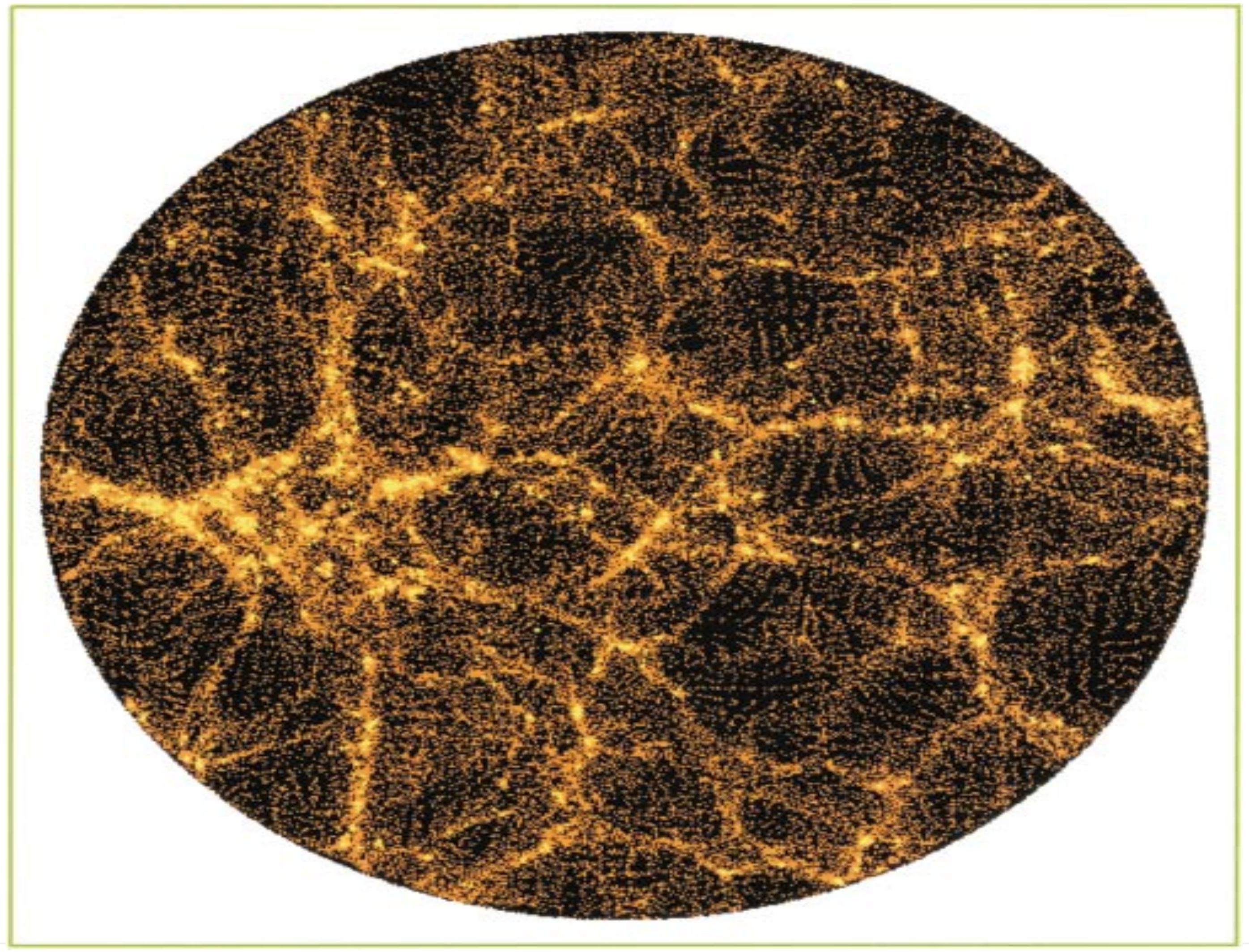
Fields Metal '82

Made contributions in differential equations, also to the Calabi conjecture in algebraic geometry, to the positive mass conjecture of general relativity theory, and to real and complex Monge-Ampère equation

letters to nature

A reconstruction of the initial conditions of the Universe by optimal mass transportation

Uriel Frisch*, Sabino Matarrese†, Roya Mohayaee‡* & Andrei Sobolevski§*



viscosity (discovered by Maxwell) does not operate, so that a non-collisional mechanism involving a small-scale gravitational instability must be invoked.

Our reconstruction hypothesis implies that the initial positions can be obtained from the present ones by another gradient map: $\mathbf{q} = \nabla_{\mathbf{x}}\Theta(\mathbf{x})$, where Θ is a convex potential related to Φ by a Legendre–Fenchel transform (see Methods). We denote by ρ_0 the initial mass density (which can be treated as uniform) and by $\rho(\mathbf{x})$ the final one. Mass conservation implies $\rho_0\, d^3q = \rho(\mathbf{x})\, d^3x$. Thus, the ratio of final to initial density is the jacobian of the inverse lagrangian map. This can be written as the following Monge–Ampère equation[20] for the unknown potential Θ:

$$\det(\nabla_{x_i}\nabla_{x_j}\Theta(\mathbf{x})) = \rho(\mathbf{x})/\rho_0 \qquad (1)$$

where 'det' stands for determinant.

We emphasize that no information about the dynamics of matter other than the reconstruction hypothesis is needed for our method, whose degree of success depends crucially on how well this hypothesis is satisfied. Exact reconstruction is obtained, for example, for the Zel'dovich approximation (before particle trajectories cross) and for adhesion-model dynamics (at arbitrary times).

We note that our Monge–Ampère equation for self-gravitating matter may be viewed as a nonlinear generalization of a Poisson equation (used for reconstruction in ref. 4), to which it reduces if particles have moved very little from their initial positions.

It has been discovered recently that the map generated by the solution to the Monge–Ampère equation (1) is the (unique) solution to an optimization problem[21] (see also refs 22 and 23).

★ Brenier theorem (1991)

Under reasonable conditions the optimal map is

$$z \mapsto x = \nabla u(z)$$

Monge-Ampère Equation

$$\frac{p(z)}{q(\nabla u(z))} = \det\left(\frac{\partial^2 u}{\partial z_i \partial z_j}\right)$$

The physics behind: fluid control

$$\frac{p(\boldsymbol{z})}{q(\nabla u(\boldsymbol{z}))} = \det\left(\frac{\partial^2 u}{\partial z_i \partial z_j}\right)$$

Monge-Ampère Equation

Continuous-time limit

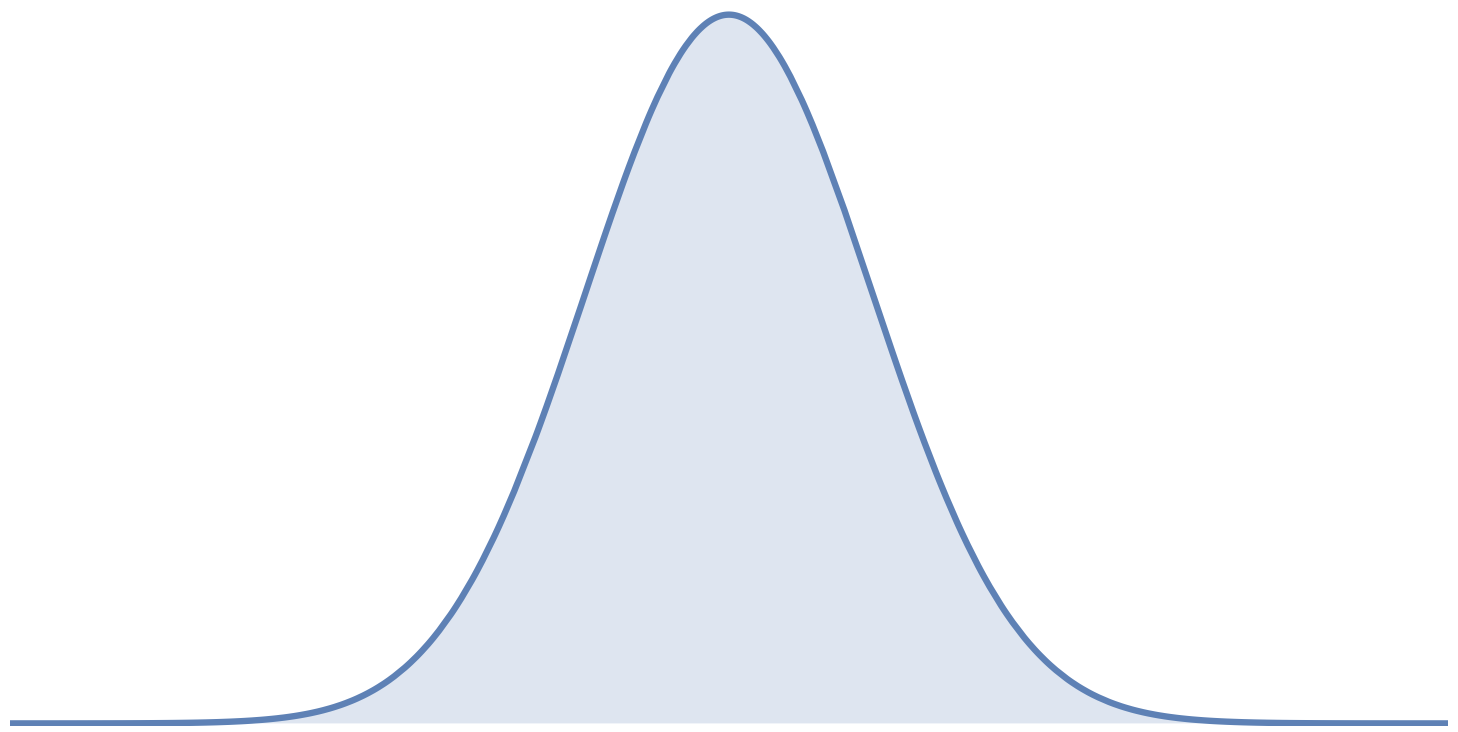
$$u(\boldsymbol{z}) = |\boldsymbol{z}|^2/2 + \epsilon\varphi(\boldsymbol{z})$$

$$\epsilon \to 0$$

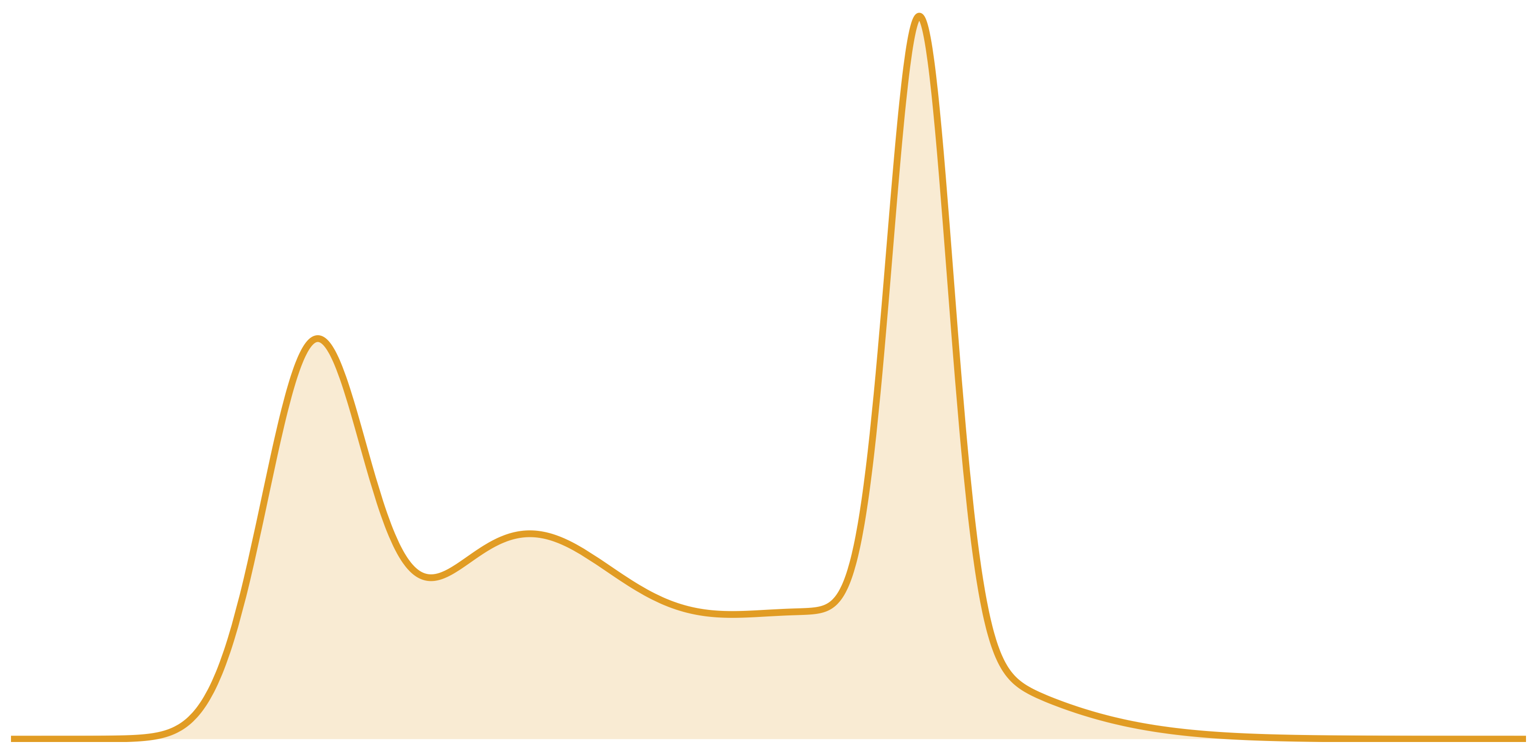
$$\frac{\partial p(\boldsymbol{x}, t)}{\partial t} + \nabla \cdot \left[p(\boldsymbol{x}, t)\nabla\varphi\right] = 0$$

Liouville Equation

(Continuity equation of compressible fluids)



Simple density



Complex density

With Linfeng Zhang, Weinan E 1809.10188

The physics behind: fluid control

$$\frac{p(\boldsymbol{z})}{q(\nabla u(\boldsymbol{z}))} = \det\left(\frac{\partial^2 u}{\partial z_i \partial z_j}\right)$$

Monge-Ampère Equation

Continuous-time limit

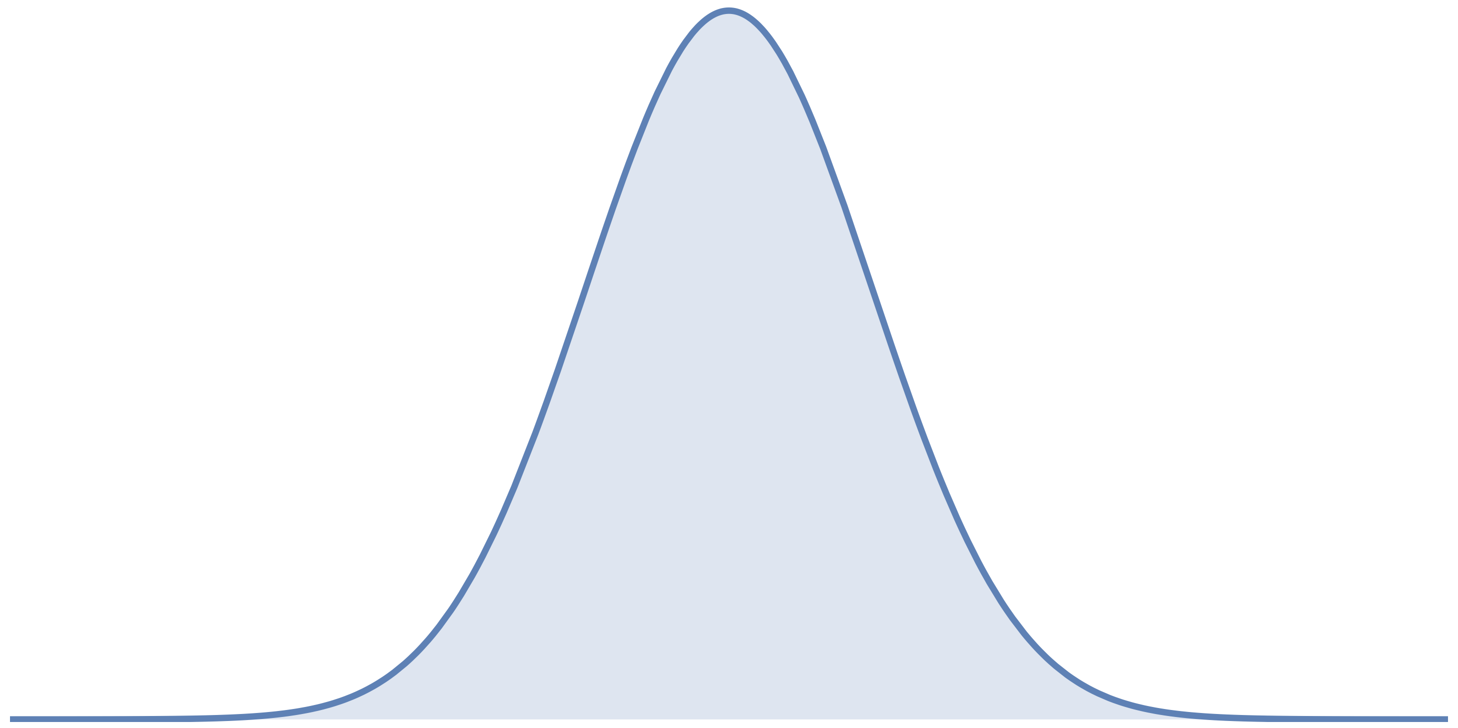
$$u(\boldsymbol{z}) = |\boldsymbol{z}|^2/2 + \epsilon\varphi(\boldsymbol{z})$$

$$\epsilon \to 0$$

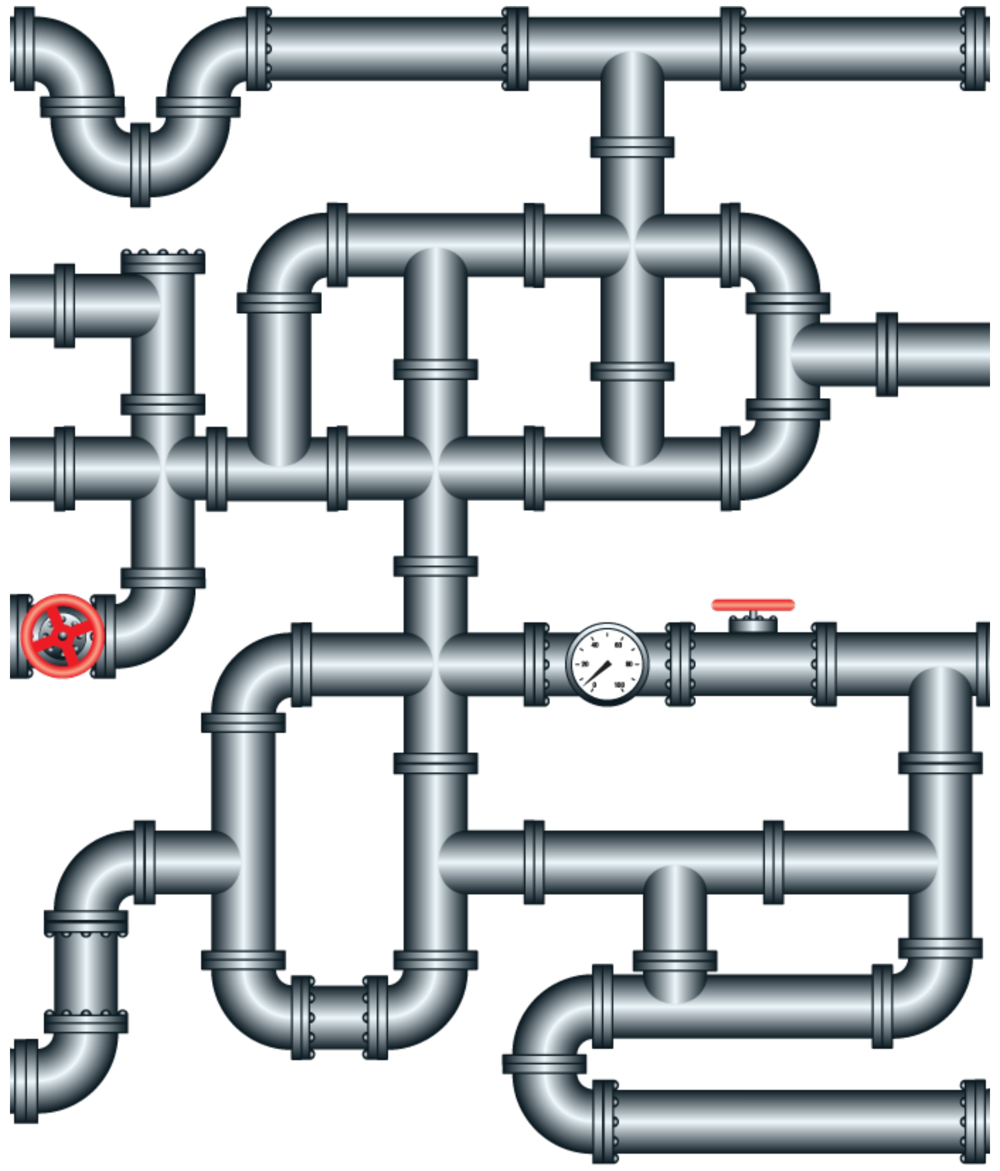
$$\frac{\partial p(\boldsymbol{x}, t)}{\partial t} + \nabla \cdot \left[p(\boldsymbol{x}, t)\nabla\varphi\right] = 0$$

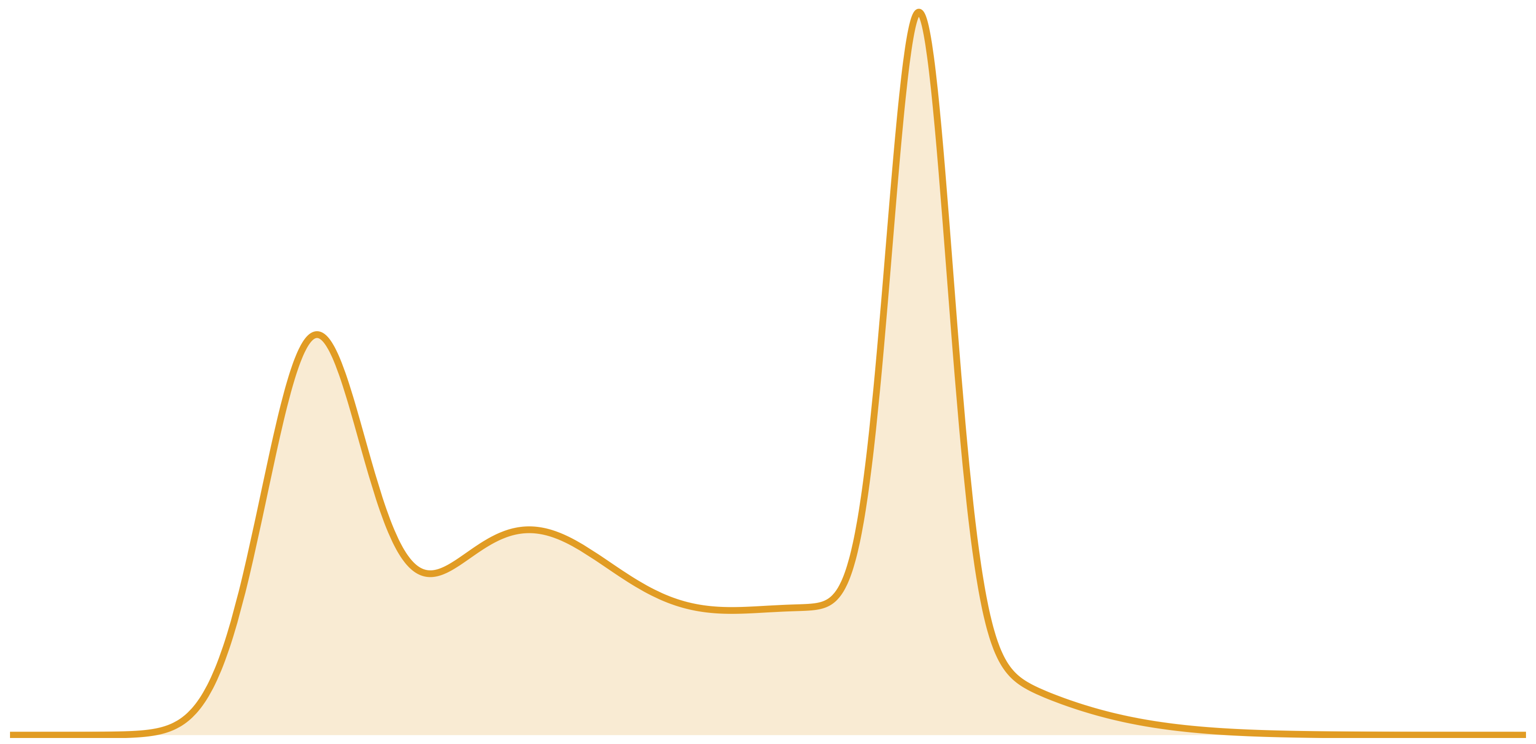
Liouville Equation

(Continuity equation of compressible fluids)



Simple density



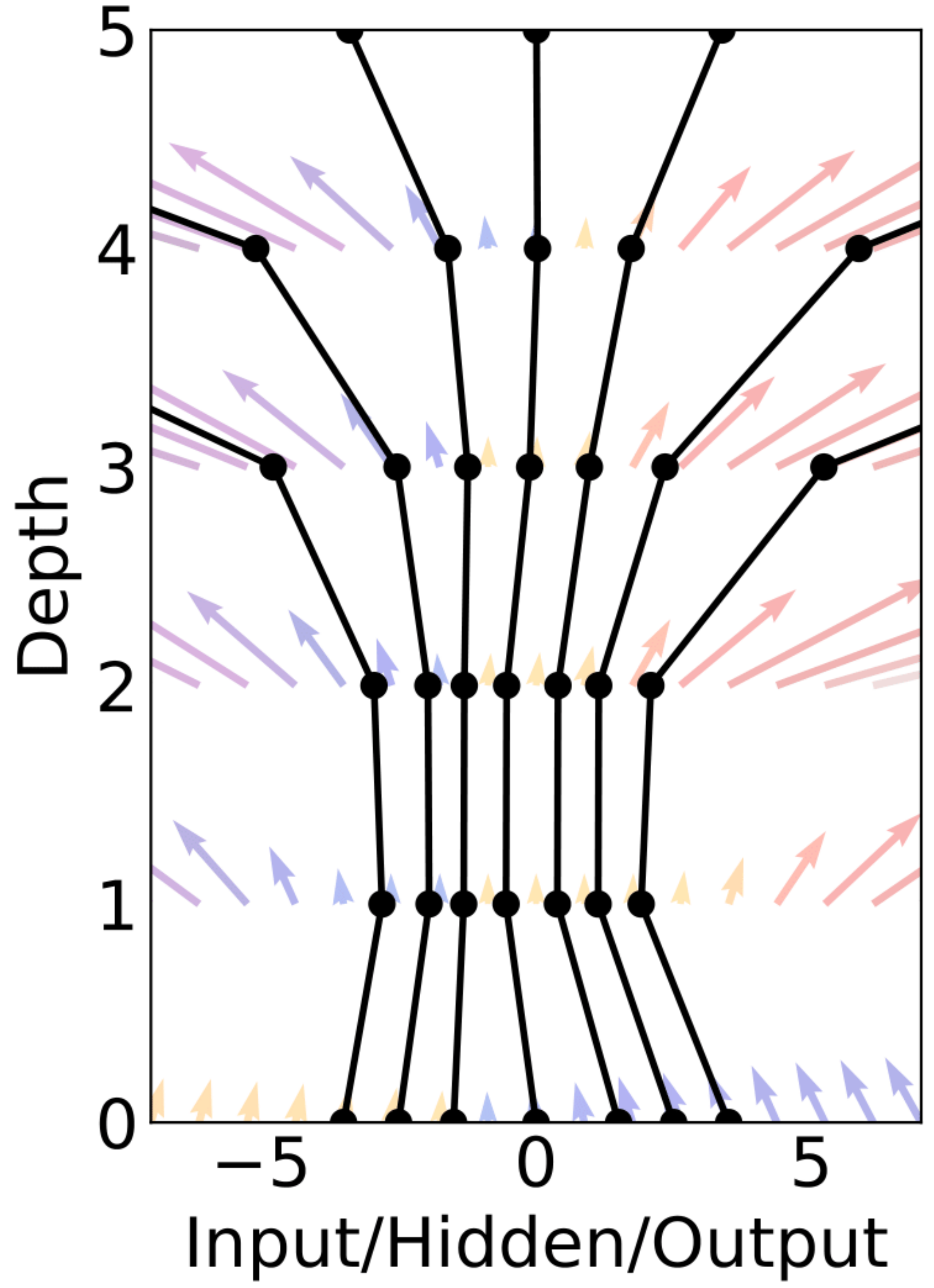


Complex density

With Linfeng Zhang, Weinan E 1809.10188

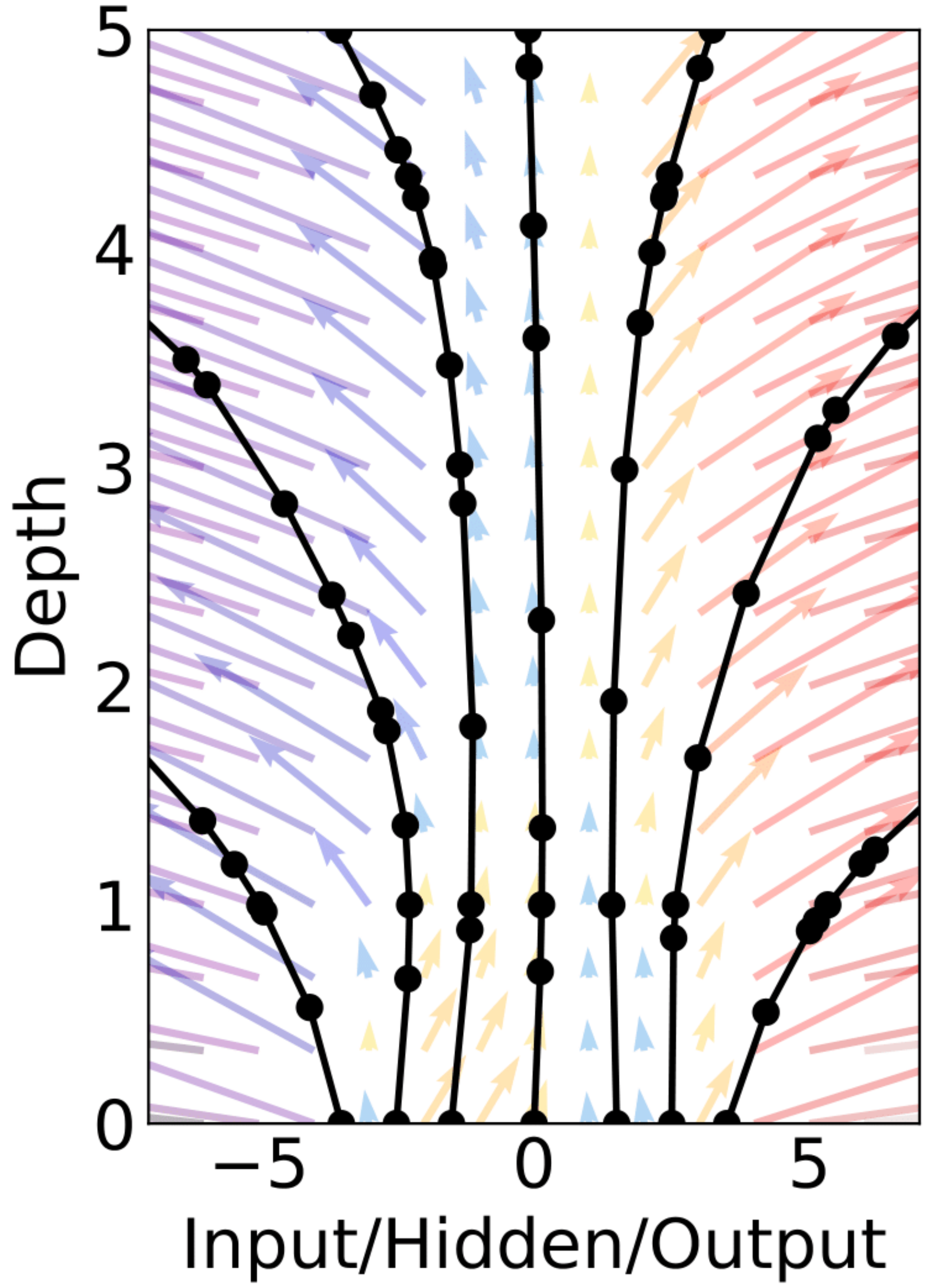
Neural Ordinary Differential Equations

Residual network



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

ODE network



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

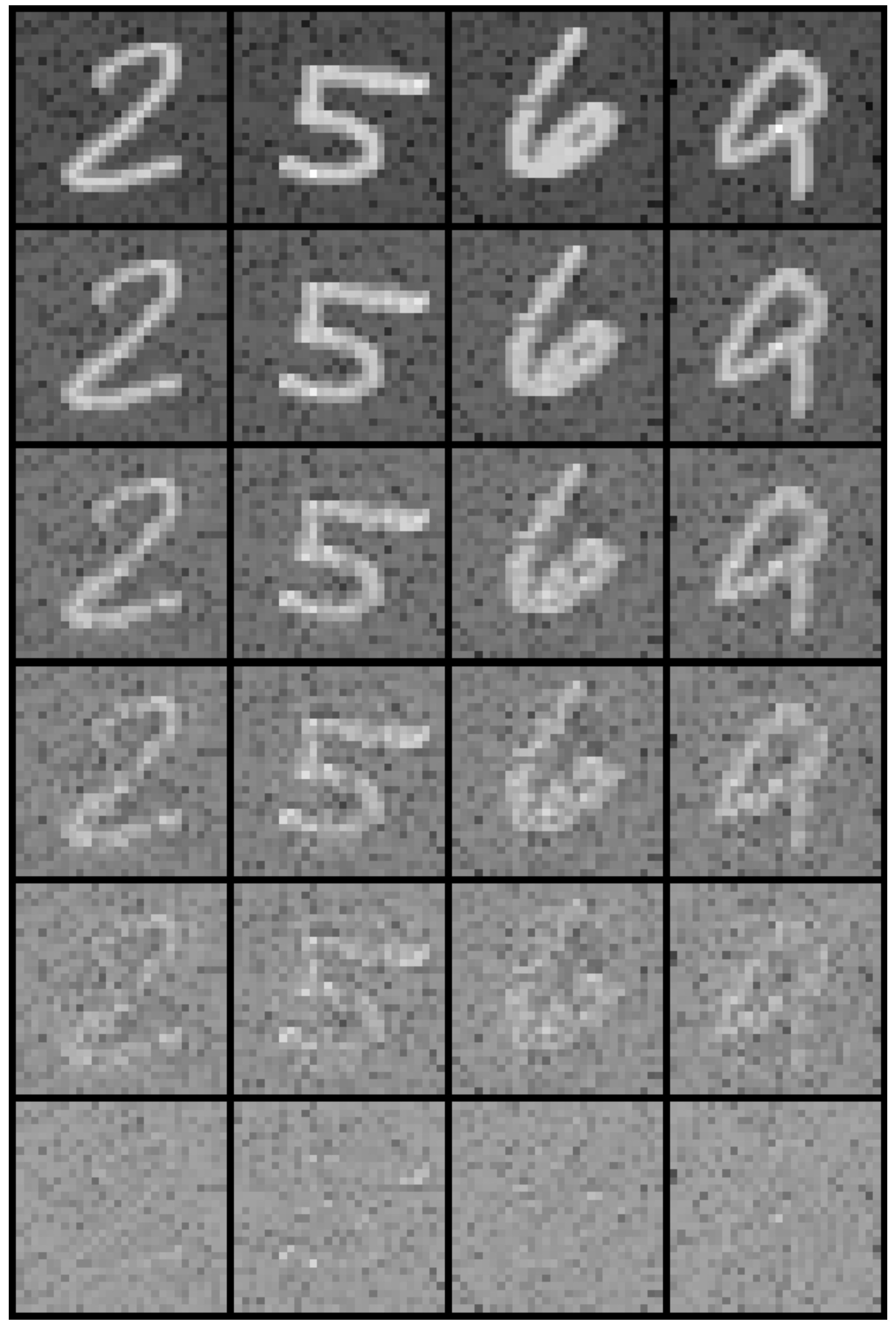
Chen et al, 1806.07366 NIPS ’18 Best paper award

cf Harbor el al 1705.03341
Lu et al 1710.10121, E 17’…

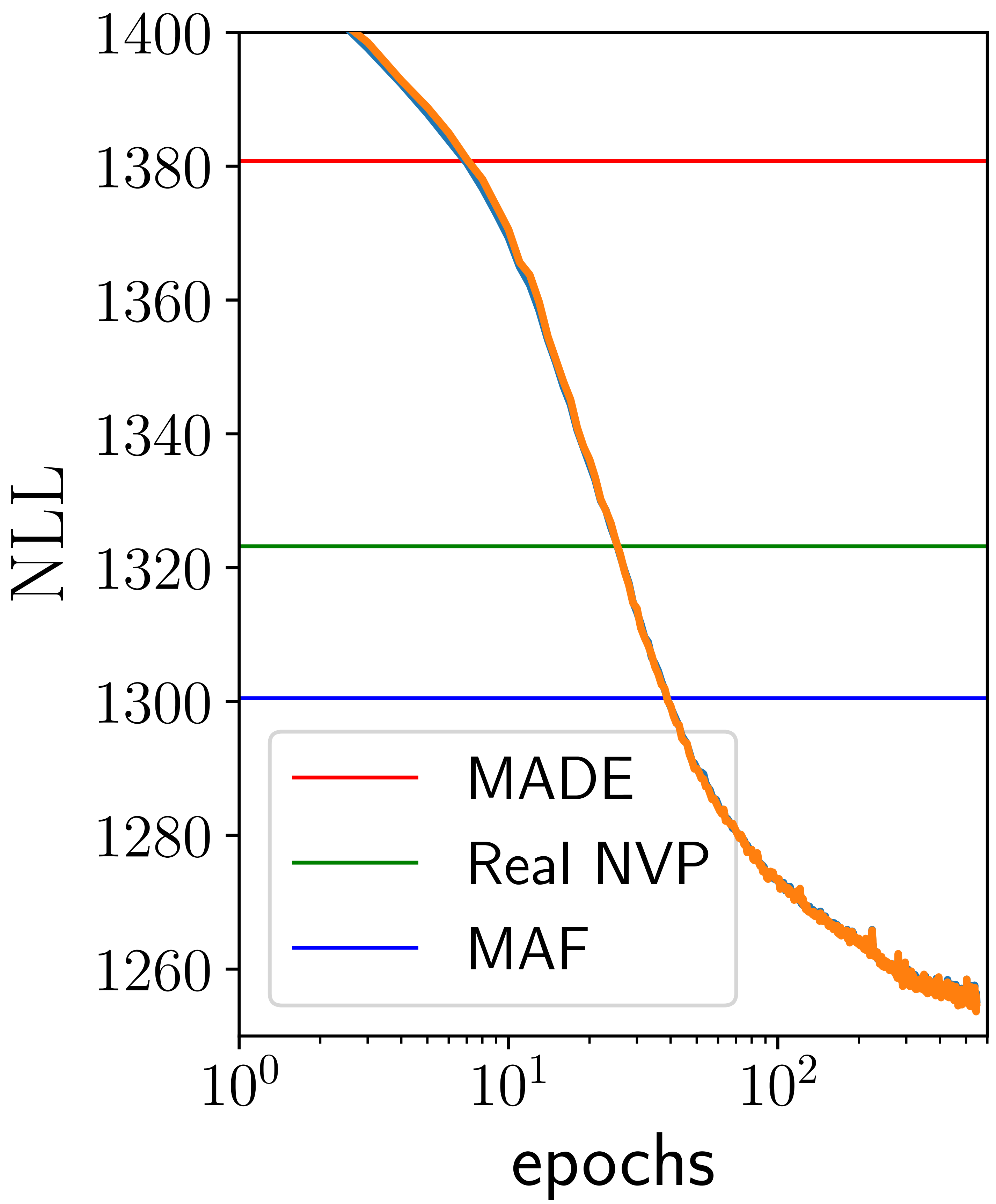
Density estimation of hand-written digits

A standard benchmark for generative models, lower is better

data space



latent space



'15

'16

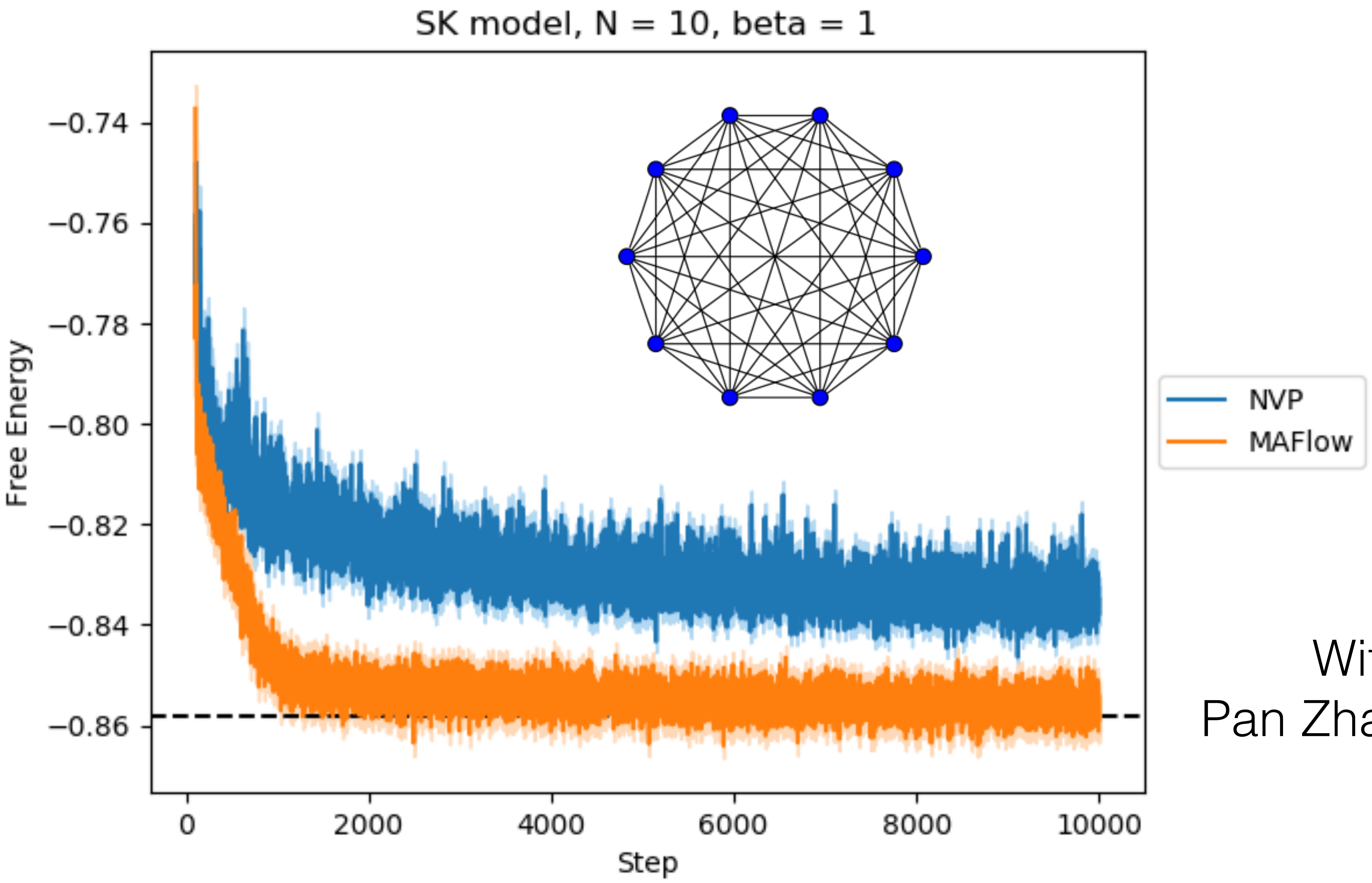
'17

'18 Our Result

See also
FFJORD 1810.01367

1 **State-of-the-art performance in unstructured density estimation**

Variational study of Sherrington-Kirkpatrick spin glasses

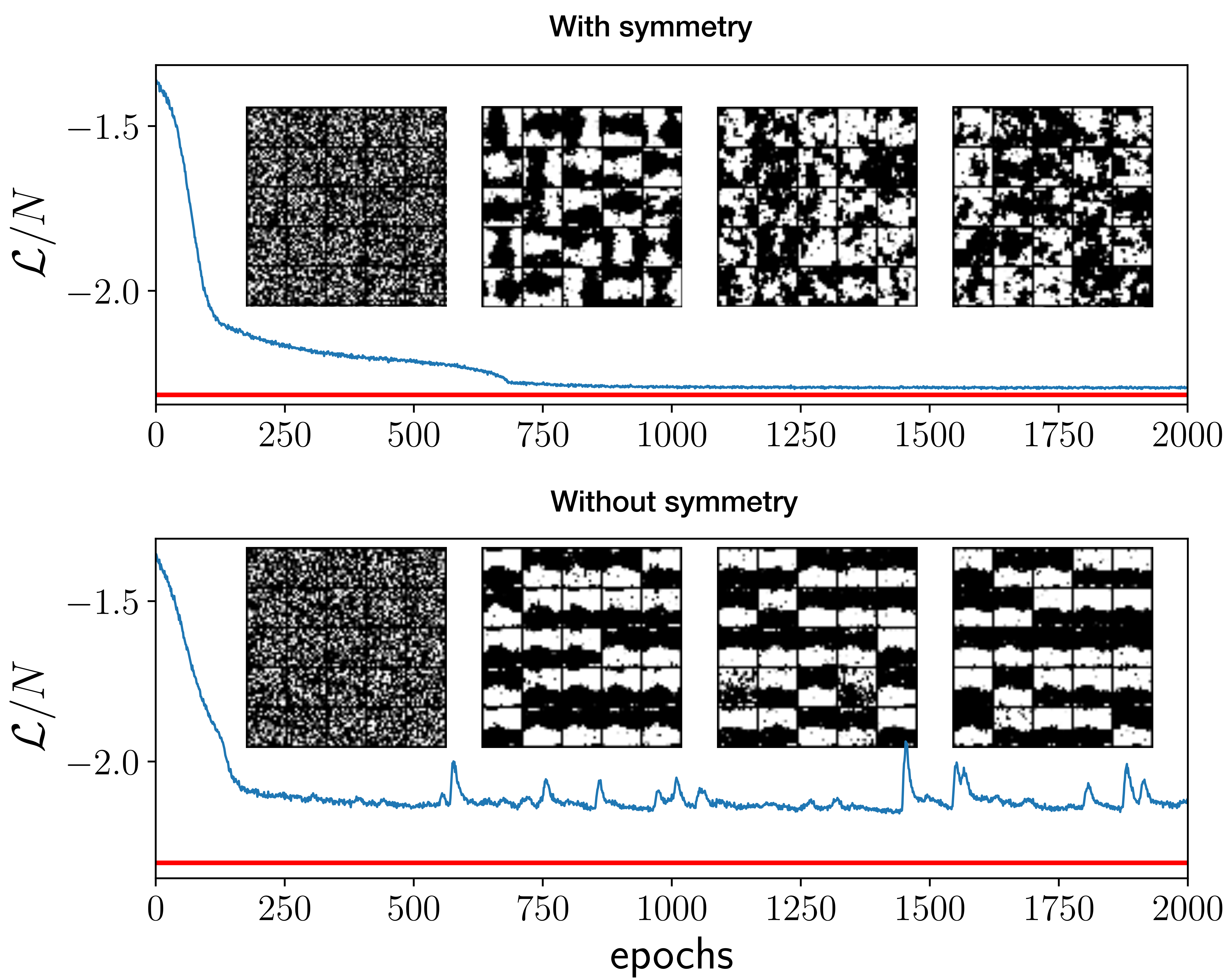


With Dian Wu,
Pan Zhang unpublished

② **Better variational energy than previous network structure**

③ **Direct sample magnetic domains respecting physical symmetry**

Importance of a symmetric flow

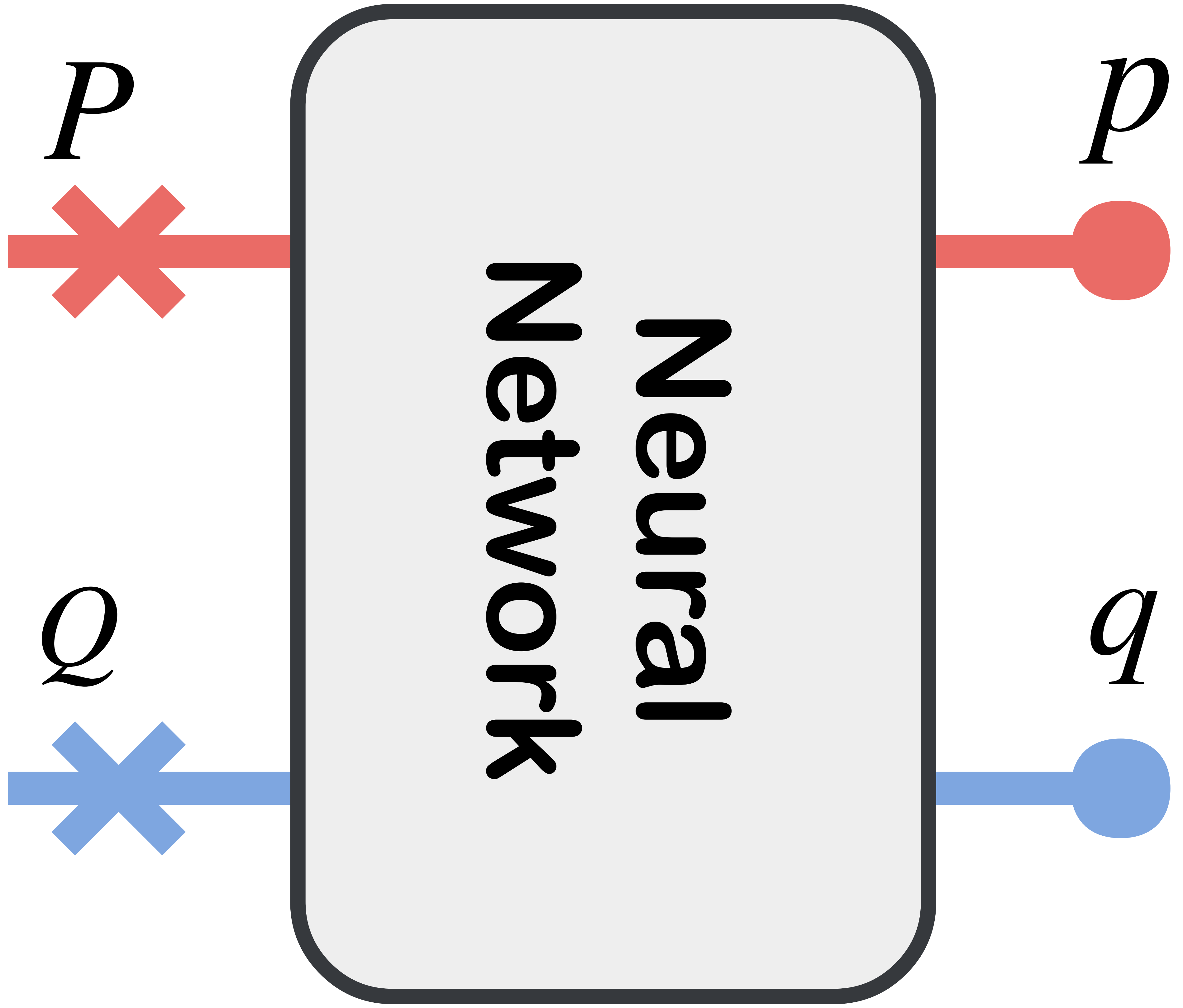


“mode collapse”

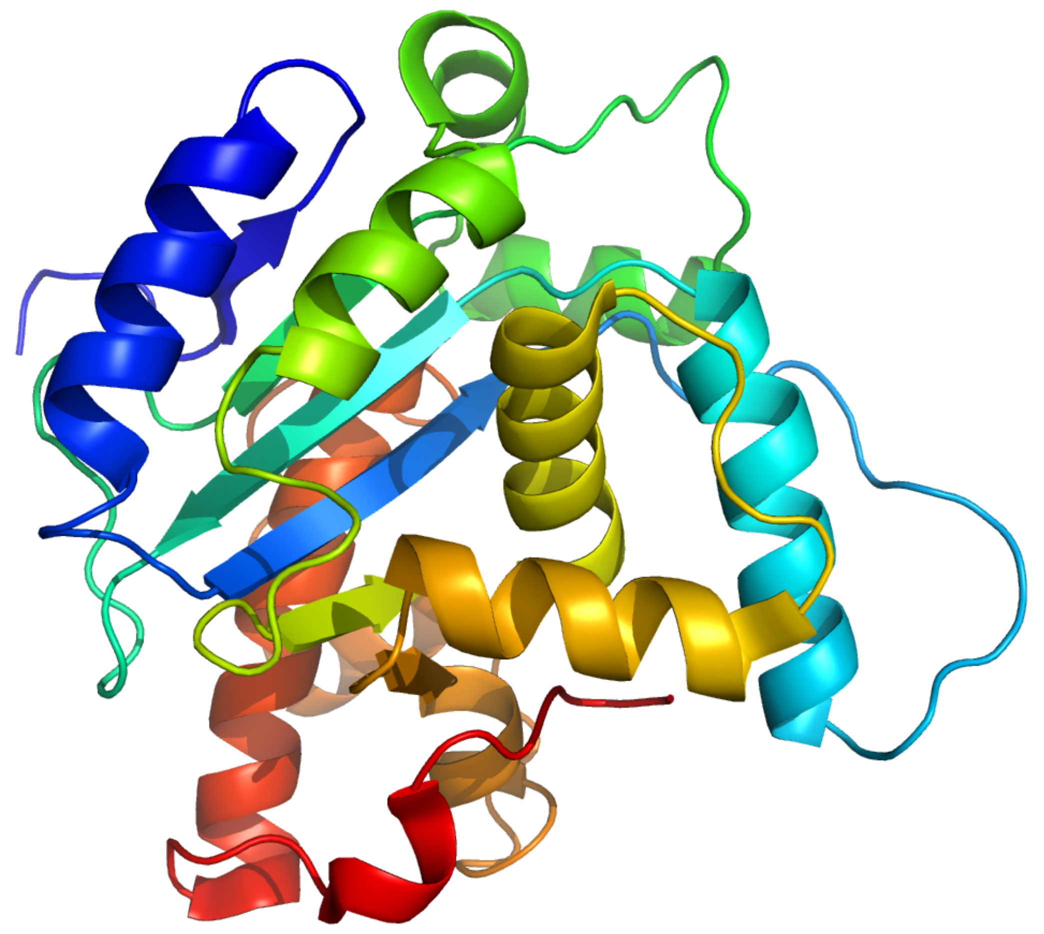
Neural Canonical Transformations

Incompressible symplectic flow in phase space

$H(P, Q) = (P^2 + Q^2)/2$



$H(p, q)$



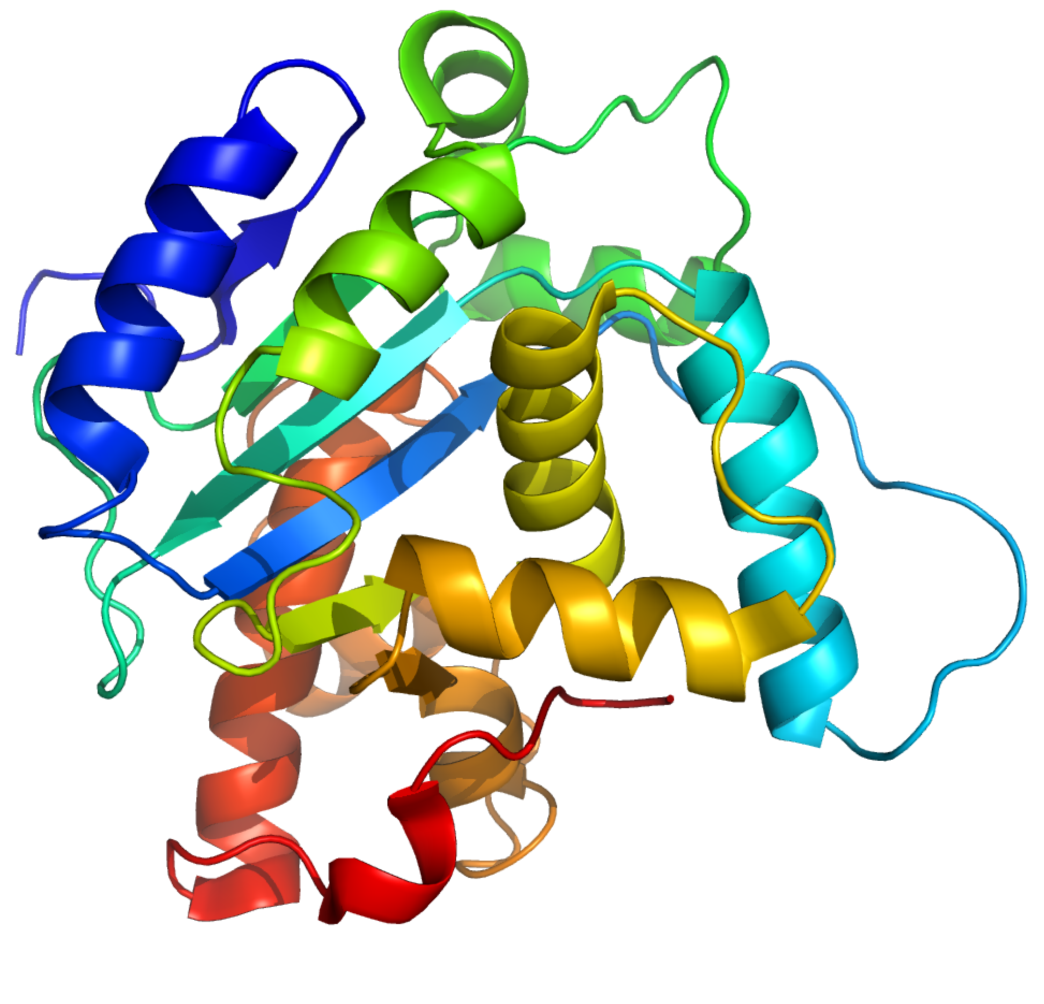
Identifying mutually independent collective modes for molecular simulations (MD, PIMD), and effective field theory

Neural Canonical Transformations

Incompressible symplectic flow in phase space

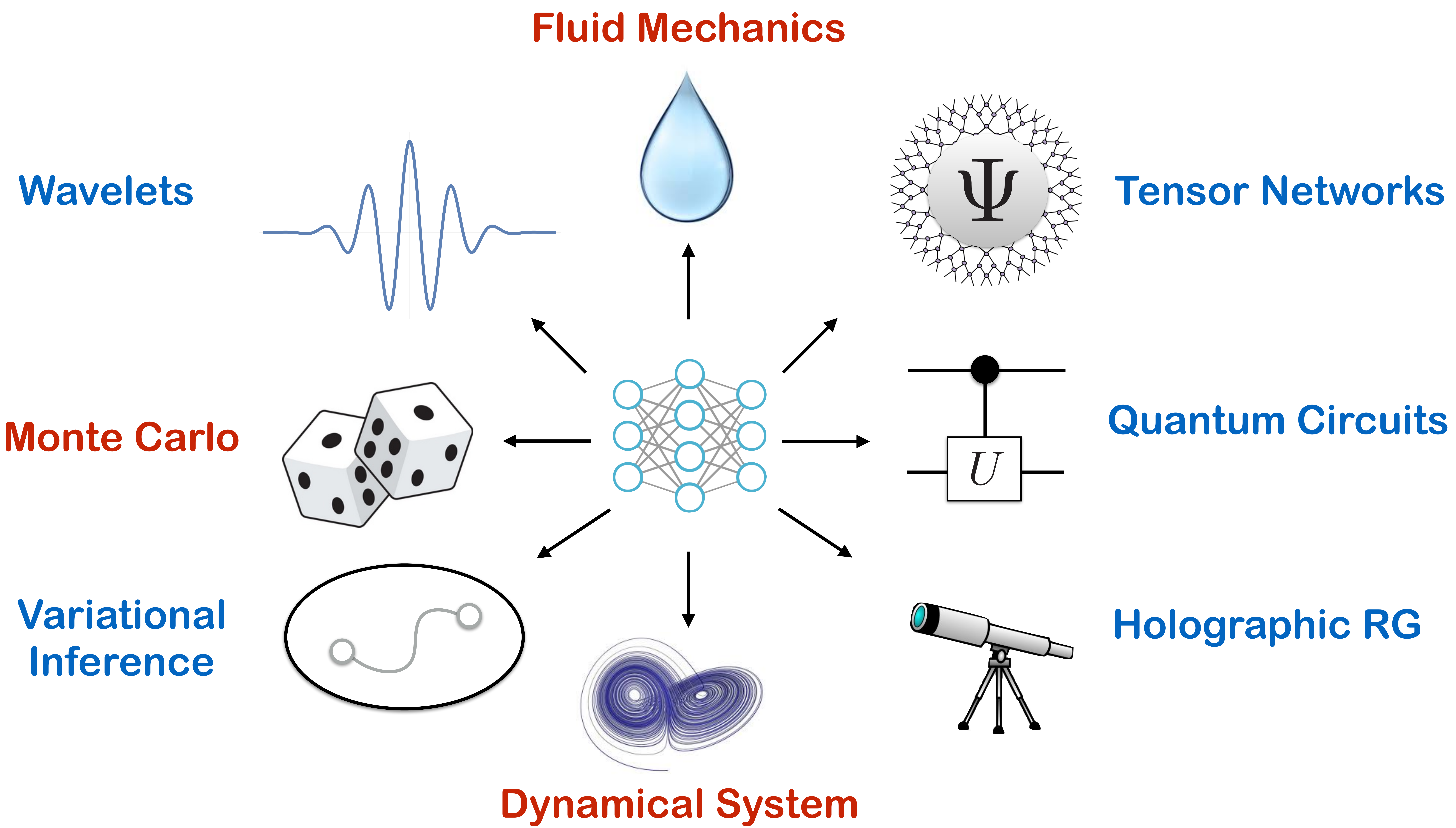
$H(P, Q) = (P^2 + Q^2)/2$

$H(p, q)$

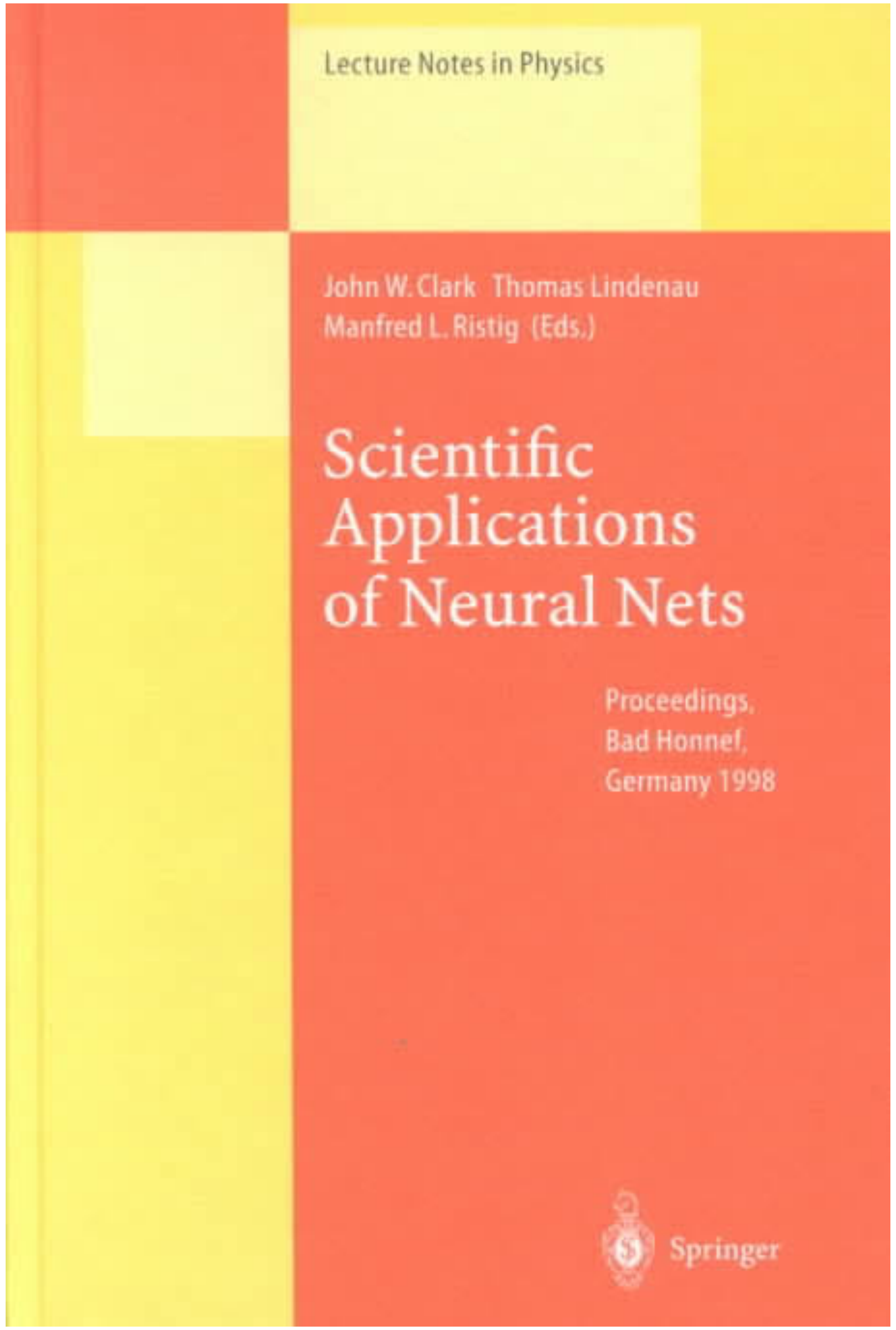


Identifying mutually independent collective modes for molecular simulations (MD, PIMD), and effective field theory

4


Fluid Mechanics
Wavelets
Tensor Networks
Monte Carlo
Quantum Circuits
U
Variational
Inference
Holographic RG
Dynamical System

But, this is not the first time we feel excited



8 Doing Science With Neural Nets: Pride and Prejudice

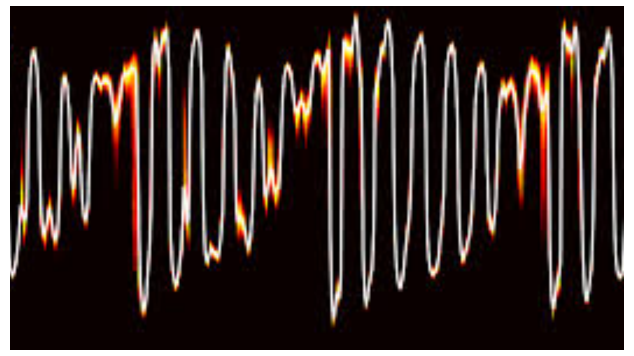
When neural networks re-emerged on the scene in the mid-80s as a new and glamorous computational paradigm, the initial reaction in some sectors of the scientific community was perhaps too enthusiastic and not sufficiently critical. There was a tendency on the part of practitioners to oversell the powers 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.

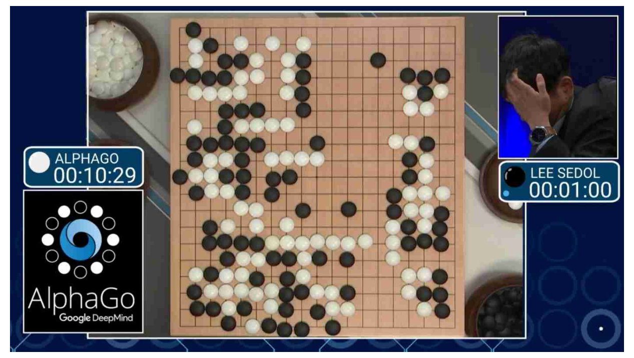
Why now, but not 20 years ago ?

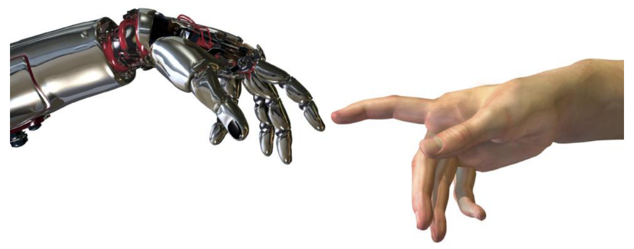
What has changed ?

What has not ?

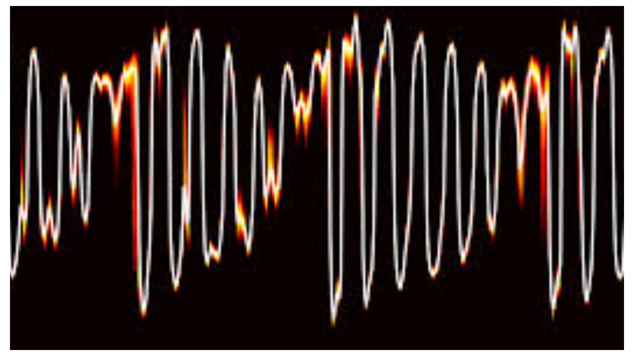
深度学习的秘诀究竟是什么？

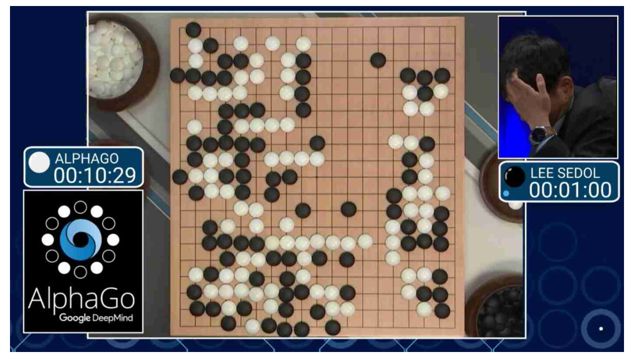


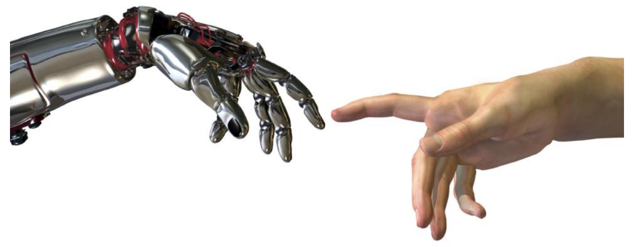




深度学习的秘诀究竟是什么？



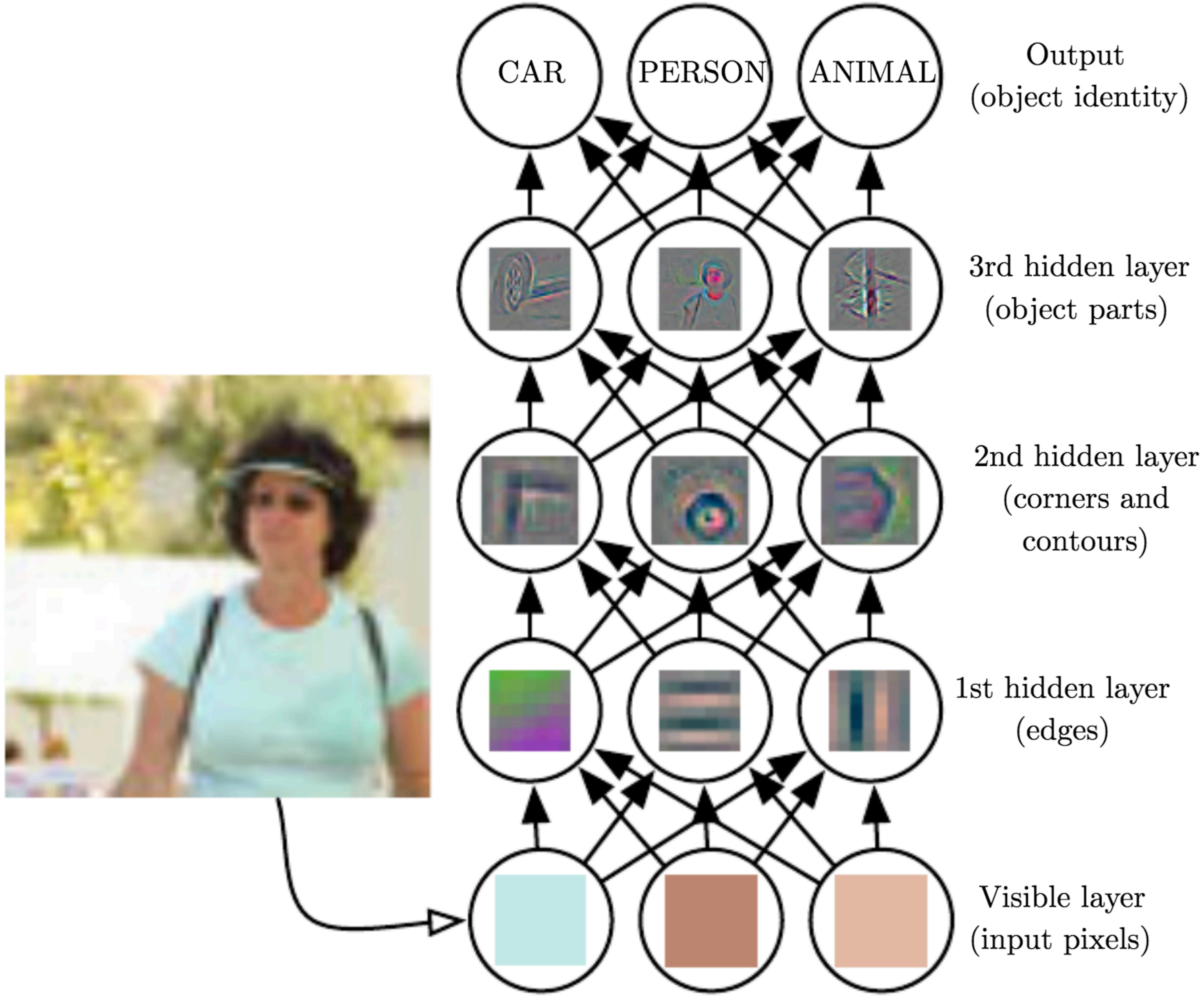




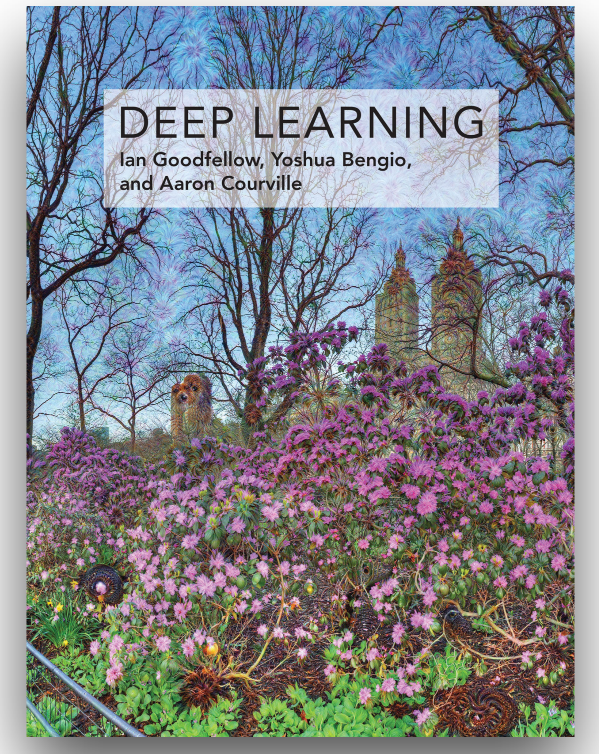
核心思想： 表示学习

关键技术： 微分编程

Representation Leanring



Page 6
Figure 1.2



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

Magic of learning representation

Neural style transfer

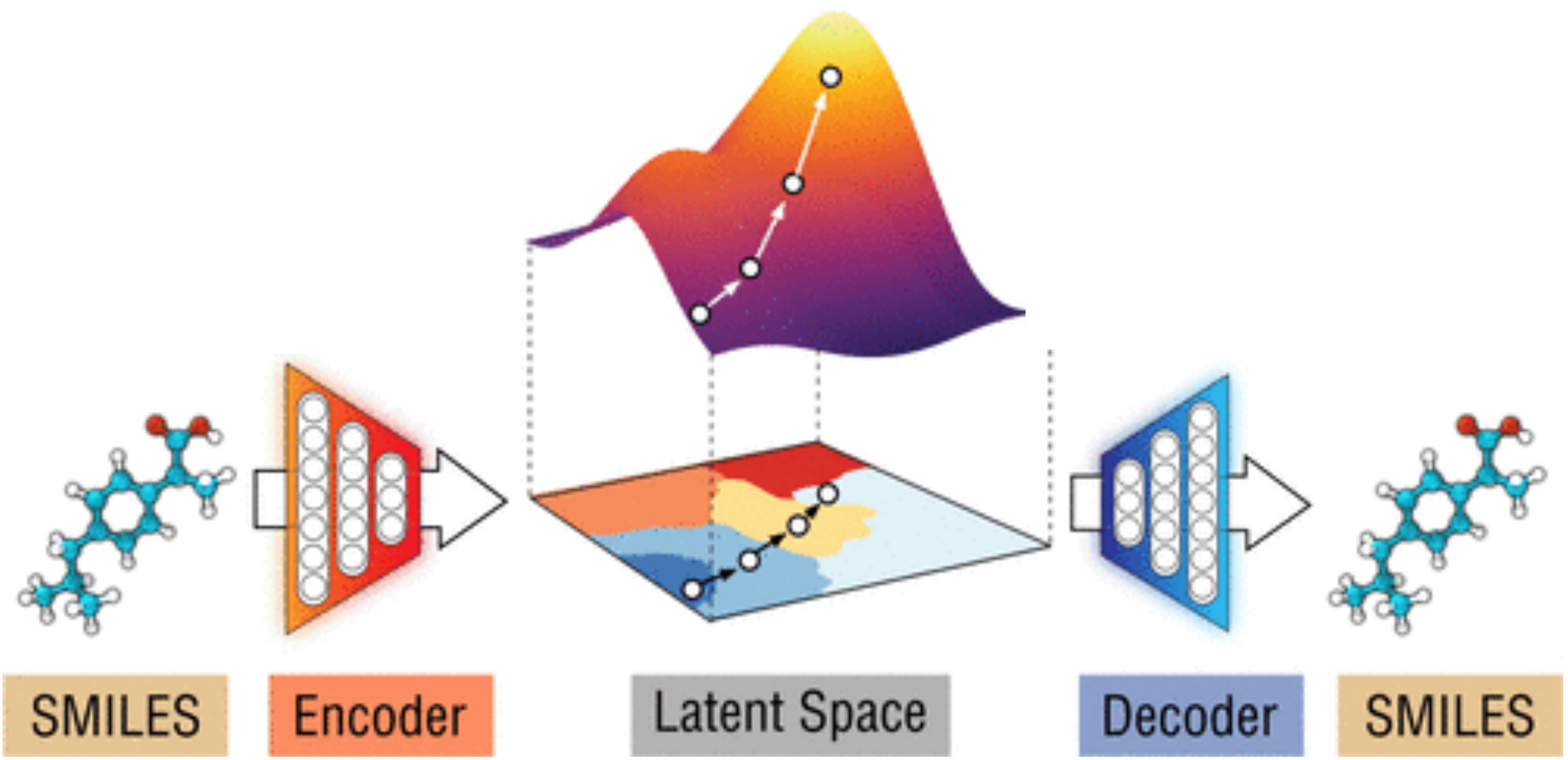
Gatys et al, 1508.06576

Latent space interpolation

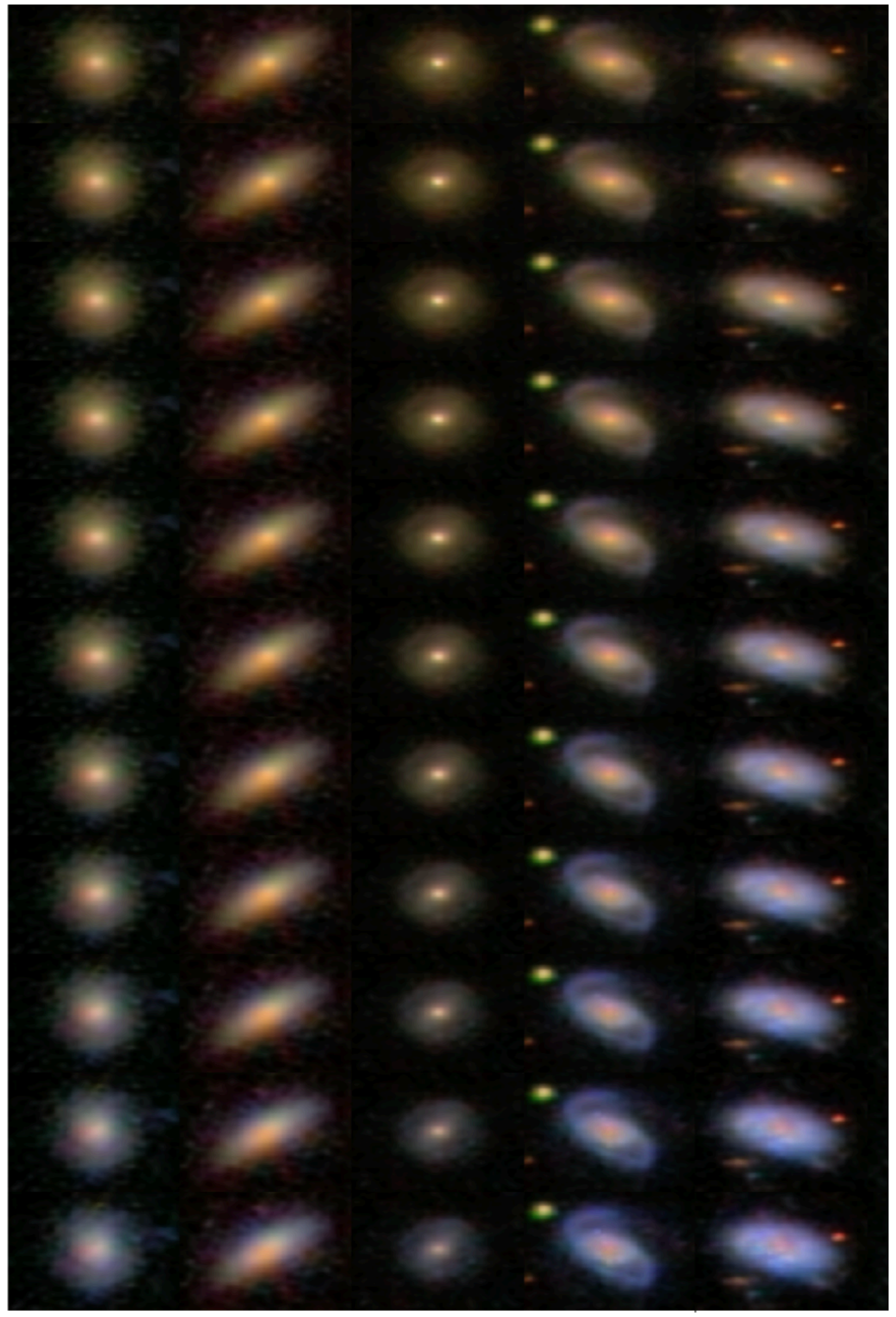
Glow 1807.03039

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

Learning representation for science

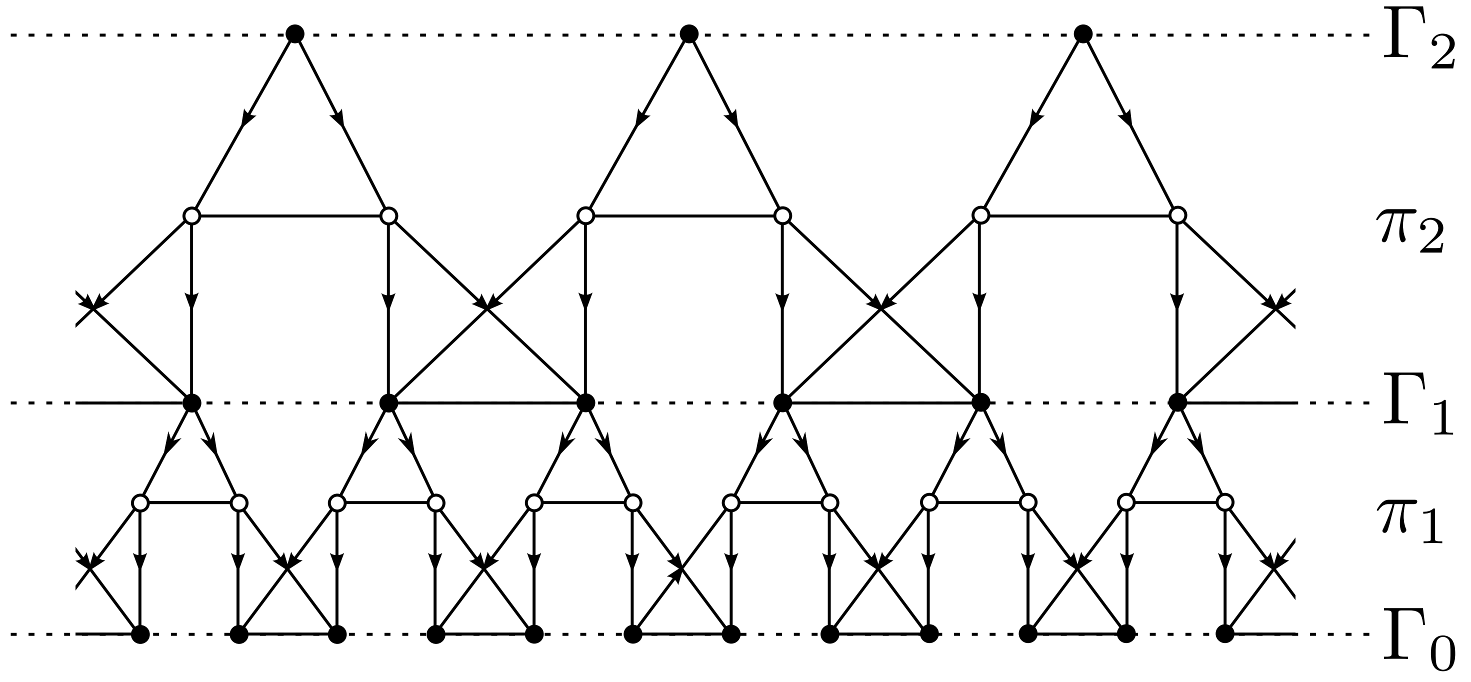


Automatic chemical design,
Gomez-Bombarelli et al,1610.02415

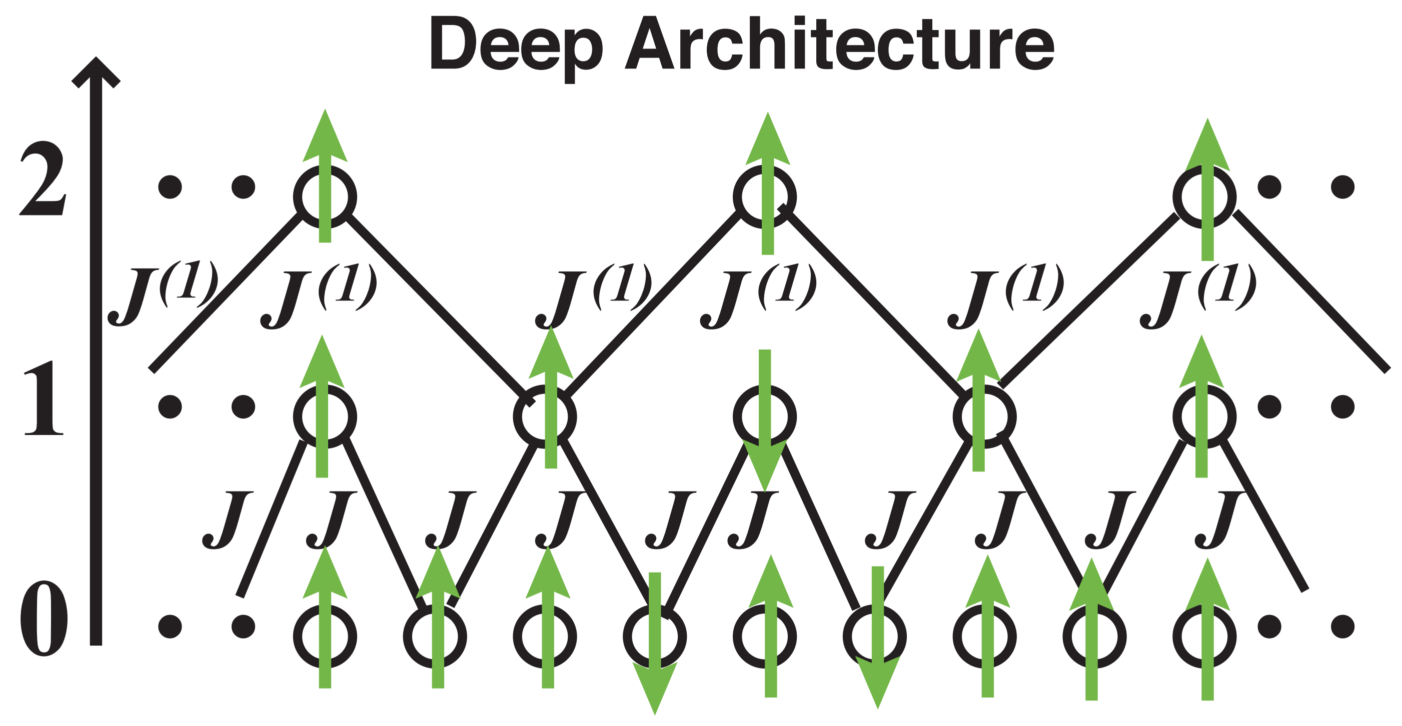


Galaxy evolution
Schawinski et al, 1812.01114

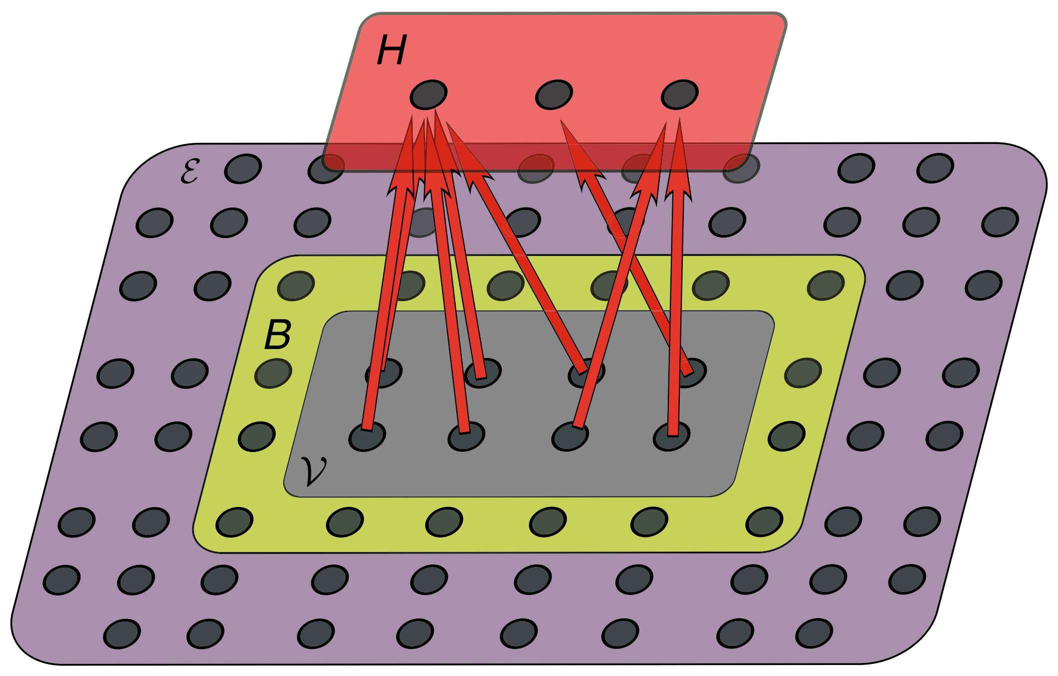
Deep Learning and Renormalization Group



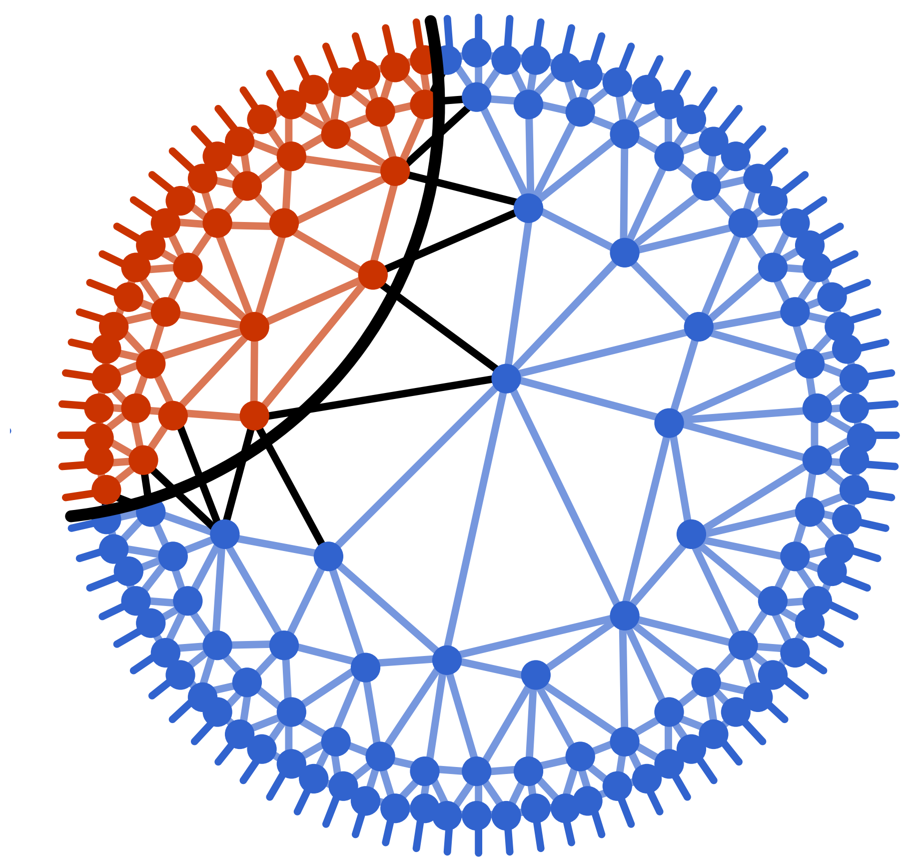
Bény, 1301.3124



Mehta and Schwab, 1410.3831



Koch-Janusz and Ringel, 1704.06279

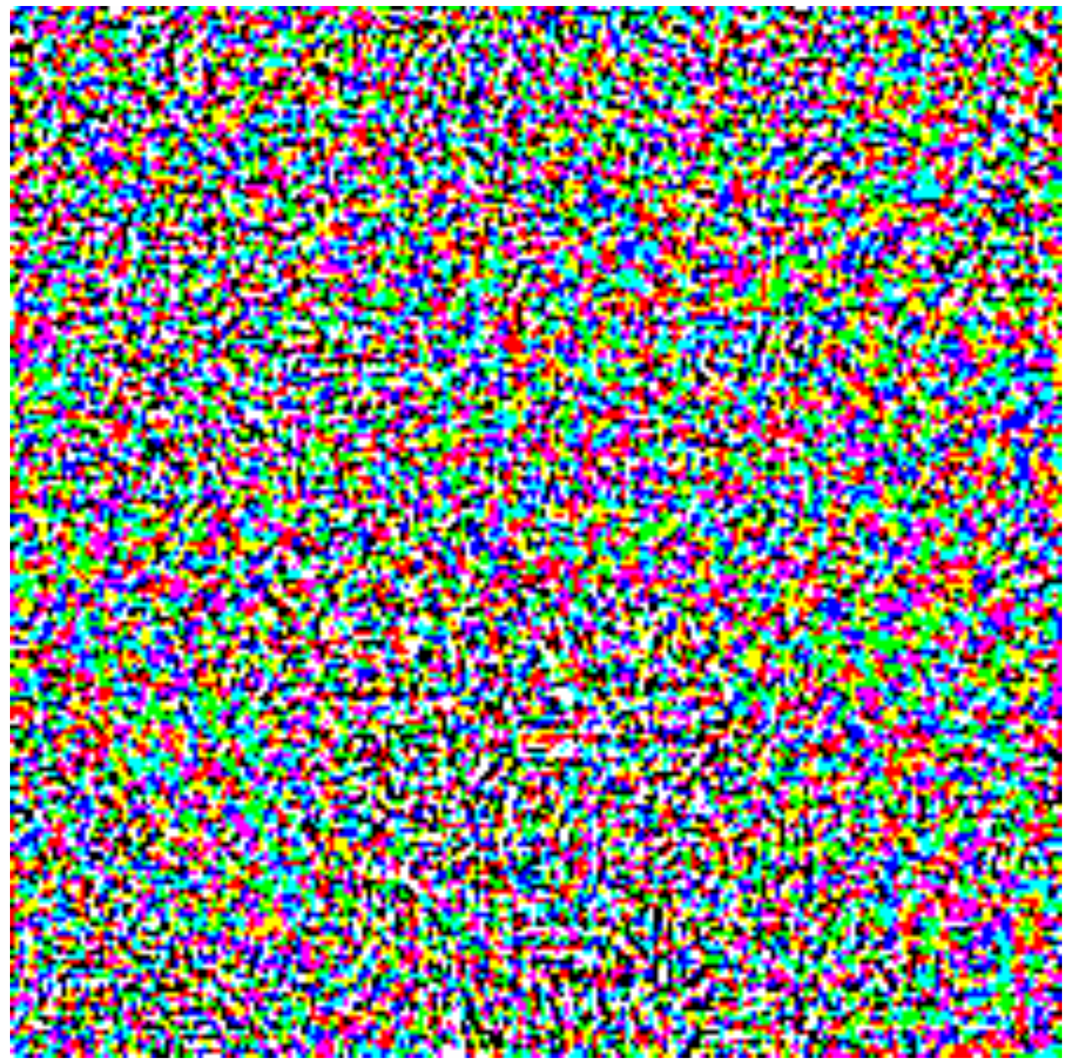


You, Yang, Qi, 1709.01223

and more…

Deep Learning and Renormalization Group

$+ .007 \times$



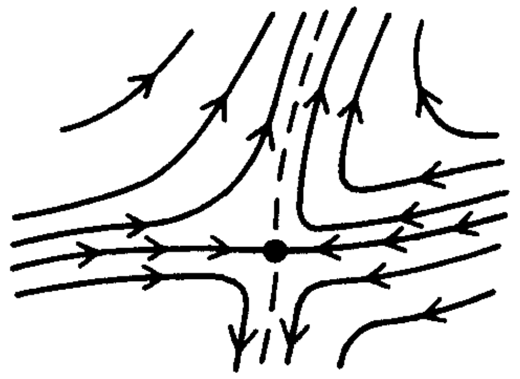
$=$

Panda
58% confidence

Goodfellow et al, 2014

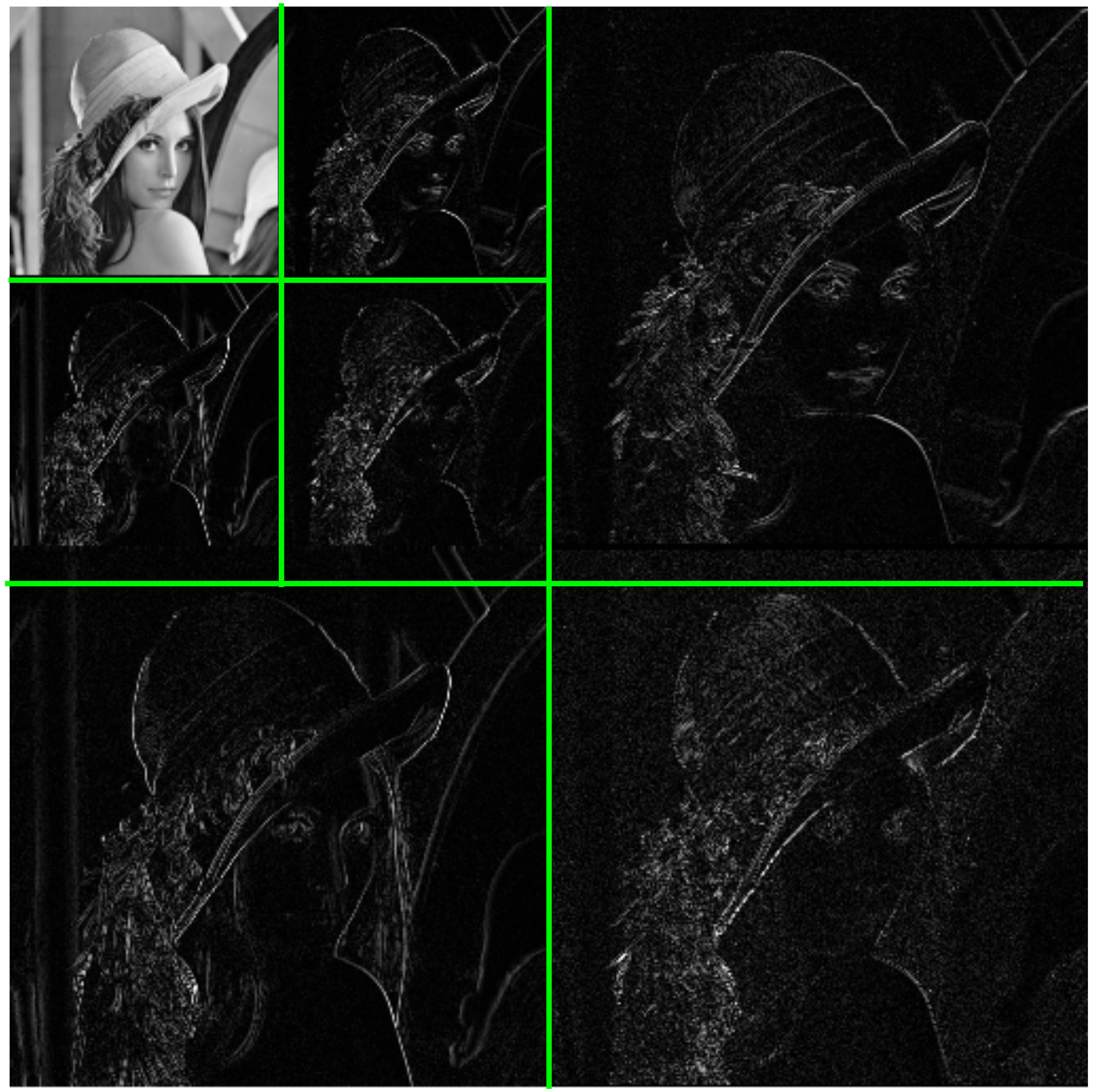
Gibbon
99% confidence

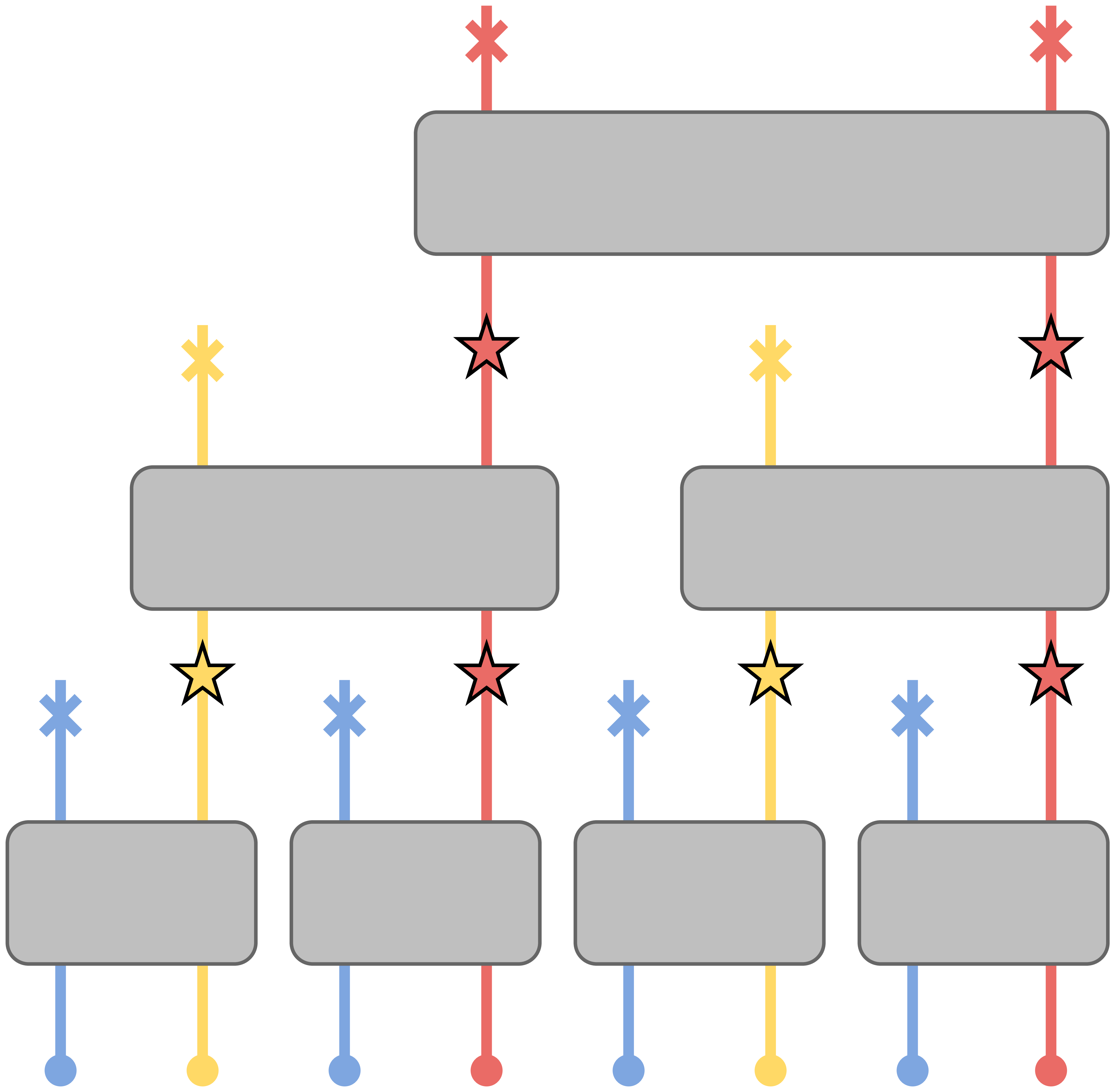
Vulnerability of deep learning, Kenway, 1803.06111 & 1803.10995



and more…

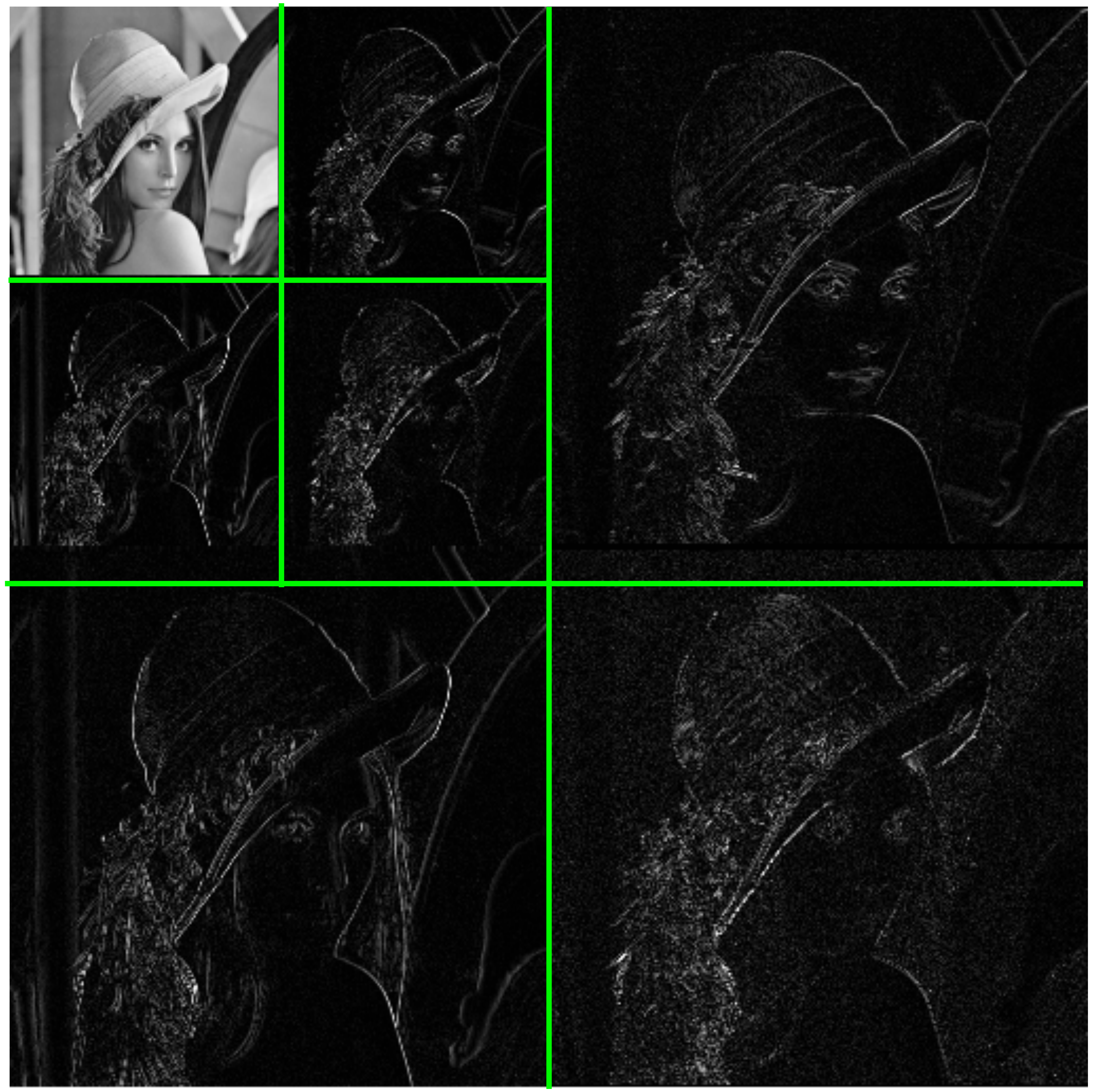
Wavelet transformation for Lena and Ising

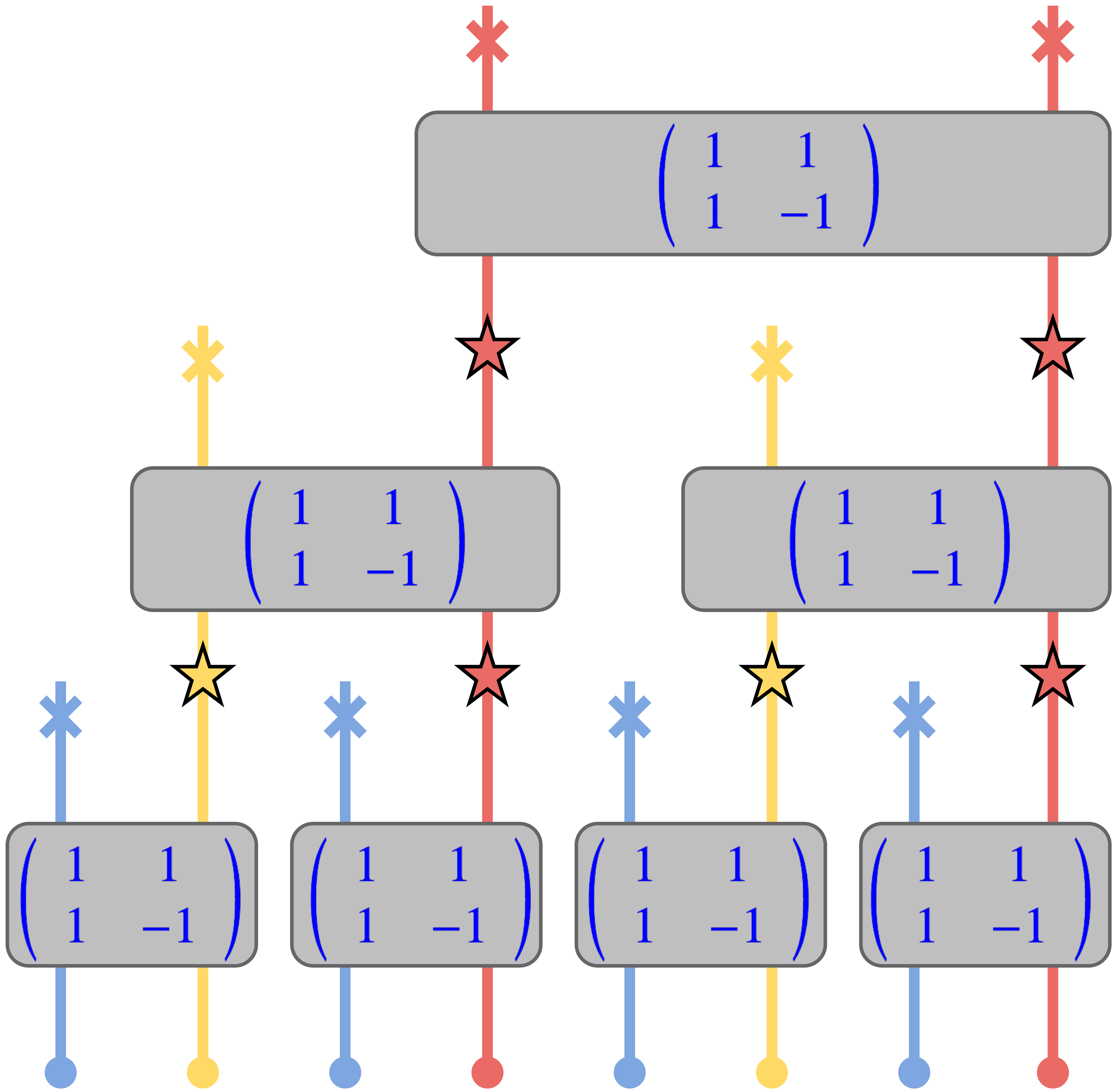




Guy, Wavelets & RG1999+ White, Evenbly, Qi, Wavelets, MERA, and holographic mapping 2013+

Wavelet transformation for Lena and Ising

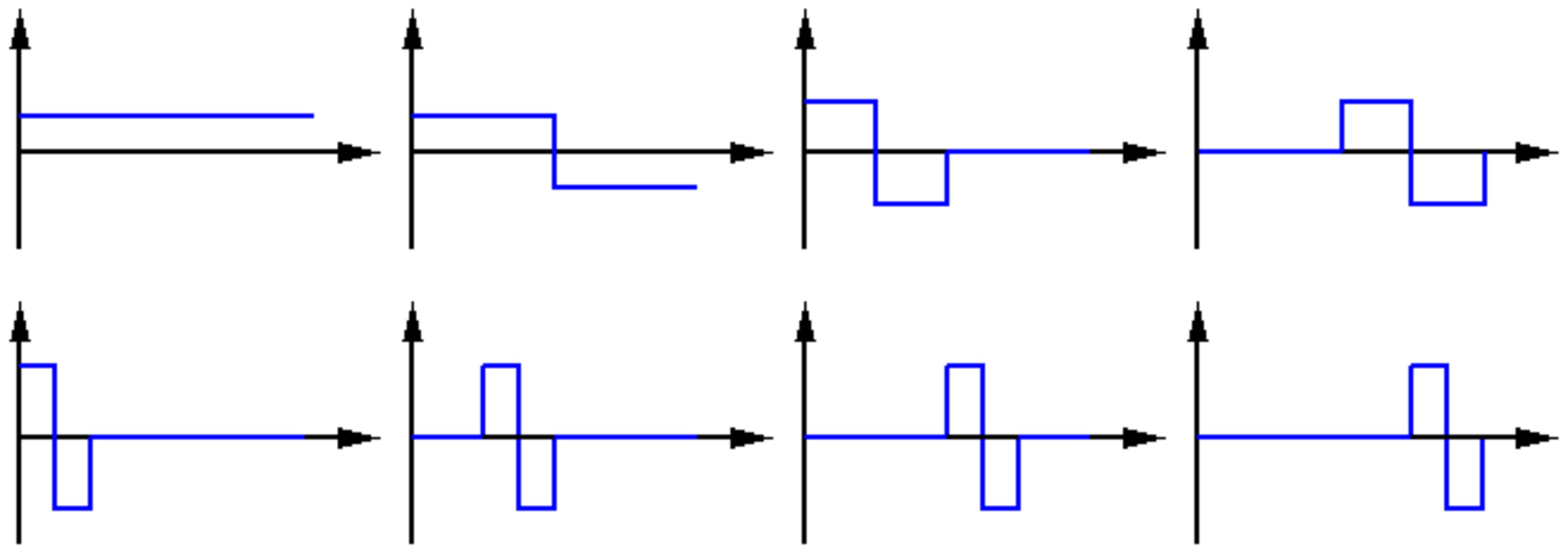


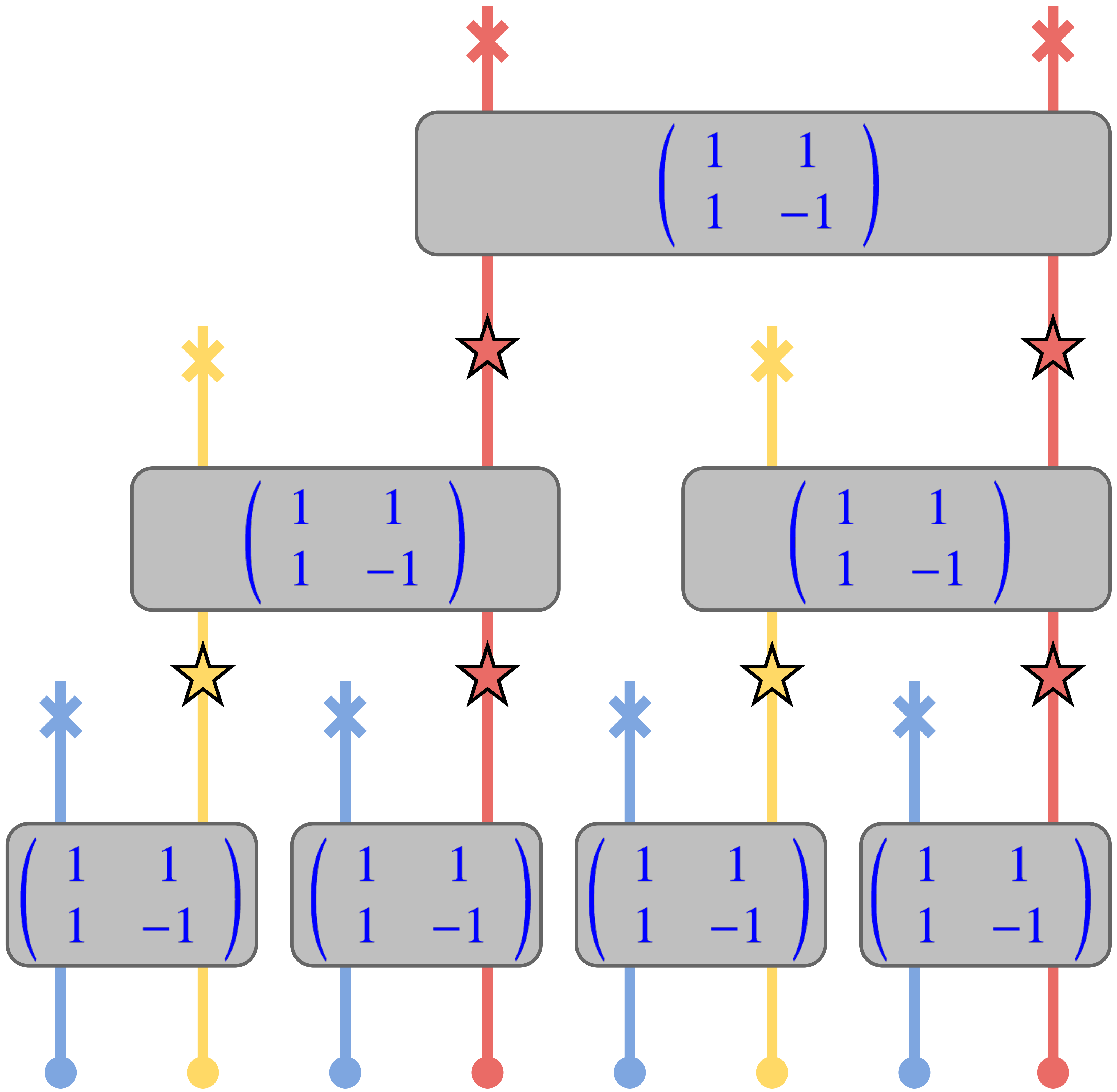


Guy, Wavelets & RG1999+ White, Evenbly, Qi, Wavelets, MERA, and holographic mapping 2013+

Wavelet transformation for Lena and Ising

$$H_8 = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & -1 & -1 & -1 & -1 \\ 1 & 1 & -1 & -1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 1 & -1 & -1 \\ 1 & -1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & -1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & -1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & -1 \end{bmatrix}.$$

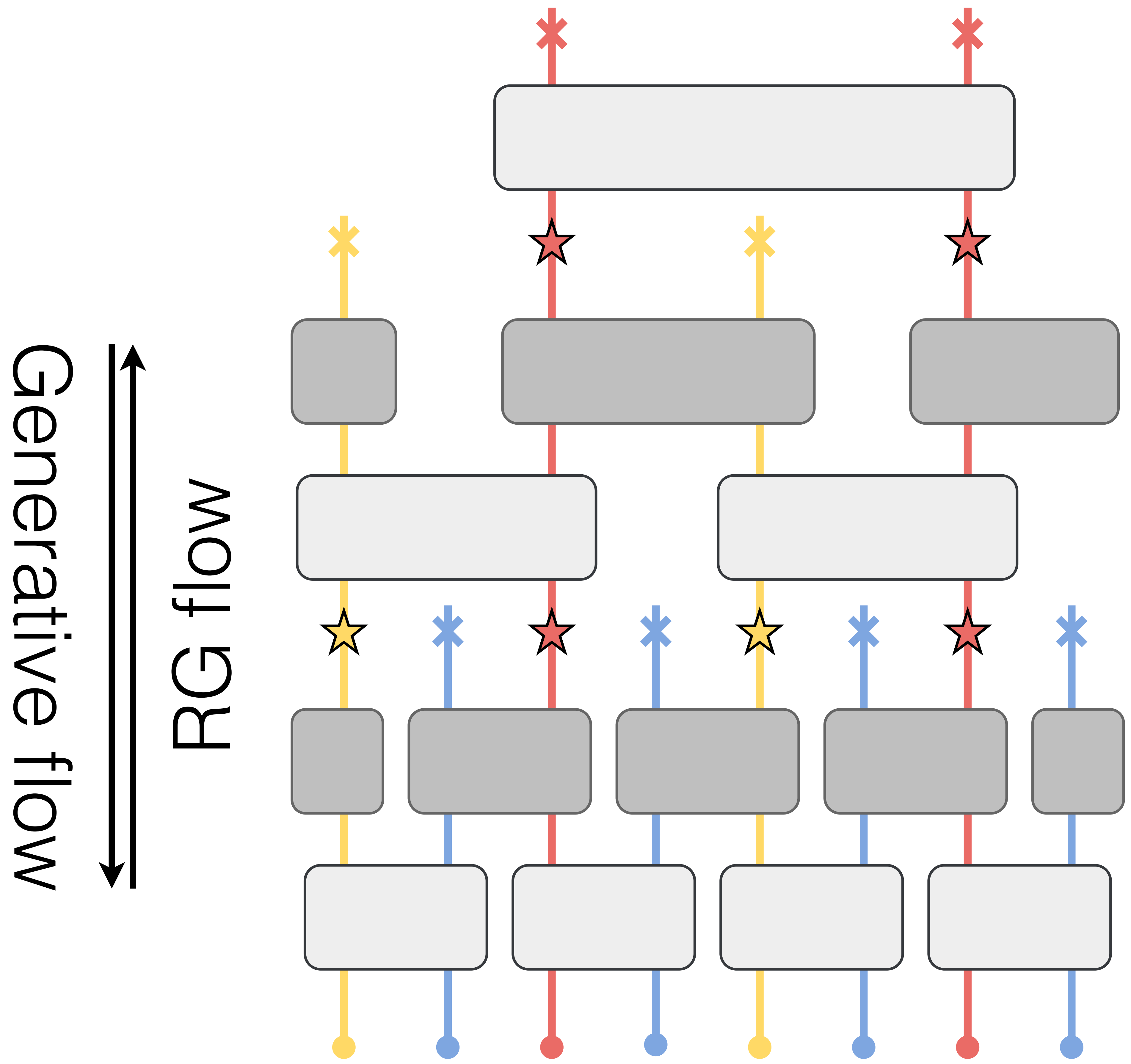




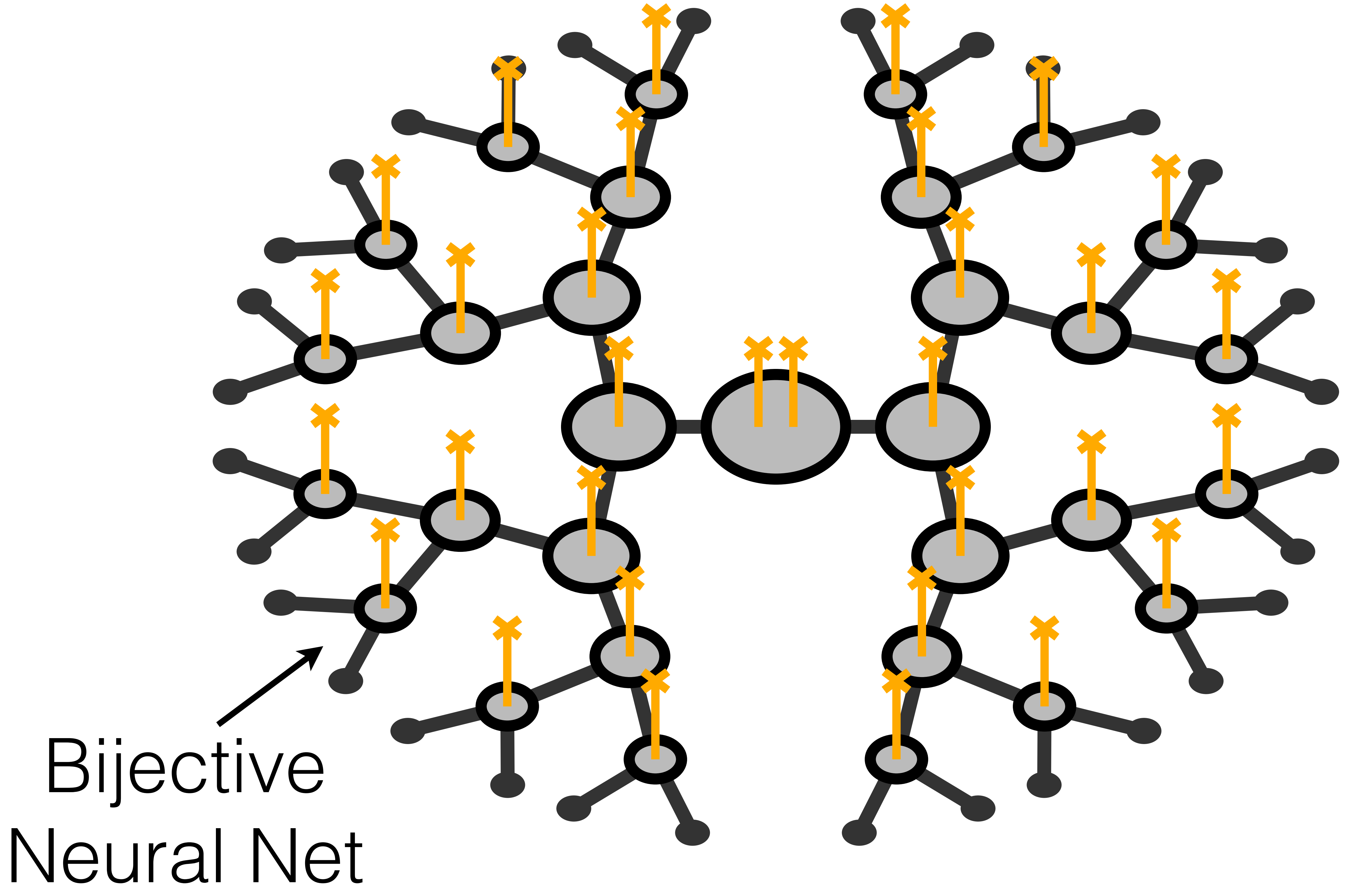
Guy, Wavelets & RG1999+ White, Evenbly, Qi, Wavelets, MERA, and holographic mapping 2013+

Neural Renormalization Group Flow

Normalizing flow with multiscale network structures



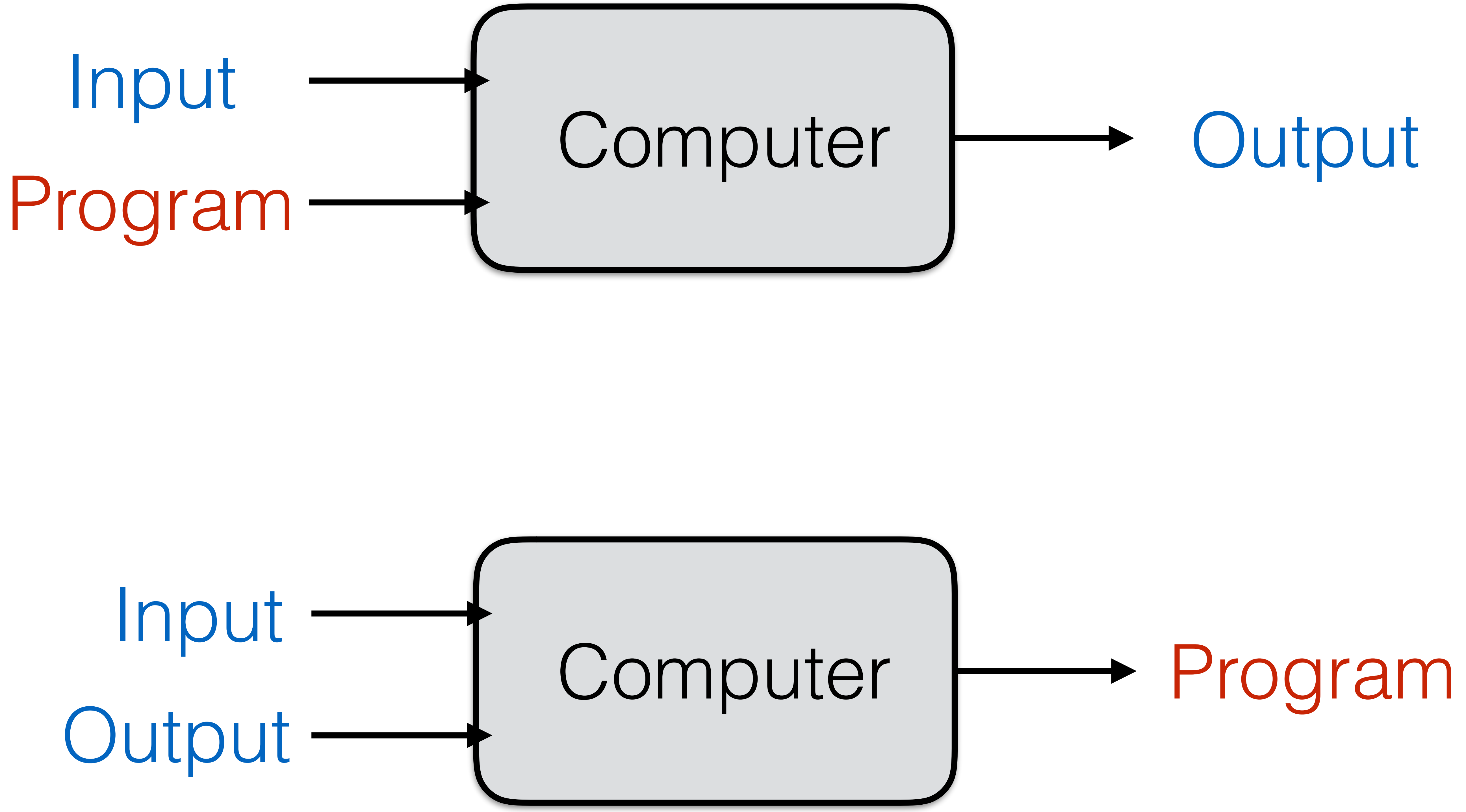
Swingle 0905.1317, Qi 1309.6282 and more



Nonlinear & adaptive generalizations of wavelets
And, a fresh approach to holographic duality

With Shuo-Hui Li
1802.02840

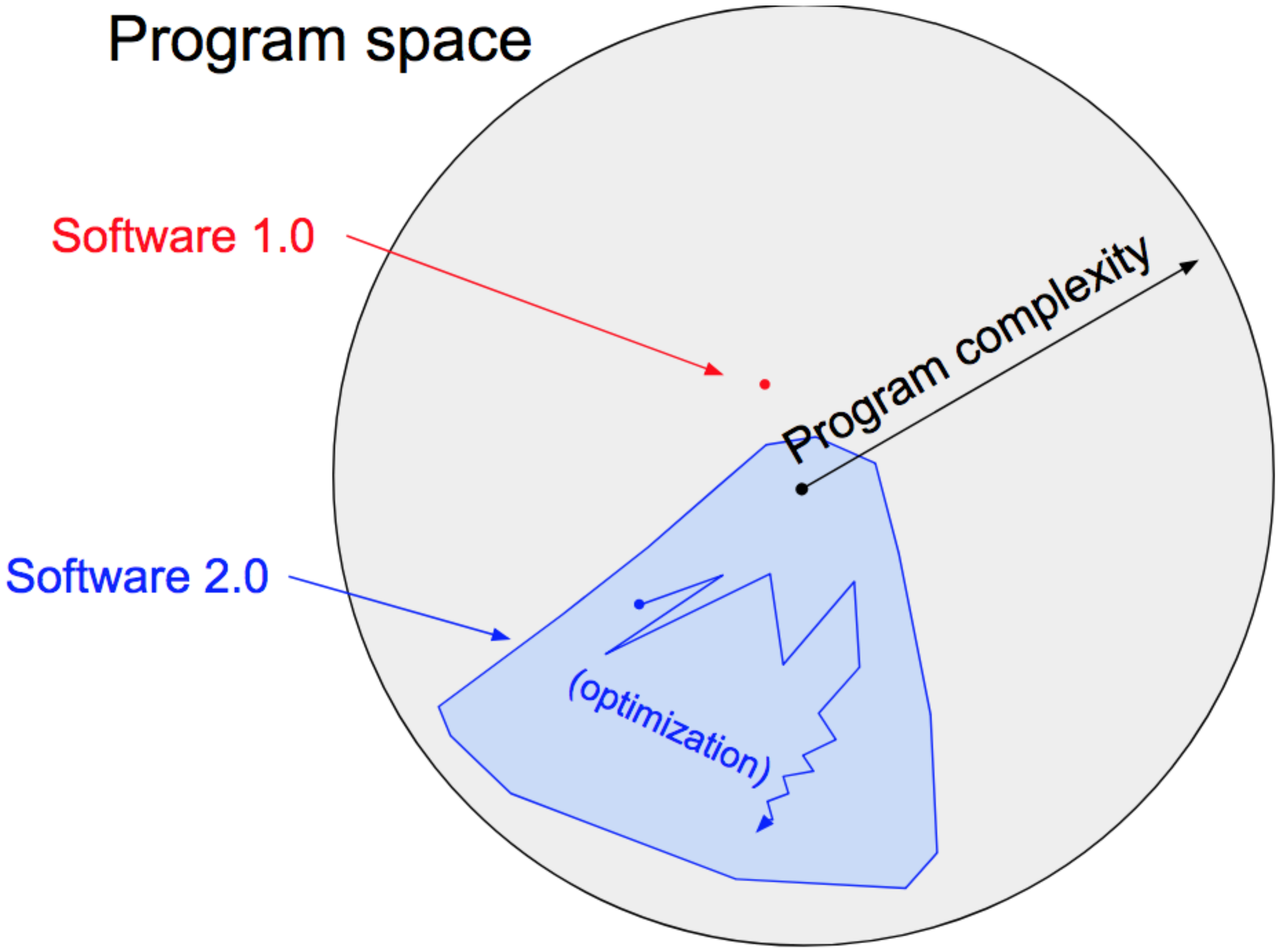
Differentiable Programming



Andrej Karpathy

Director of AI at Tesla. Previously Research Scientist at OpenAI and PhD student at Stanford. I like to train deep neural nets on large datasets.

https://medium.com/@karpathy/software-2-0-a64152b37c35



A new paradigm for programming computers

Software 2.0

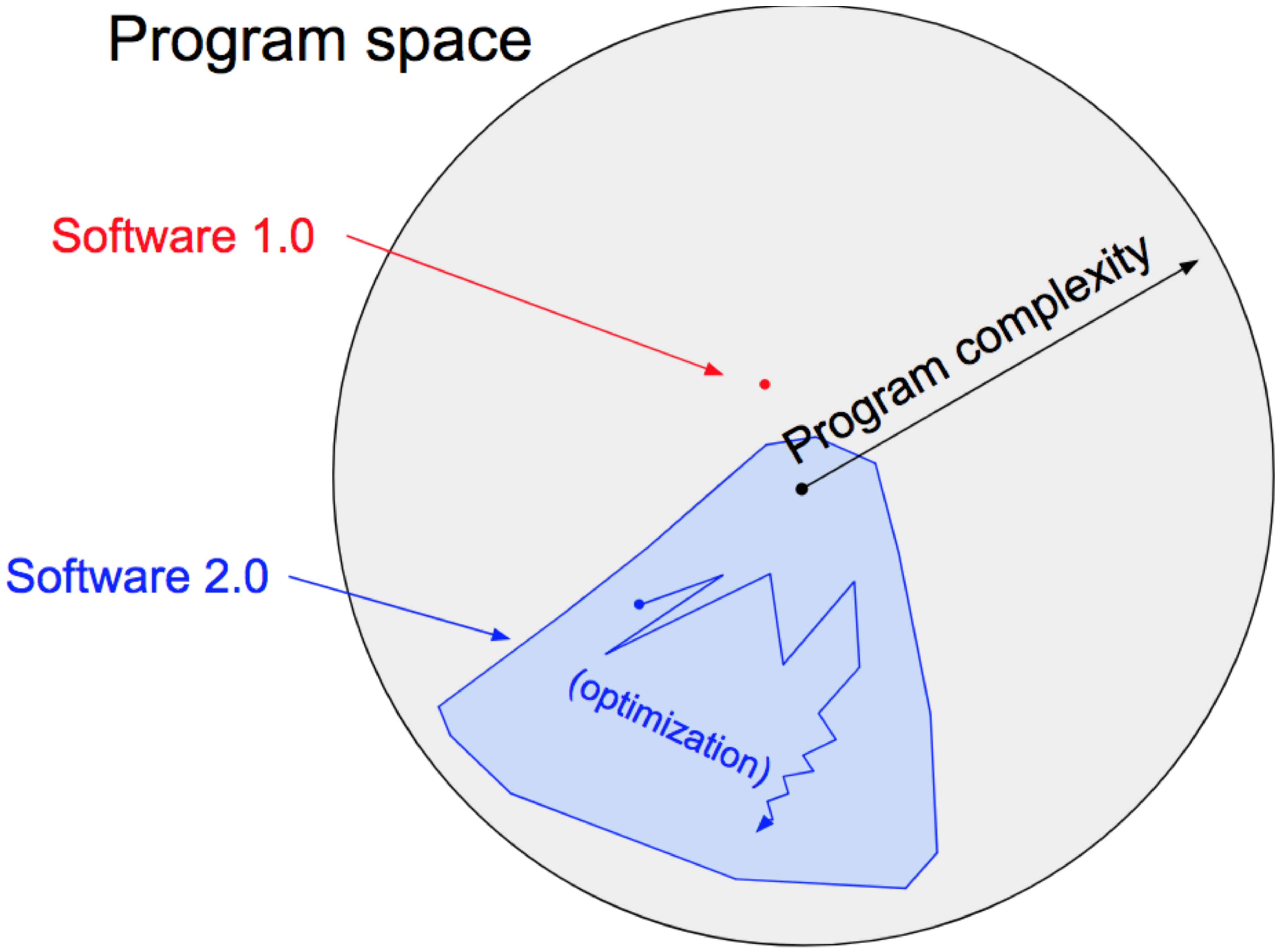
Benefits compared to 1.0

- Computationally homogeneous
- Simple to bake into silicon
- Constant running time
- Constant memory usage
- Highly portable & agile
- Modules can meld into an optimal whole
- Better than humans

Andrej Karpathy

Director of AI at Tesla. Previously Research Scientist at OpenAI and PhD student at Stanford. I like to train deep neural nets on large datasets.

https://medium.com/@karpathy/software-2-0-a64152b37c35



Writing software 2.0 by searching in the program space

Differentiable Scientific Programming

- Most linear algebra operations (Eigen, SVD!) are differentiable
- Loop/Condition/Sort/Permutations are also differentiable
- ODE integrators are differentiable with O(1) memory
- Differentiable ray tracer and Differentiable fluid simulations
- Differentiable Monte Carlo/Tensor Network/Functional RG/ Dynamical Mean Field Theory/Density Functional Theory…

Differentiable programming is more than training neural networks

Differentiable Eigensolver

$$H\Psi = \Psi\Lambda$$

Forward mode: What happen if $H = H + dH$? Perturbation theory

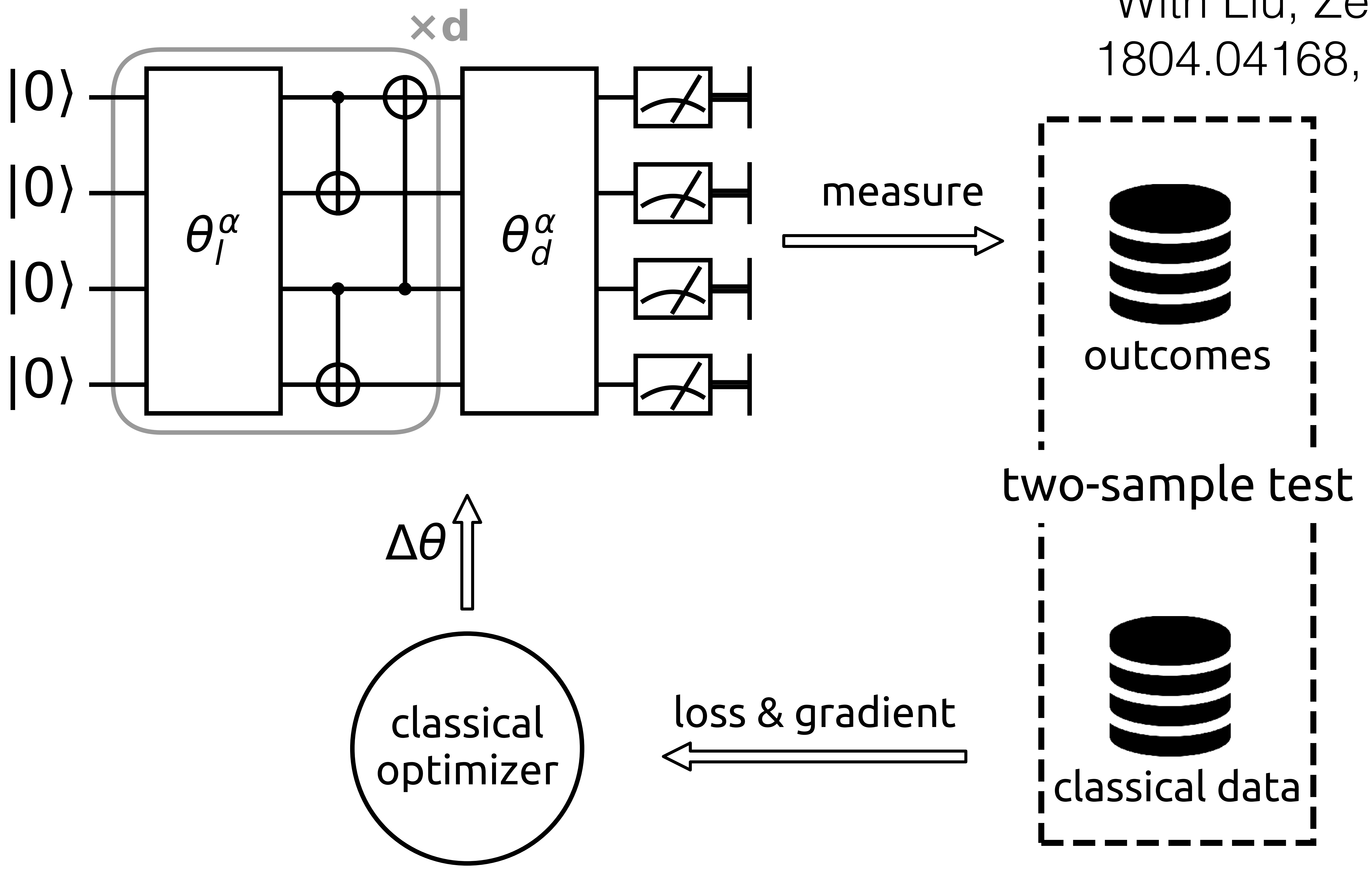
Reverse mode: How should I change H given $\partial\mathscr{L}/\partial\Psi$ and $\partial\mathscr{L}/\partial\Lambda$? **Inverse perturbation theory!**

Hamiltonian engineering via differentiable programming

https://github.com/wangleiphy/DL4CSRC/tree/master/2-ising see also Fujita et al, 1705.05372

Differentiable Quantum Programming

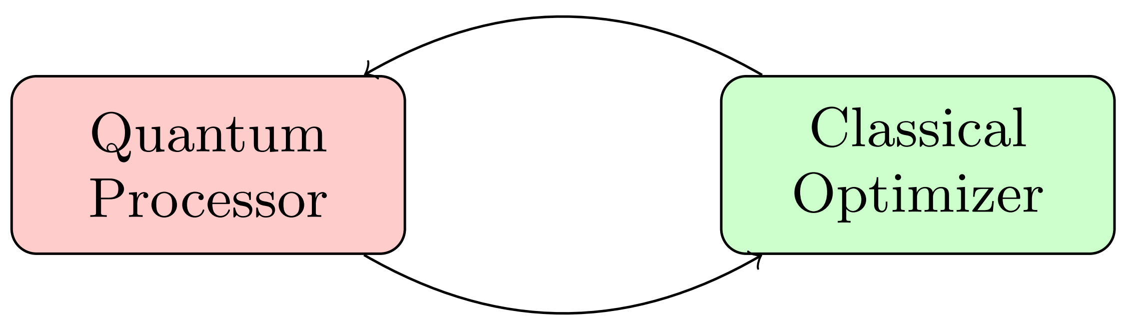
With Liu, Zeng, Wu, Hu
1804.04168, 1808.03425



Train the quantum circuit as a probabilistic generative model

Quantum sampling complexity underlines the "quantum supremacy"

Differentiable Quantum Programming

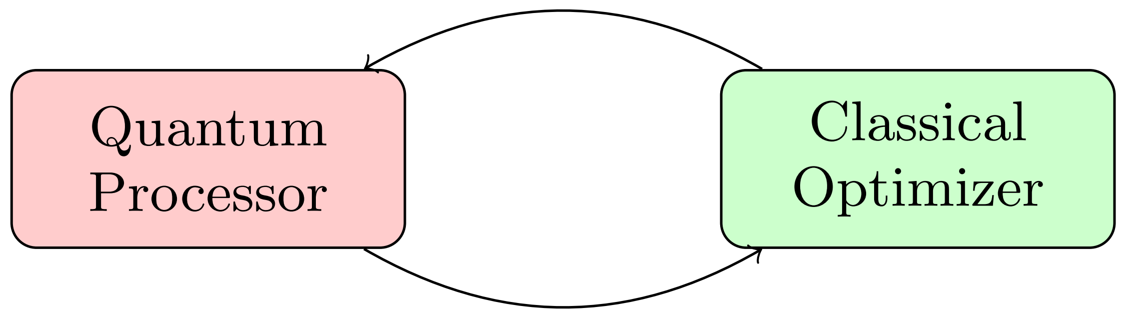


- Variational quantum eigensovler (VQE)
- Quantum approximate optimization algorithm (QAOA)
- Quantum pattern recognition
- Quantum circuit Born machine (QCBM)

…

Quantum circuit classifier	Farhi, Neven, 1802.06002 Havlicek et al, 1804.11326
Born machine experiment	Benedetti, Garcia-Pintos, Nam, Perdomo-Ortiz, 1801.07686
TNS inspired circuit architecture	Huggins, Patel, Whaley, Stoudenmire, 1803.11537
Quantum generative model	Gao, Zhang, Duan, 1711.02038
Quantum adversarial training	Dallaire-Demers, Lloyd, Benedetti 1804.08641,1804.09139, 1806.00463

Differentiable Quantum Programming



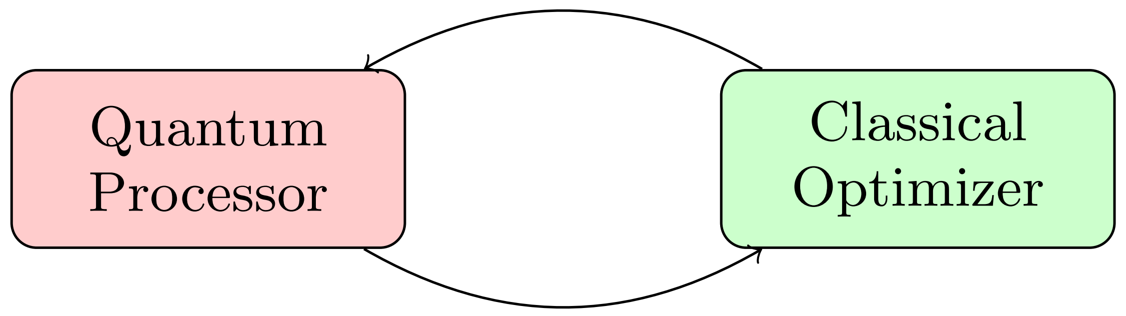
It is a paradigm beyond quantum-classical hybrid

- Variational quantum eigensovler (VQE)
- Quantum approximate optimization algorithm (QAOA)
- Quantum pattern recognition
- Quantum circuit Born machine (QCBM)

…

Quantum circuit classifier	Farhi, Neven, 1802.06002 Havlicek et al, 1804.11326
Born machine experiment	Benedetti, Garcia-Pintos, Nam, Perdomo-Ortiz, 1801.07686
TNS inspired circuit architecture	Huggins, Patel, Whaley, Stoudenmire, 1803.11537
Quantum generative model	Gao, Zhang, Duan, 1711.02038
Quantum adversarial training	Dallaire-Demers, Lloyd, Benedetti 1804.08641,1804.09139, 1806.00463

Differentiable Quantum Programming



It is a paradigm beyond quantum-classical hybrid

- Variational quantum eigensovler (VQE)
- Quantum approximate optimization algorithm (QAOA)
- Quantum pattern recognition
- Quantum circuit Born machine (QCBM)

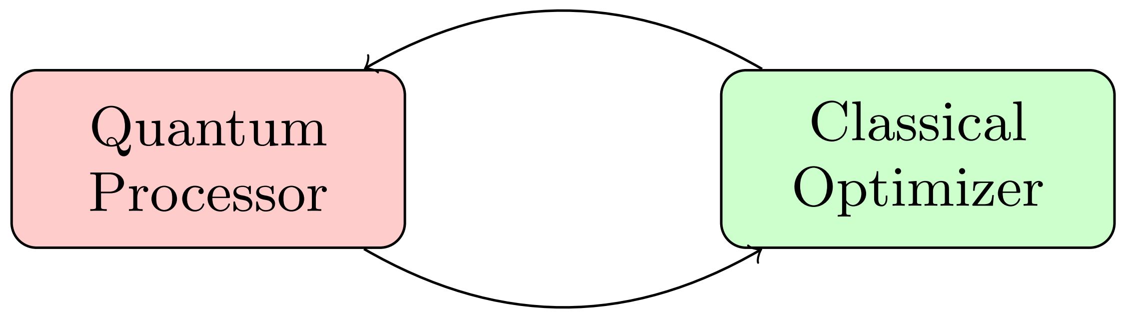
…

Short term：
What can we do with circuits of limited depth ?

Long term:
Are we really good at programing quantum computers ?

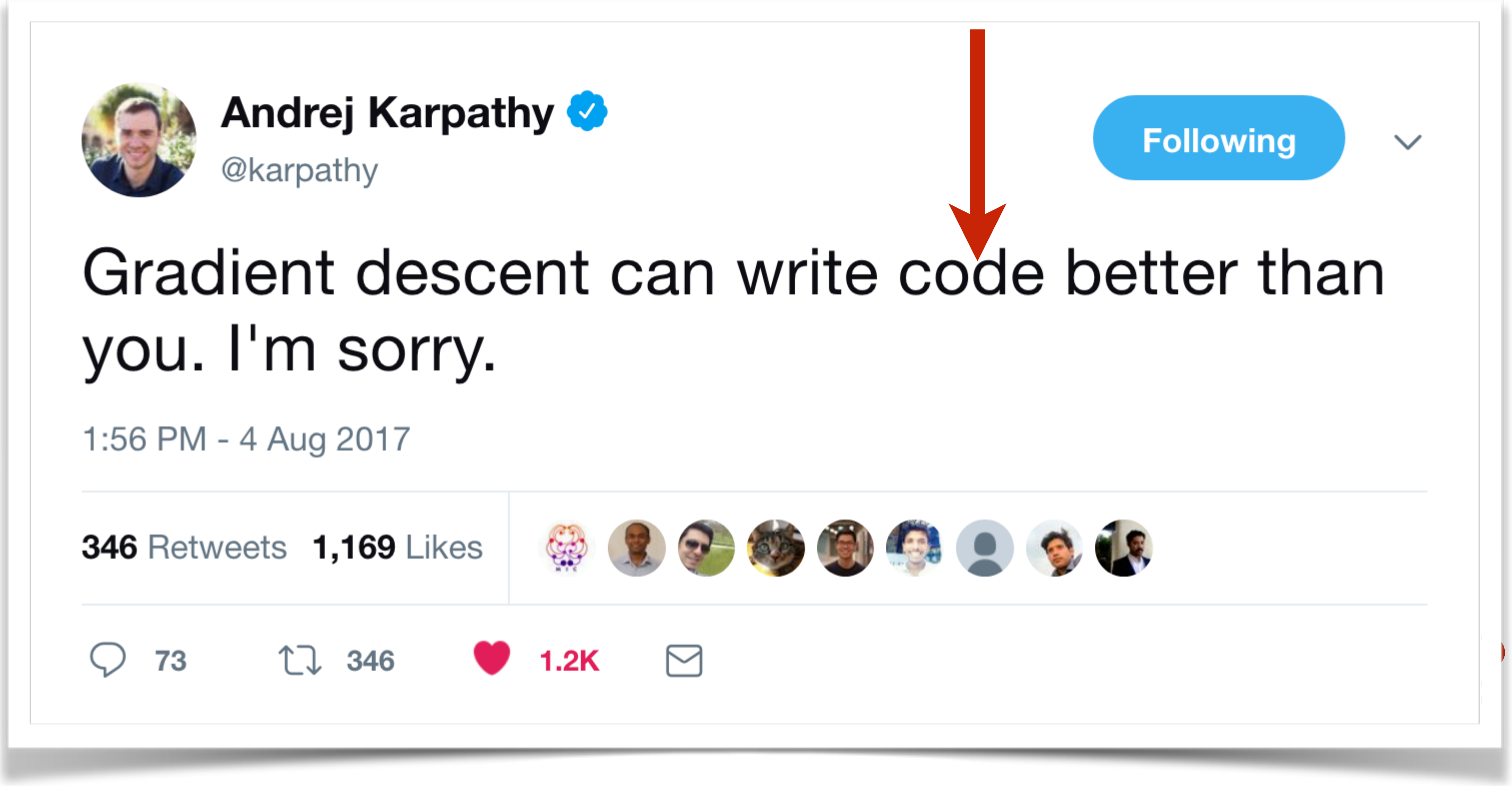
Quantum circuit classifier	Farhi, Neven, 1802.06002 Havlicek et al, 1804.11326
Born machine experiment	Benedetti, Garcia-Pintos, Nam, Perdomo-Ortiz, 1801.07686
TNS inspired circuit architecture	Huggins, Patel, Whaley, Stoudenmire, 1803.11537
Quantum generative model	Gao, Zhang, Duan, 1711.02038
Quantum adversarial training	Dallaire-Demers, Lloyd, Benedetti 1804.08641,1804.09139, 1806.00463

Differentiable Quantum Programming



It is a paradigm beyond quantum-classical hybrid

Quantum code

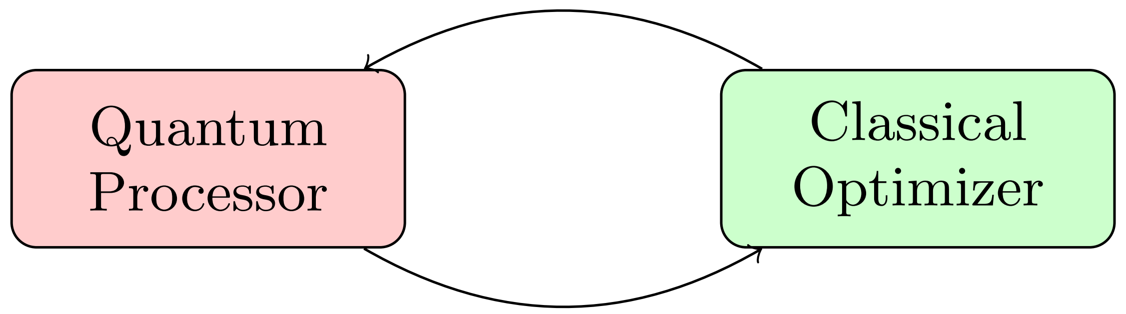


- Variational quantum eigensovler (VQE)
- Quantum approximate optimization algorithm (QAOA)
- Quantum pattern recognition
- Quantum circuit Born machine (QCBM)

…

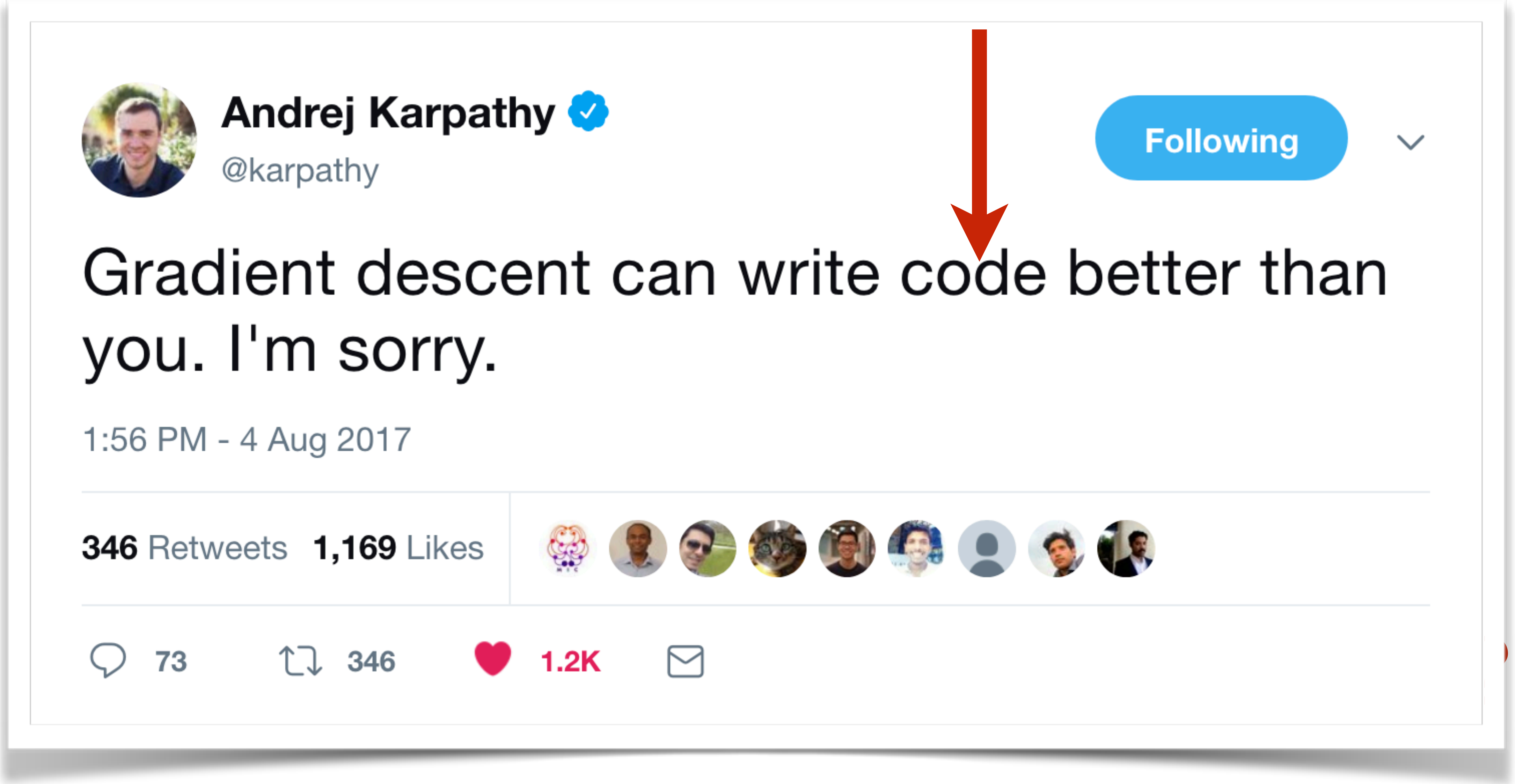
Quantum circuit classifier	Farhi, Neven, 1802.06002 Havlicek et al, 1804.11326
Born machine experiment	Benedetti, Garcia-Pintos, Nam, Perdomo-Ortiz, 1801.07686
TNS inspired circuit architecture	Huggins, Patel, Whaley, Stoudenmire, 1803.11537
Quantum generative model	Gao, Zhang, Duan, 1711.02038
Quantum adversarial training	Dallaire-Demers, Lloyd, Benedetti 1804.08641,1804.09139, 1806.00463

Differentiable Quantum Programming



It is a paradigm beyond quantum-classical hybrid

Quantum code



- Variational quantum eigensovler (VQE)
- Quantum approximate optimization algorithm (QAOA)
- Quantum pattern recognition
- Quantum circuit Born machine (QCBM)

…

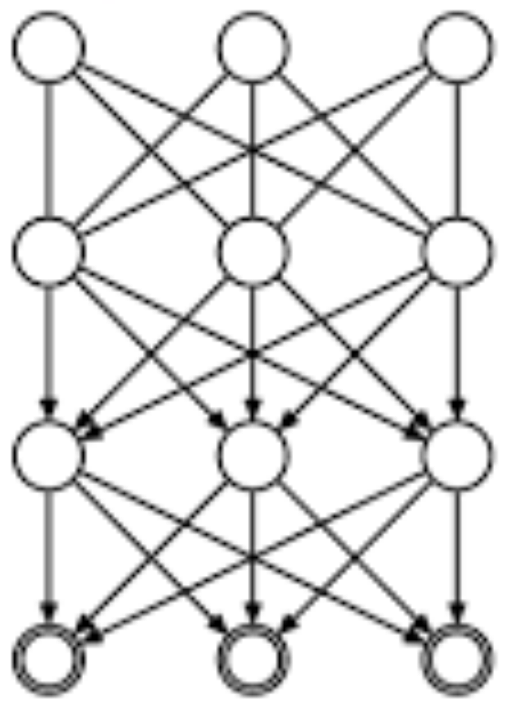
Quantum circuit
Born machine ex
TNS inspired cir
Quantum genera
Quantum advers

Quantum Software 2.0

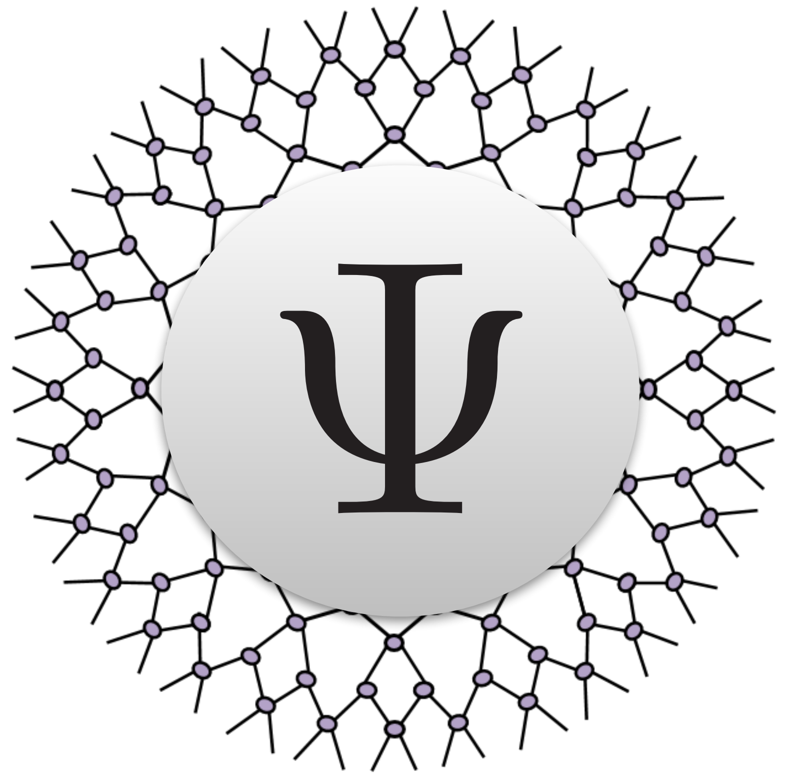
1806.00463

Summary

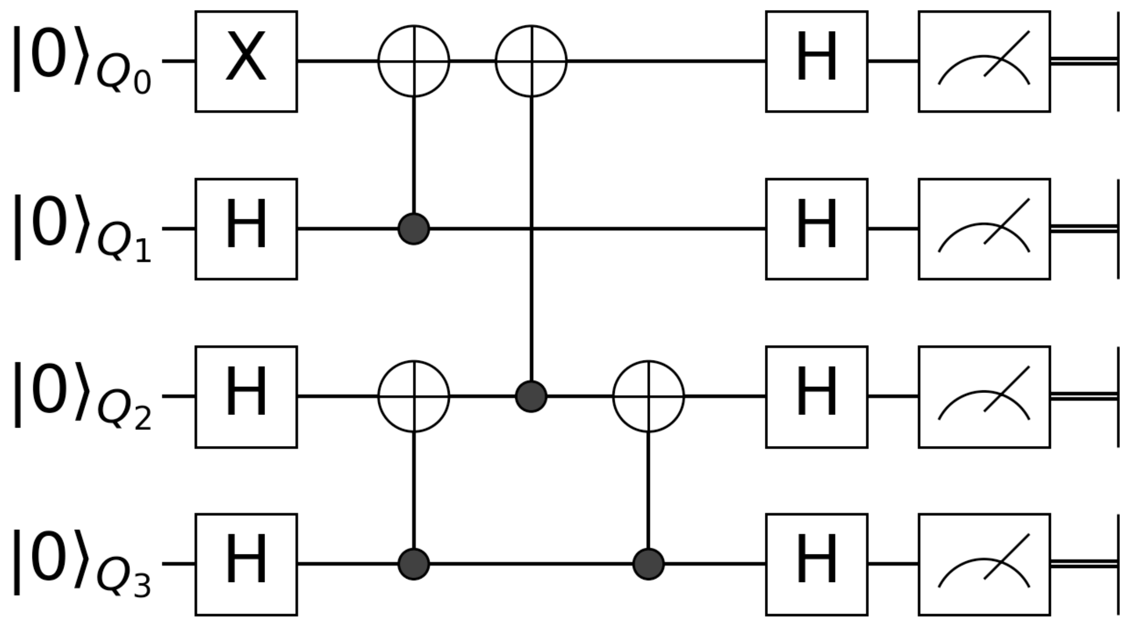
Neural Networks



Tensor Networks



Quantum Circuits

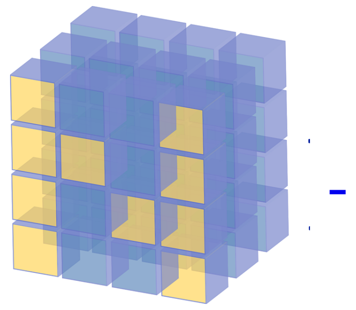


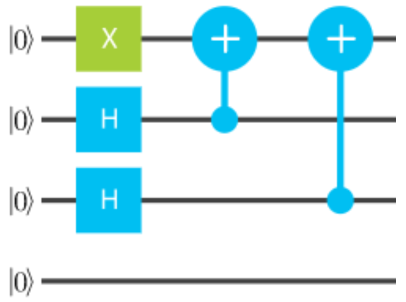
“三重境界”

1. Function Approximation
2. Probabilistic Transformation
3. Information Processing Device

Try it yourself!

https://github.com/wangleiphy/TRG
https://github.com/li012589/NeuralRG
https://github.com/wangleiphy/MongeAmpereFlow





https://github.com/GiggleLiu/QuantumCircuitBornMachine

https://github.com/QuantumBFS/Yao.jl/

Refs

1802.02840 1804.04168
1808.03425 1809.10188

Thank You!

Pan Zhang **Song Cheng** **Jin-Guo Liu** **Weinan E**
Shuo-Hui Li **Xiu-Zhe Luo** **Jinfeng Zeng** **Linfeng Zhang**

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

Google Colab
free GPU access

Lecture Note on Deep Learning and Quantum Many-Body Computation

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

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

February 14, 2018

Abstract

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

* wanglei@iphy.ac.cn

CONTENTS

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

Google Colab
free GPU access

Lecture Note on Deep Learning and Quantum Many-Body Computation

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

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

February 14, 2018

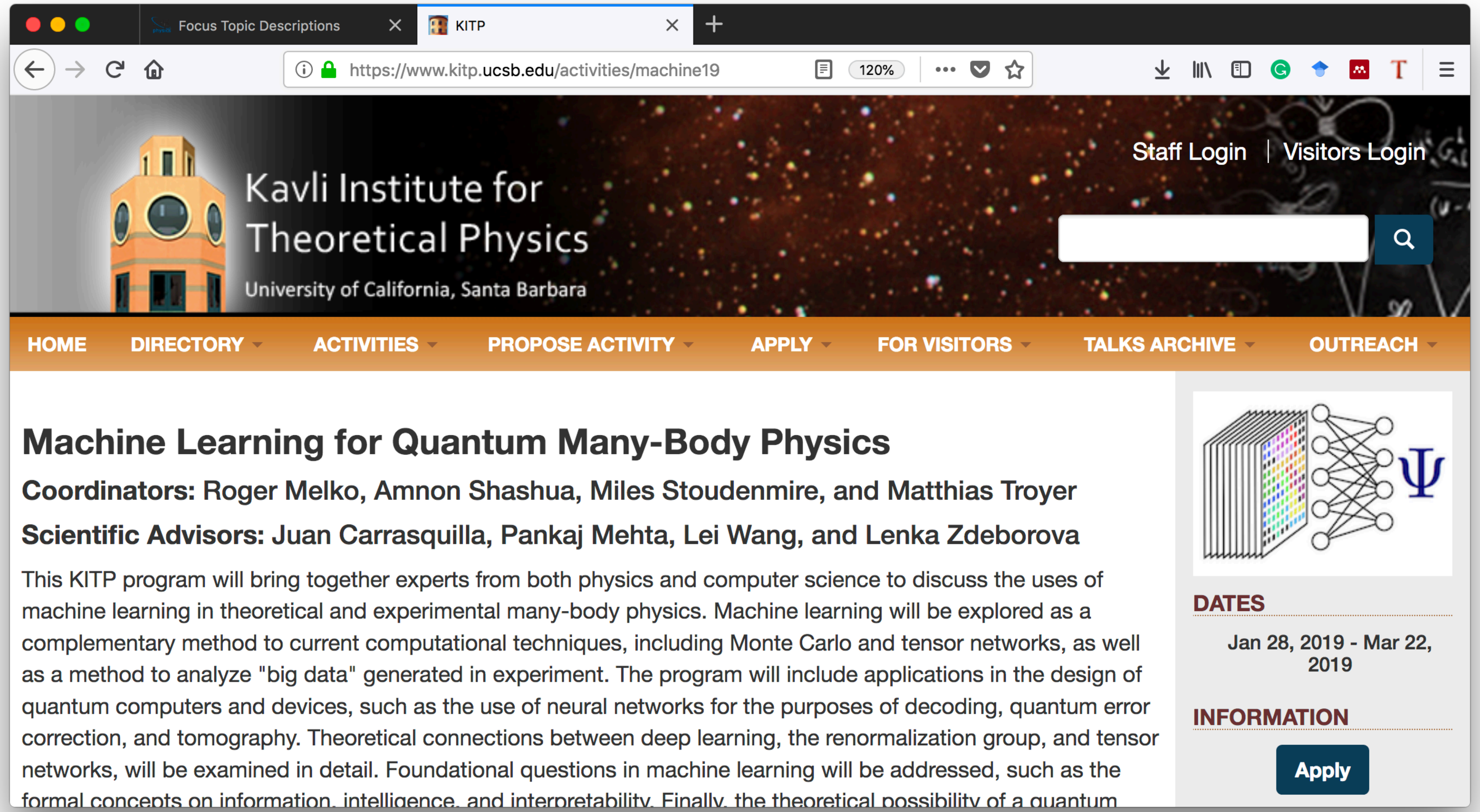
Abstract

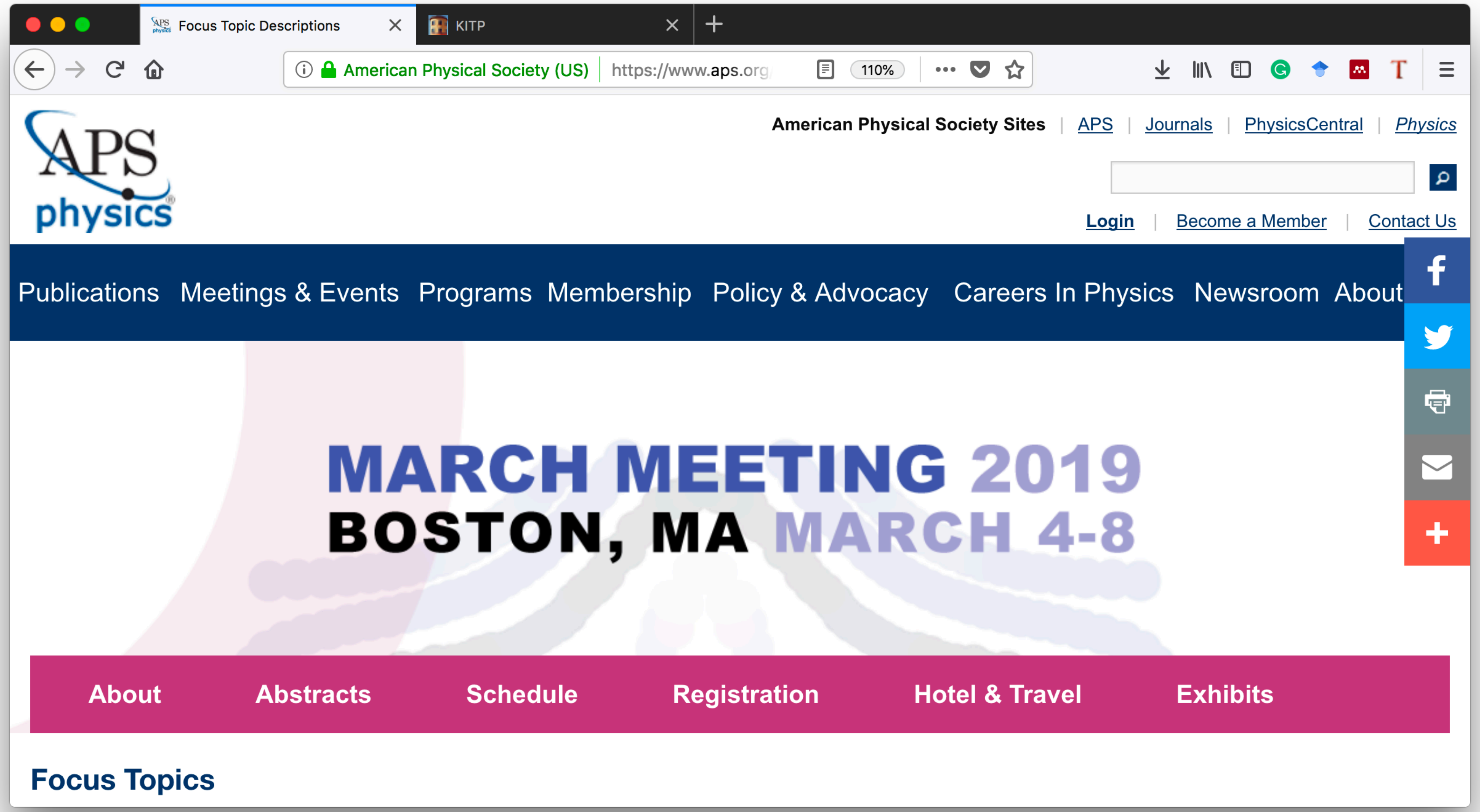
This note introduces deep learning from a computational quantum physicist's perspective. The focus is on deep learning's impacts to quantum many-body computation, and vice versa. The latest version of the note is at http://wangleiphy.github.io/. Please send comments, suggestions and corrections to the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

* wanglei@iphy.ac.cn

CONTENTS

To catch up with the latest updates





KITP, Santa Barbara Program
ML for Quantum Many-Body Physics
Jan 28-Mar 22, 2019

APS March meeting focus session
ML in Condensed Matter Physics
Boston, Mar 4-8, 2019