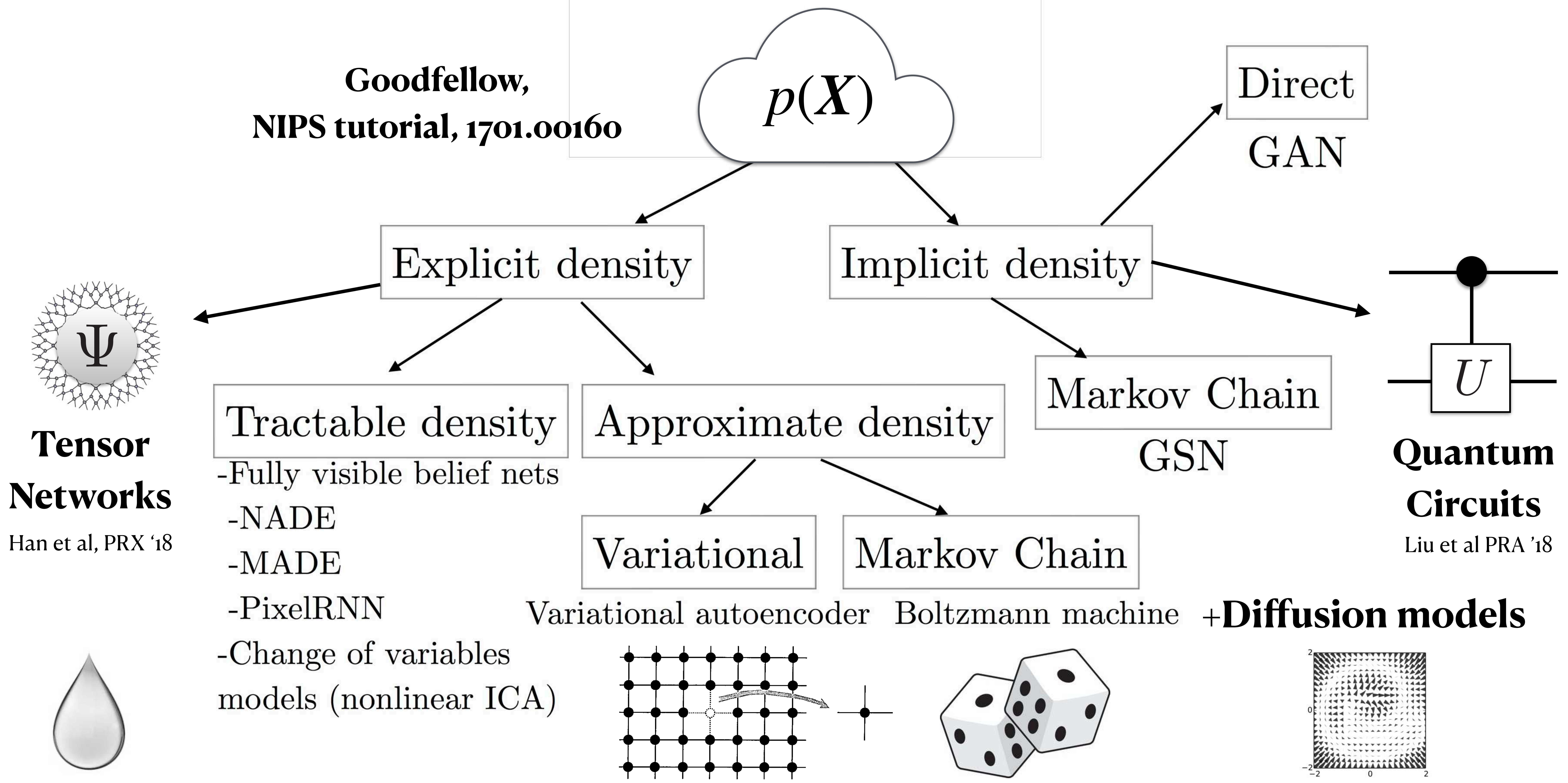
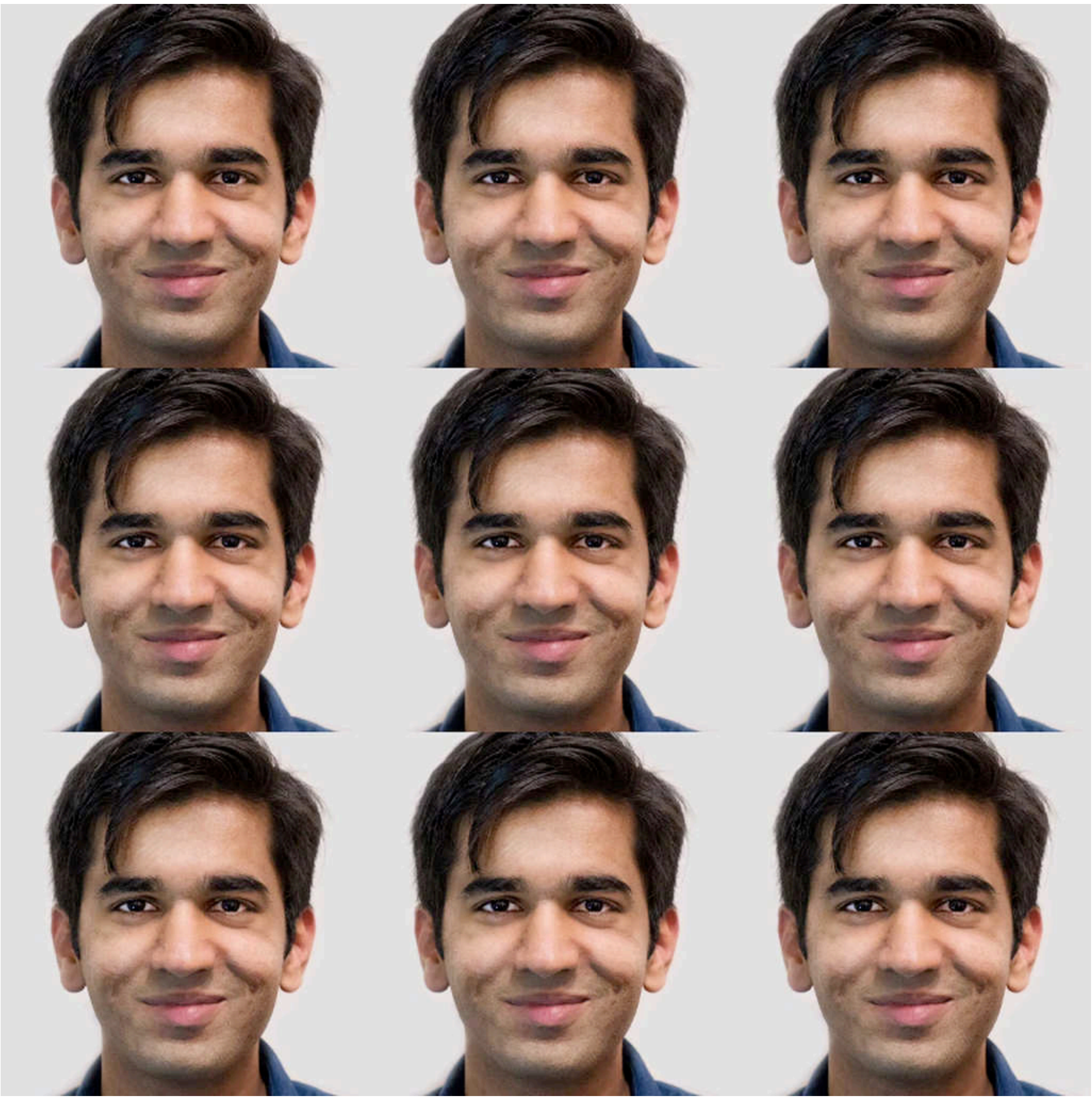
A physicist's perspective on generative models

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

Generative models and their physics genes



Normalizing flows



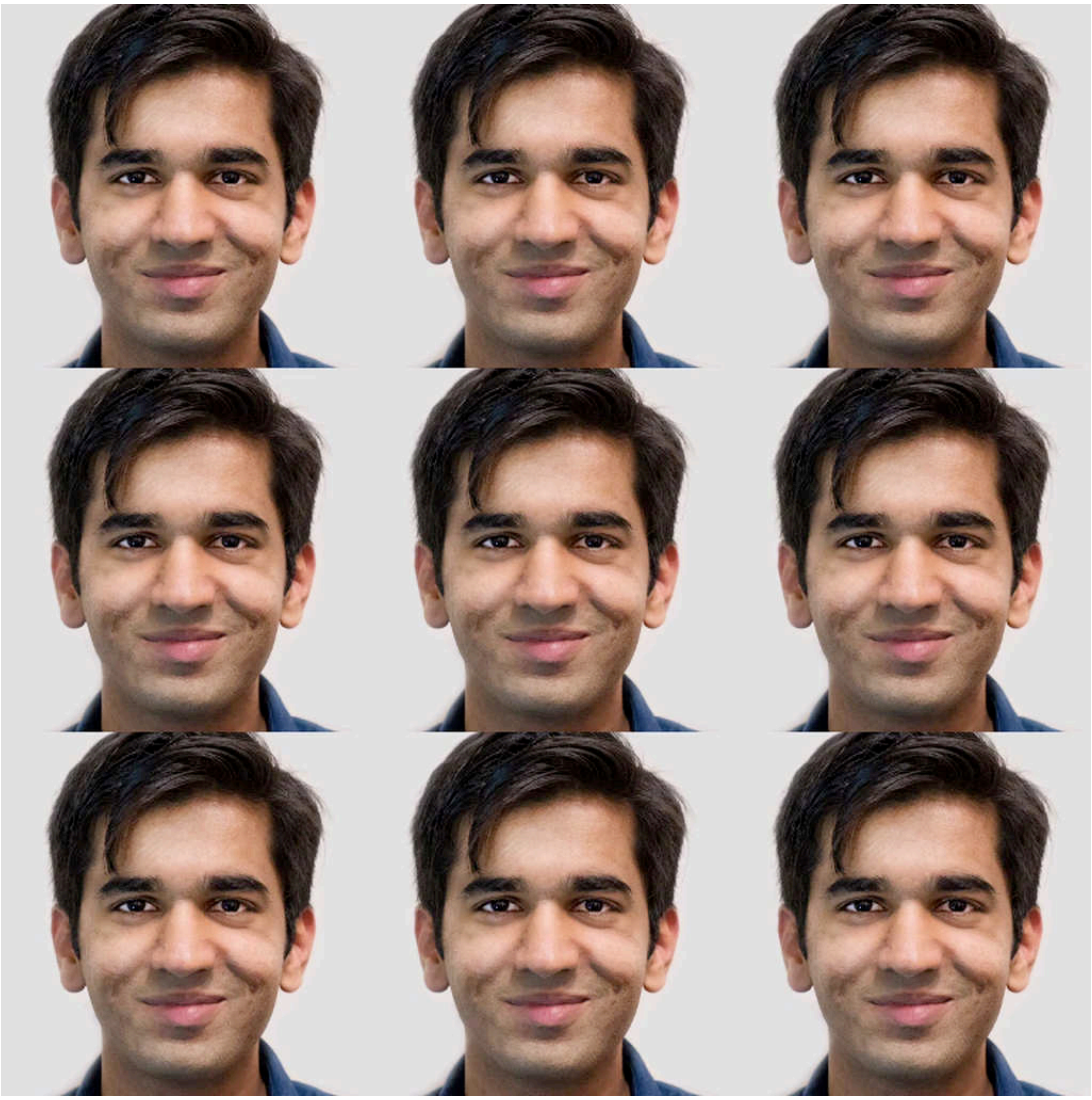
Parallel WaveNet 1711.10433

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

Glow 1807.03039

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

Normalizing flows



Parallel WaveNet 1711.10433

Glow 1807.03039

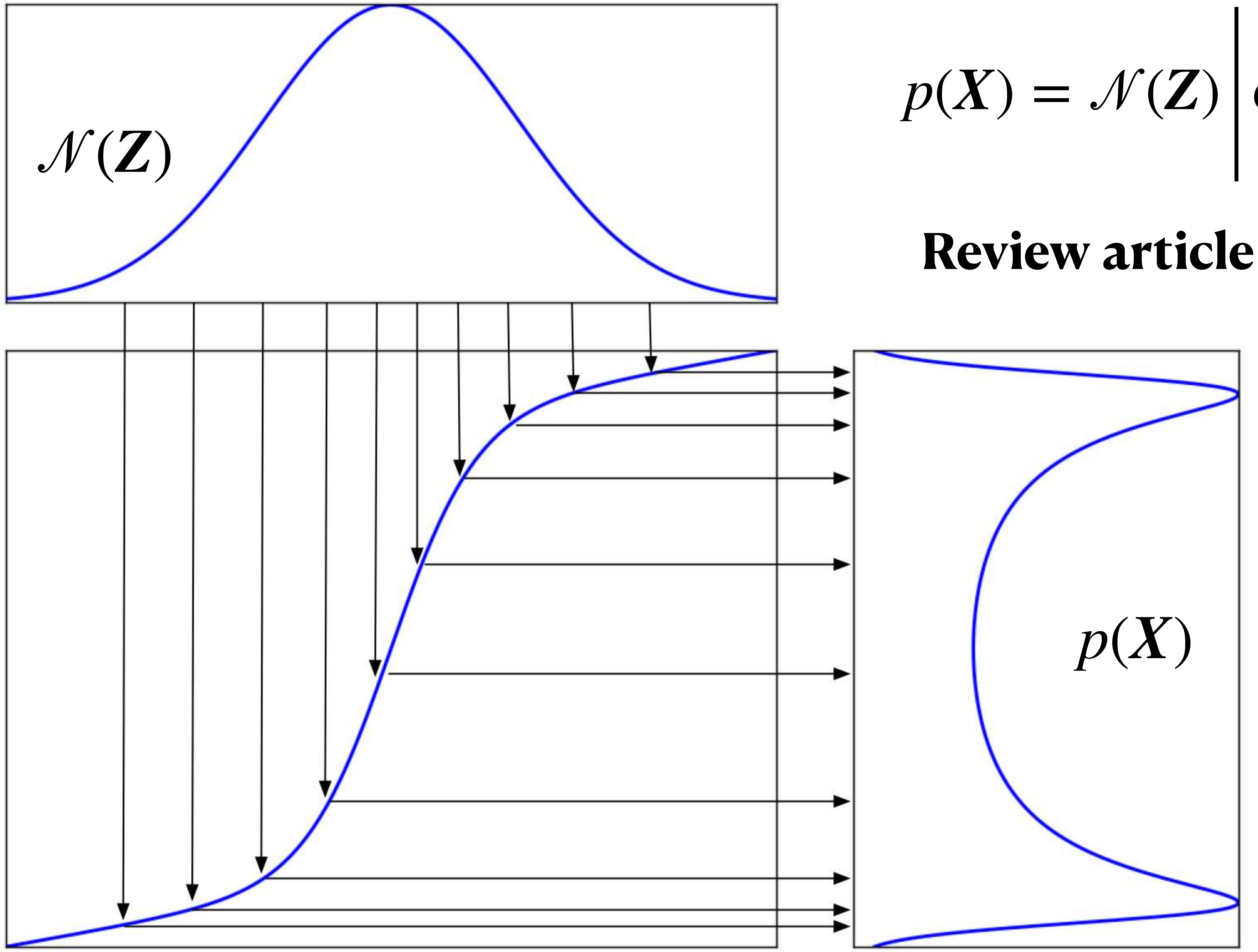
https://deepmind.com/blog/high-fidelity-speech-synthesis-wavenet/

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

Normalizing flow in a nutshell

Base density

"neural net" with 1 neuron



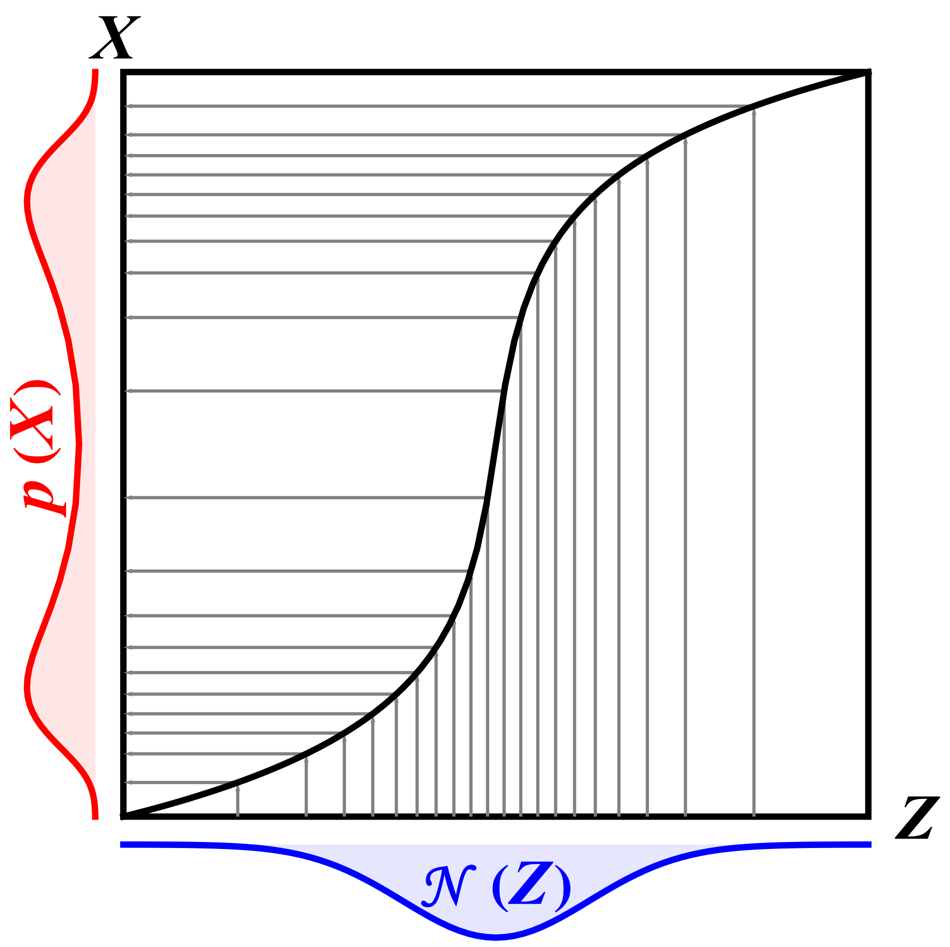
Target density

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

Review article 1912.02762

Flow model

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



Sampling

$$\boldsymbol{Z} \sim \mathcal{N}(\boldsymbol{Z})$$

$$\boldsymbol{Z} \mapsto \boldsymbol{X}$$

Normalization

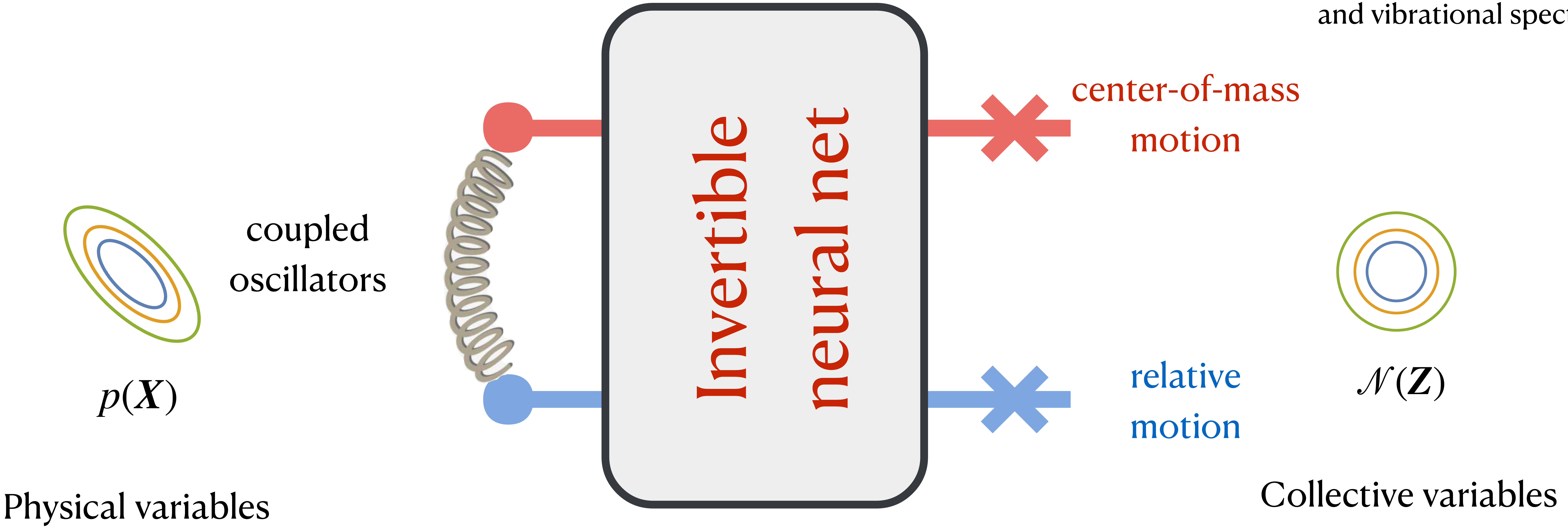
$$\int dX \left| \det \left(\frac{\partial Z}{\partial X} \right) \right| = \int dZ$$

From flow to diffusion, and back

A unified perspective to transportation-based generative model

Physics intuition of normalizing flow

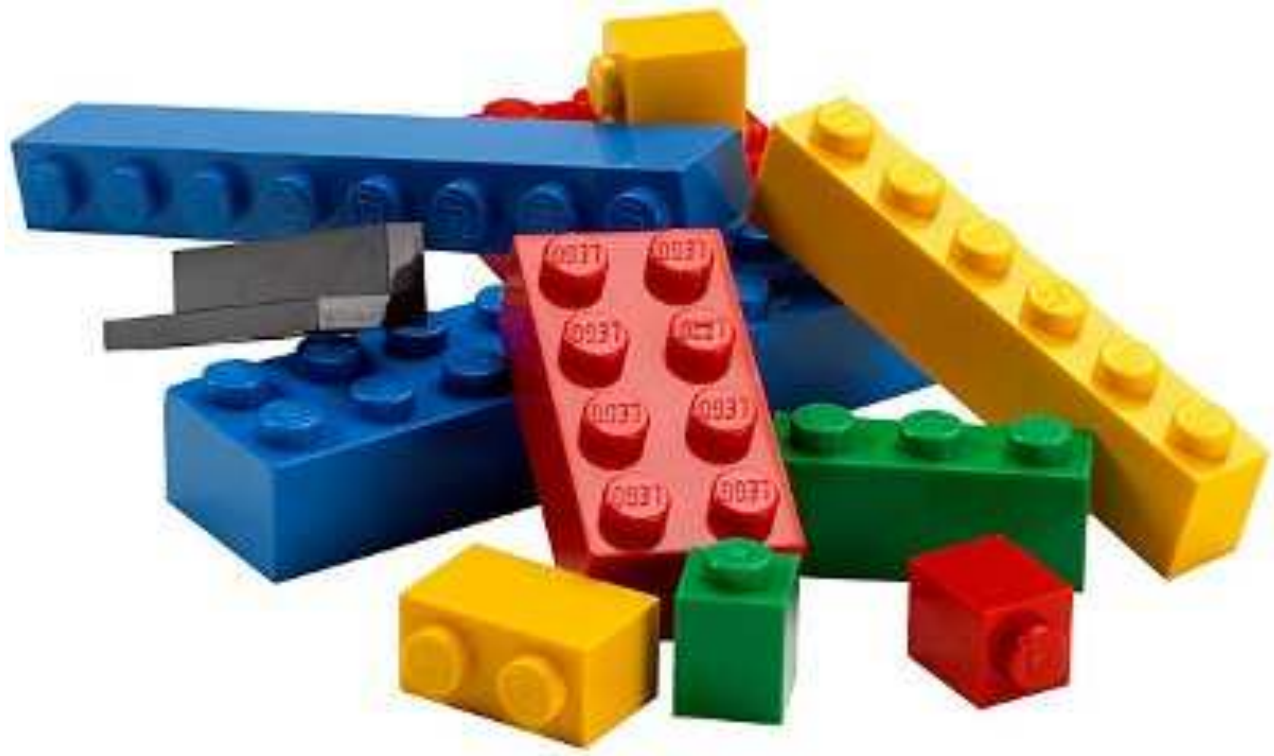
Li and LW, PRL '18
Li, Dong, Zhang, LW, PRX '20
Zhang, Wang, LW, JCP '24
for RG, canonical transformations, and vibrational spectra



High-dimensional, composable, learnable, nonlinear transformations

Flow architecture design

Composability

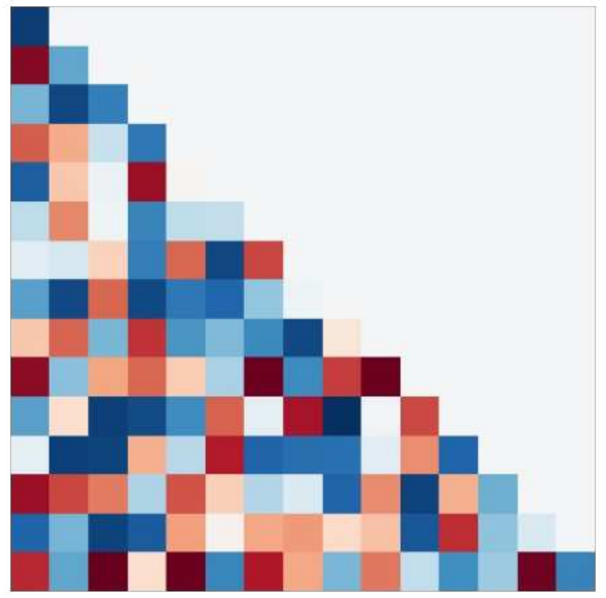


$$\boldsymbol{Z} = \mathcal{T}(\boldsymbol{X})$$

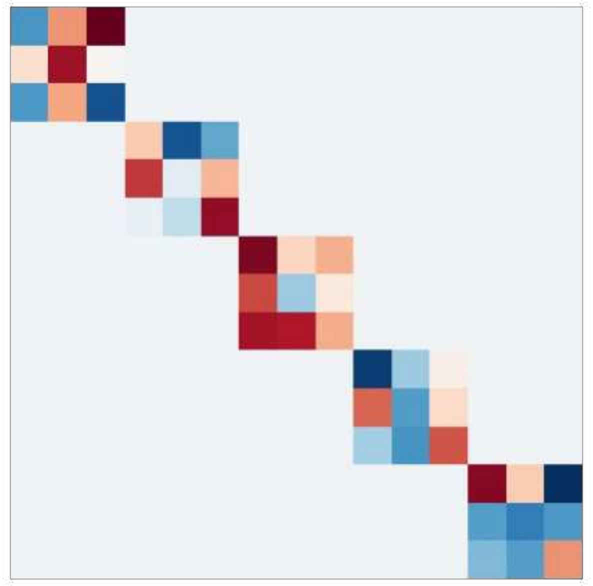
$$\mathcal{T} = \mathcal{T}_1 \circ \mathcal{T}_2 \circ \mathcal{T}_3 \circ \cdots$$

Balanced efficiency & inductive bias

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



Autoregressive



Blockwise

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

Continuous flow

Example of a building block

Forward

arbitrary neural nets

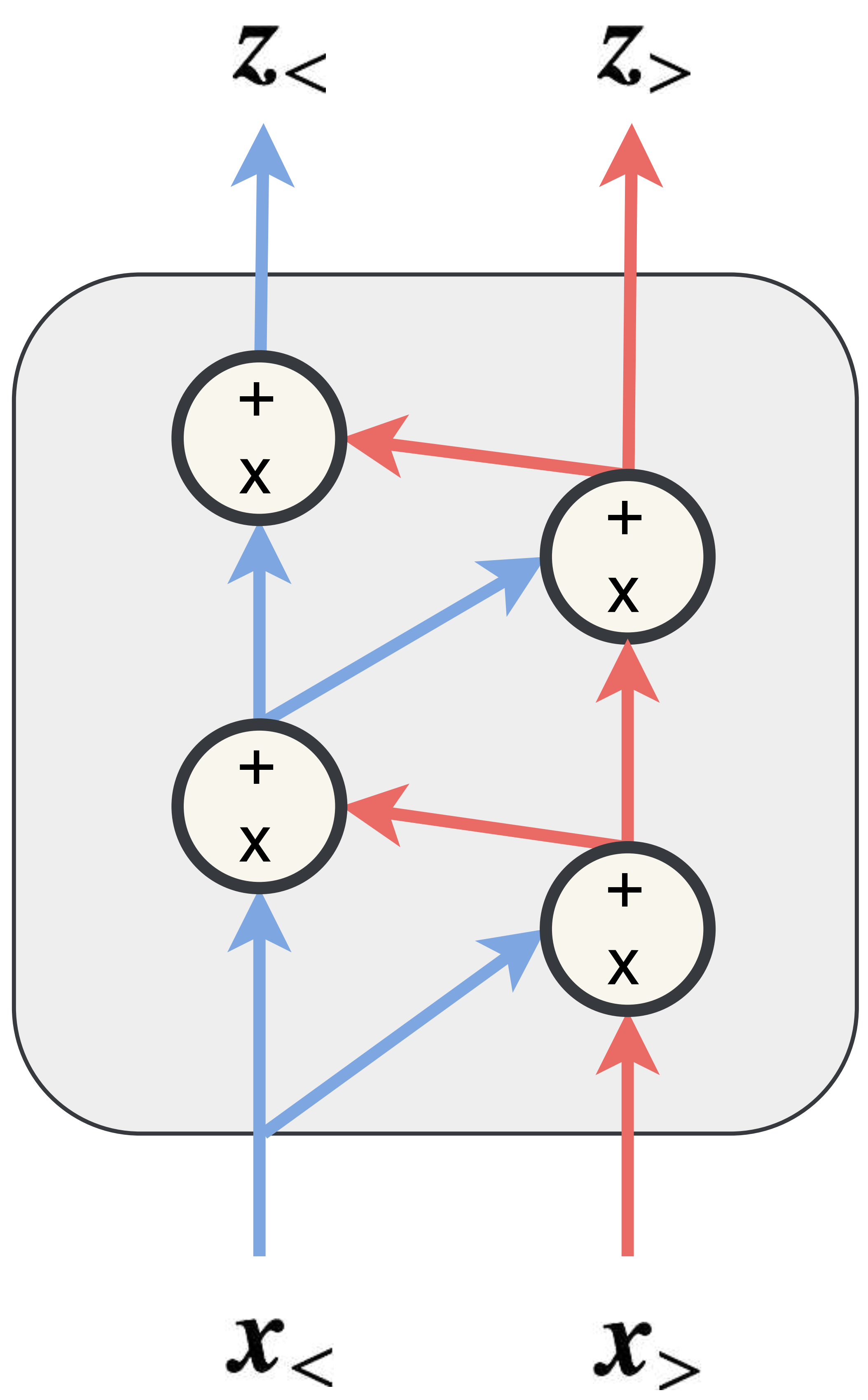
$$\begin{cases} \boldsymbol{x}_< = \boldsymbol{z}_< \\ \boldsymbol{x}_> = \boldsymbol{z}_> \odot e^{s(\boldsymbol{z}_<)} + t(\boldsymbol{z}_<) \end{cases}$$

Inverse

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

Log-Abs-Jacobian-Det

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

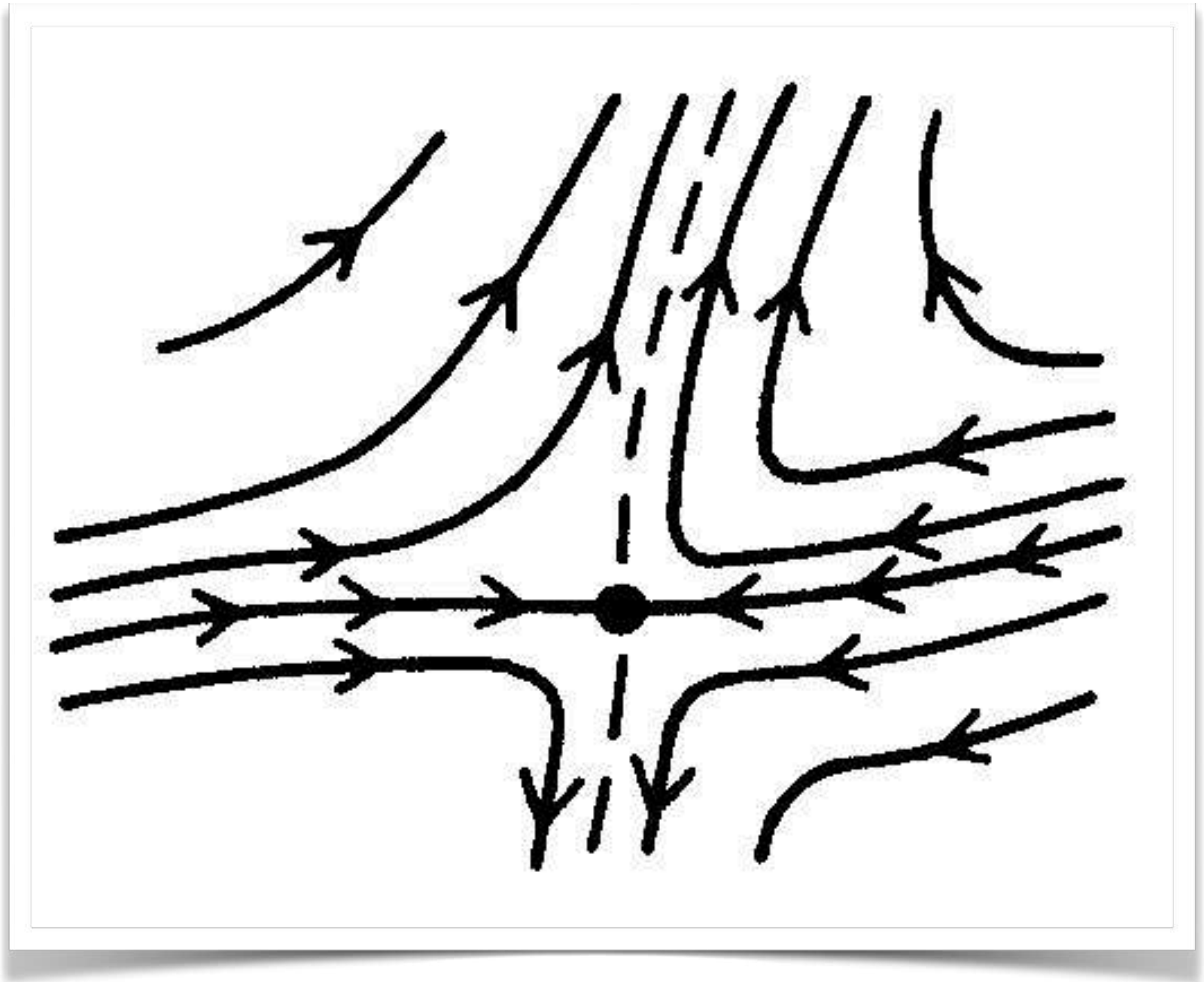


Real NVP, Dinh et al,1605.08803

Turns out to have surprising connection Störmer–Verlet integration

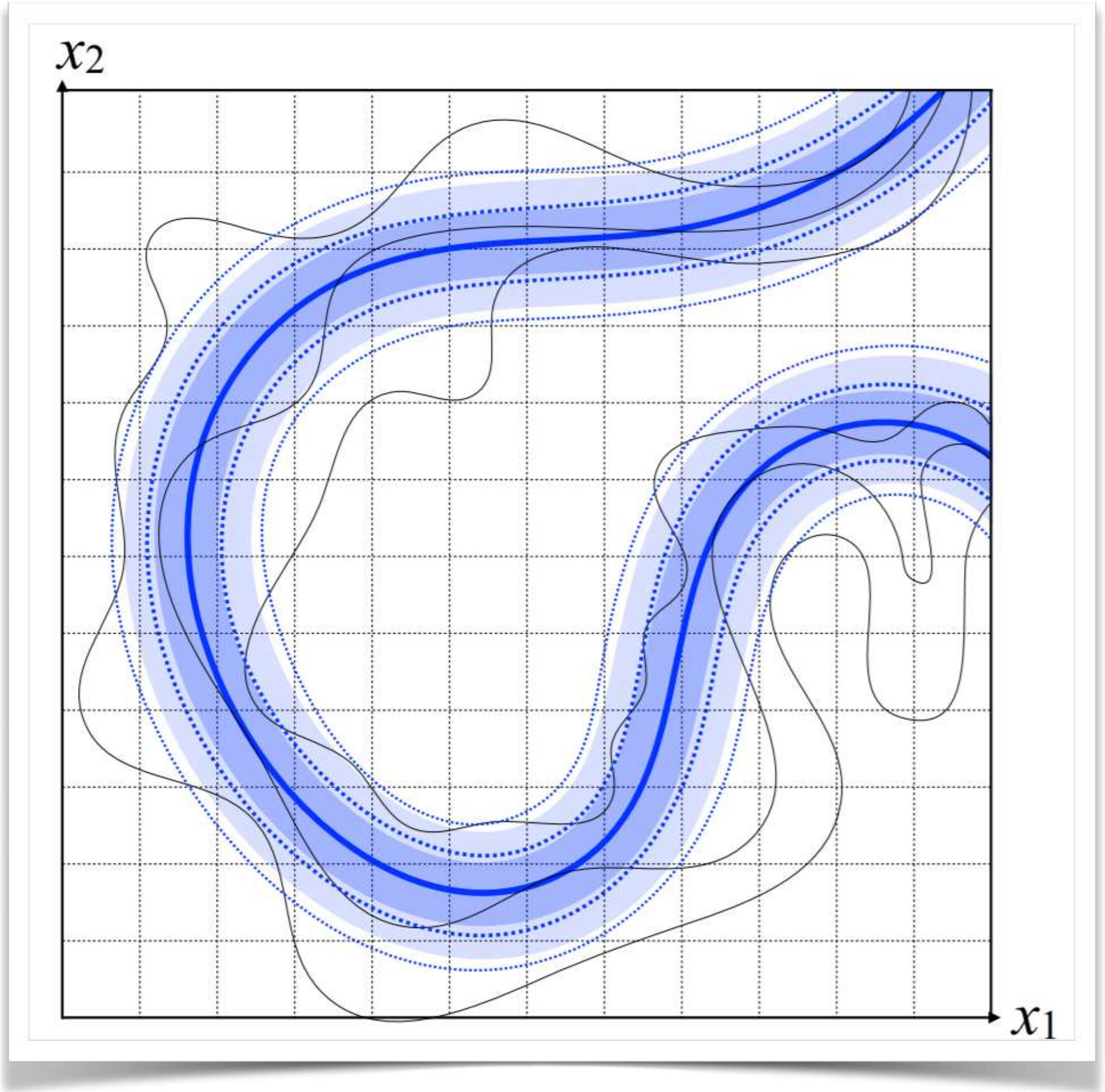
Why is flow useful for physics?

Renormalization group



Effective theory emerges upon transformation of the variables

Monte Carlo update



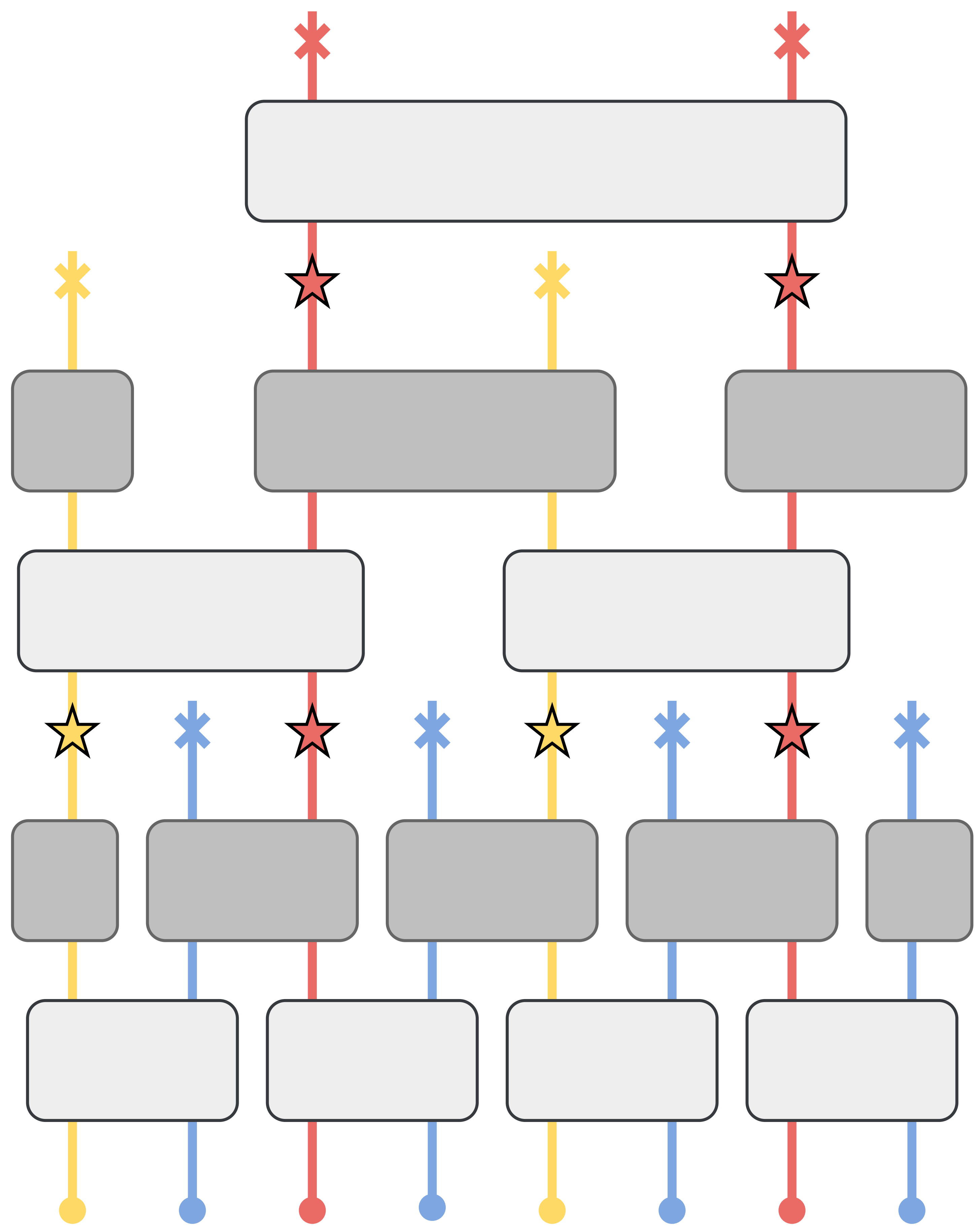
Physics happens on a manifold

Train neural nets to unfold that manifold

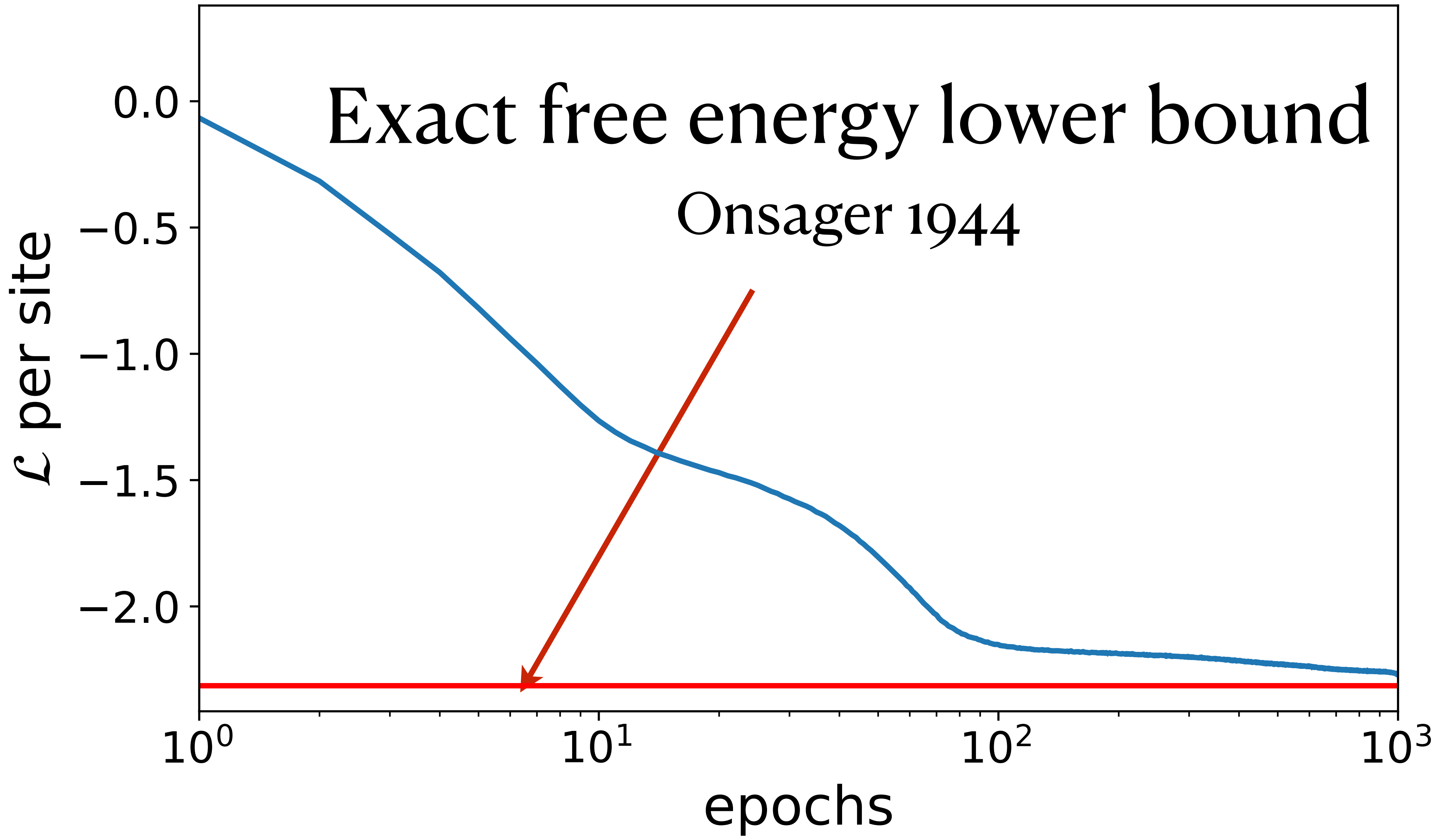
Neural network renormalization group

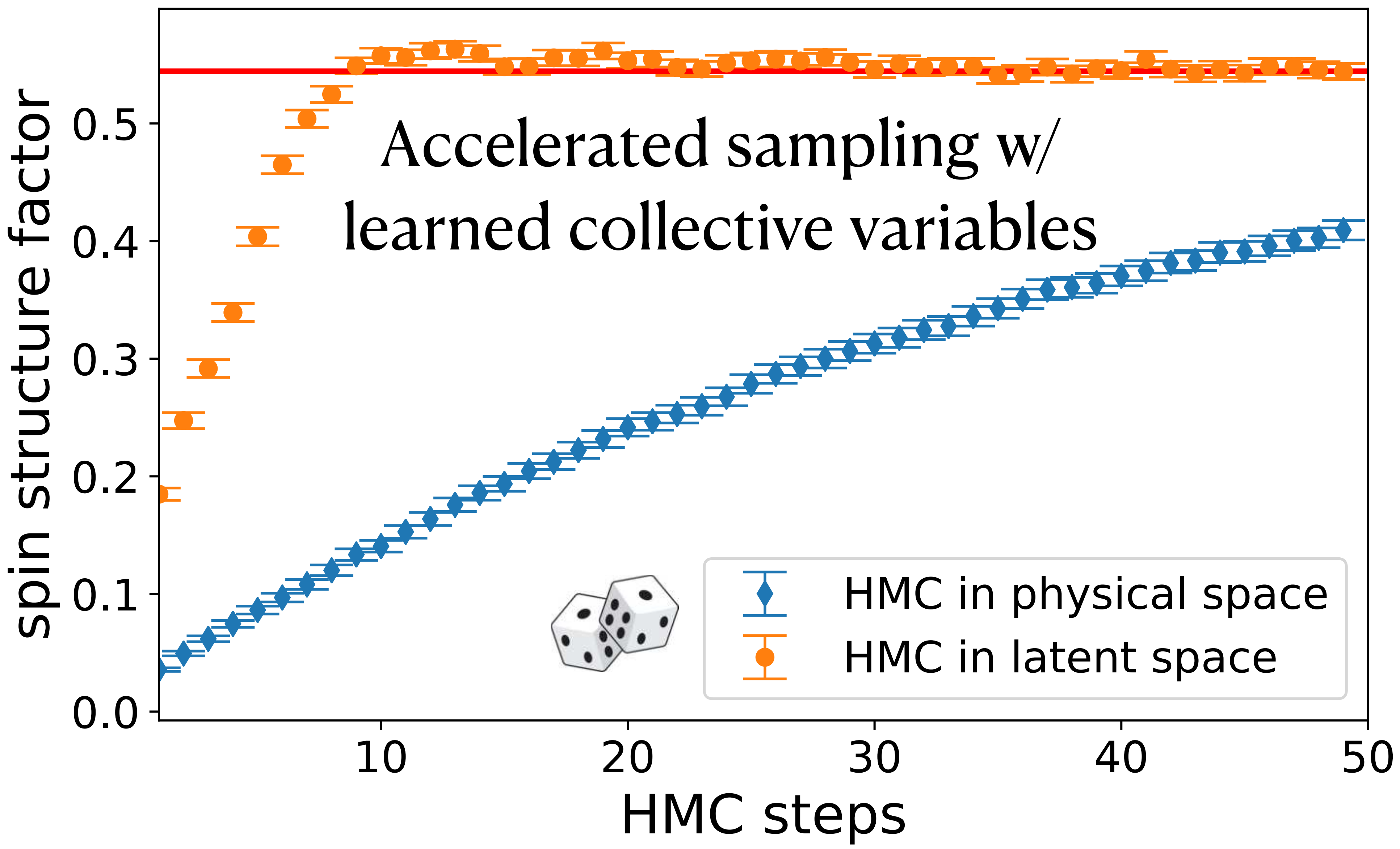
Li, LW, PRL ’18 [li012589/NeuralRG]

Collective variables

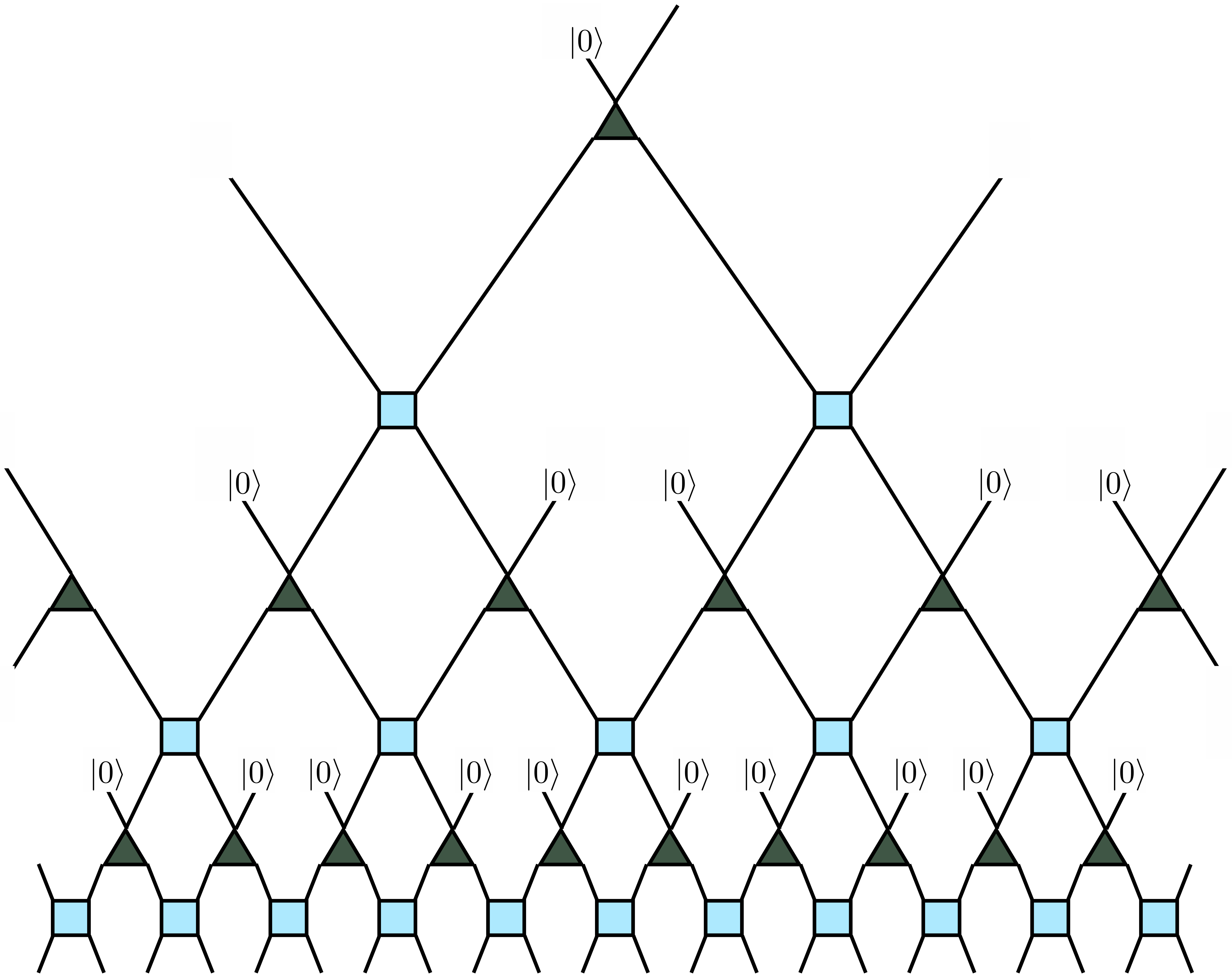


Physical variables

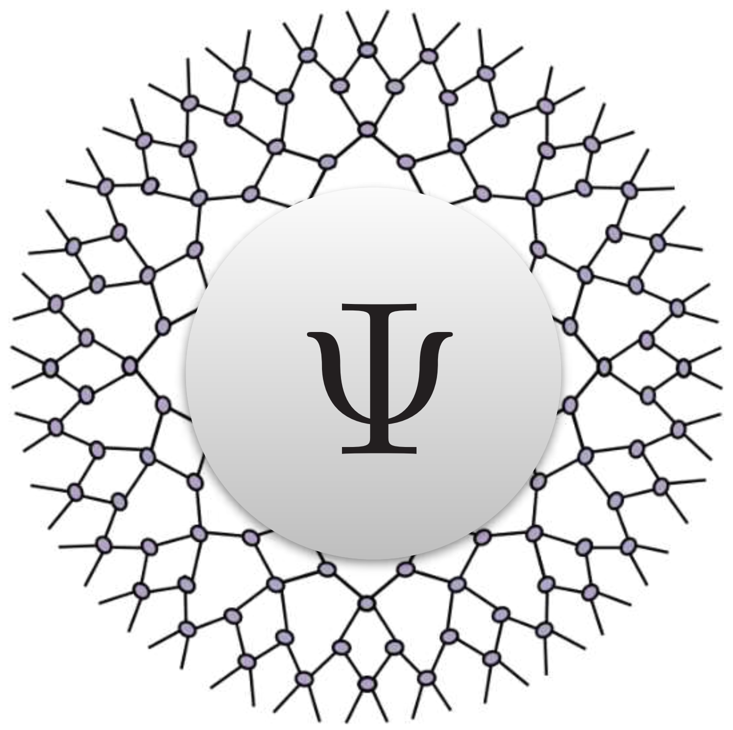




Quantum version of the architecture

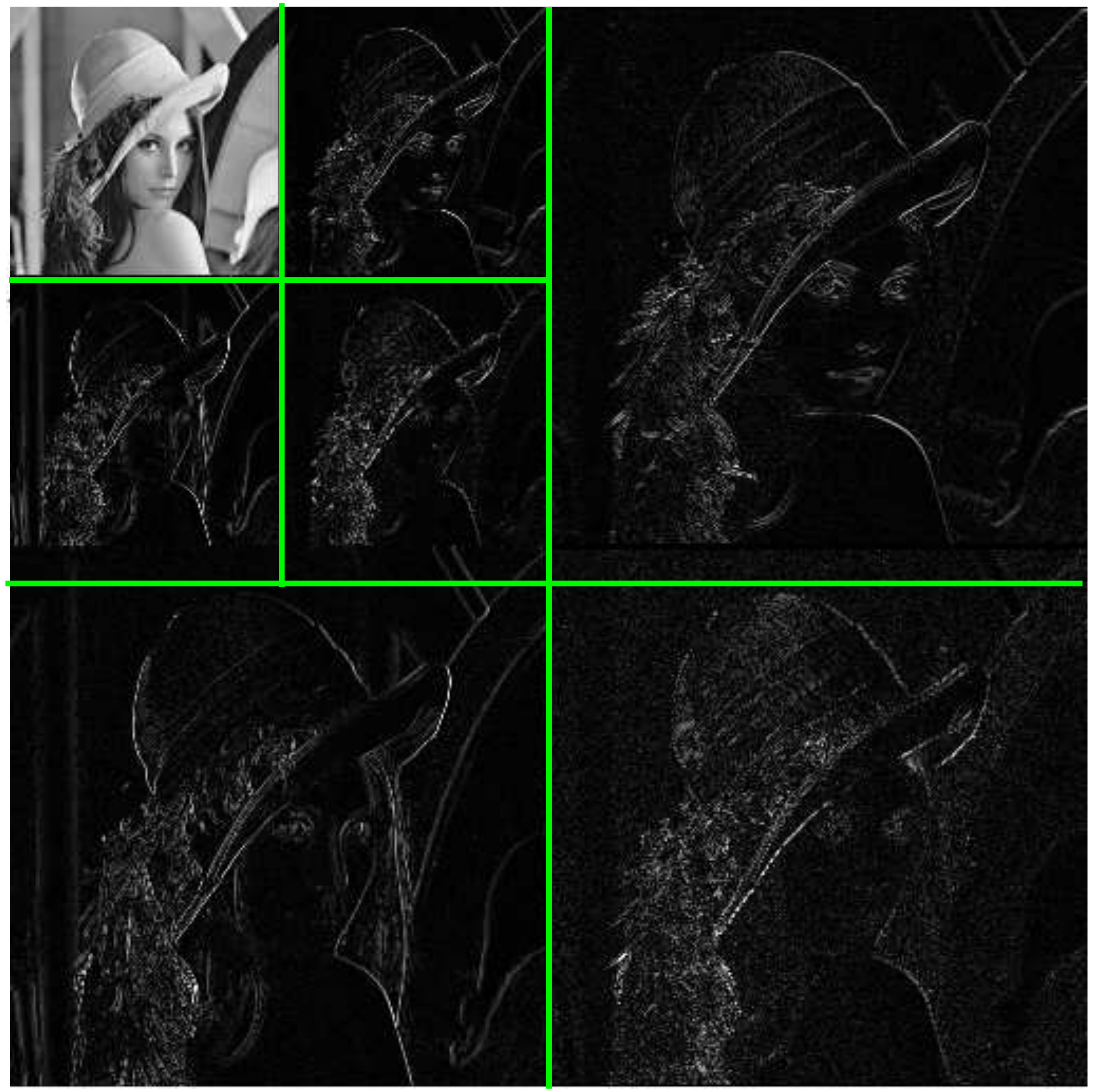


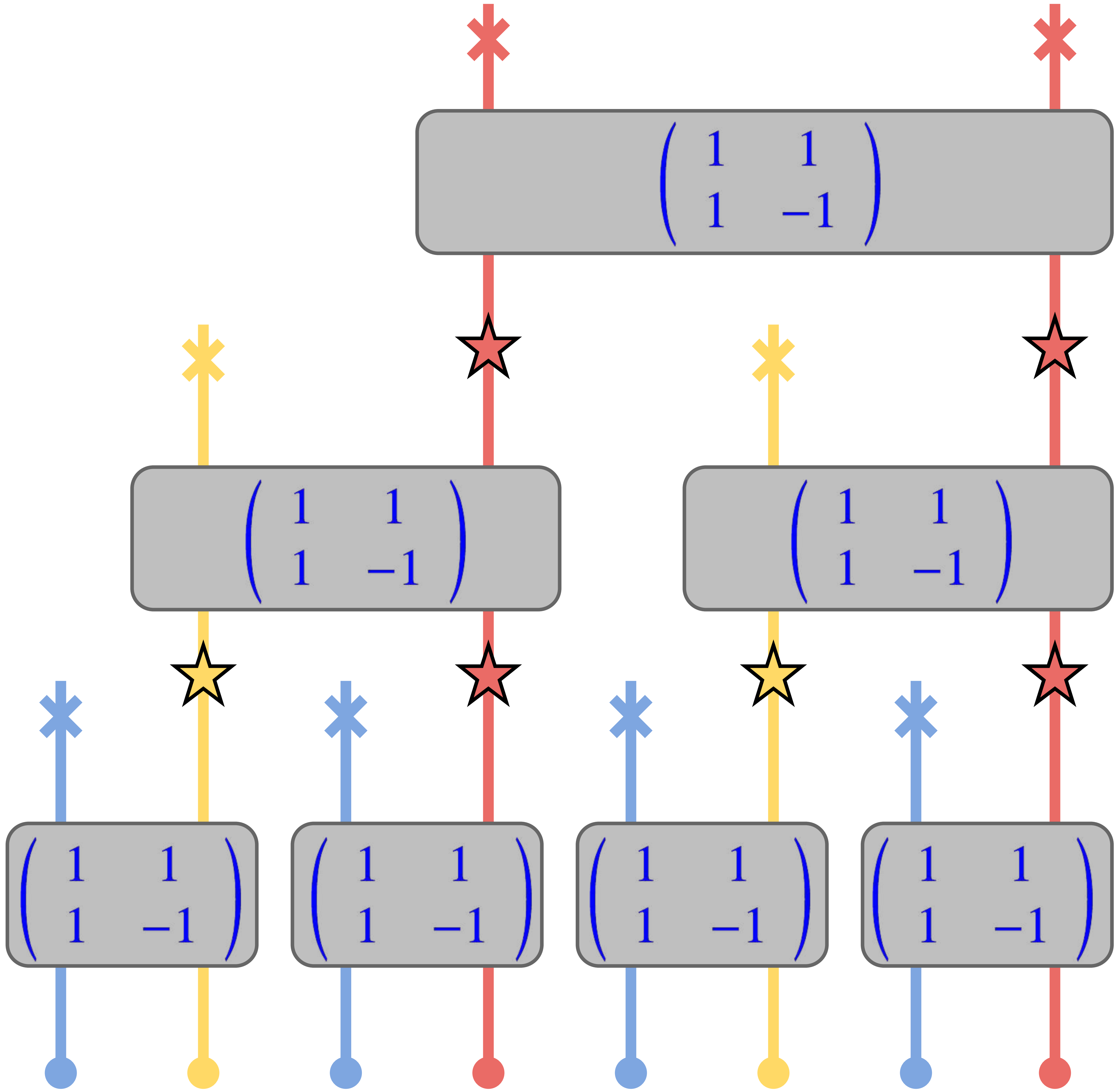
Entangled qubits



Multi-Scale
Entanglement
Renormalization
Ansatz

Connection to wavelets



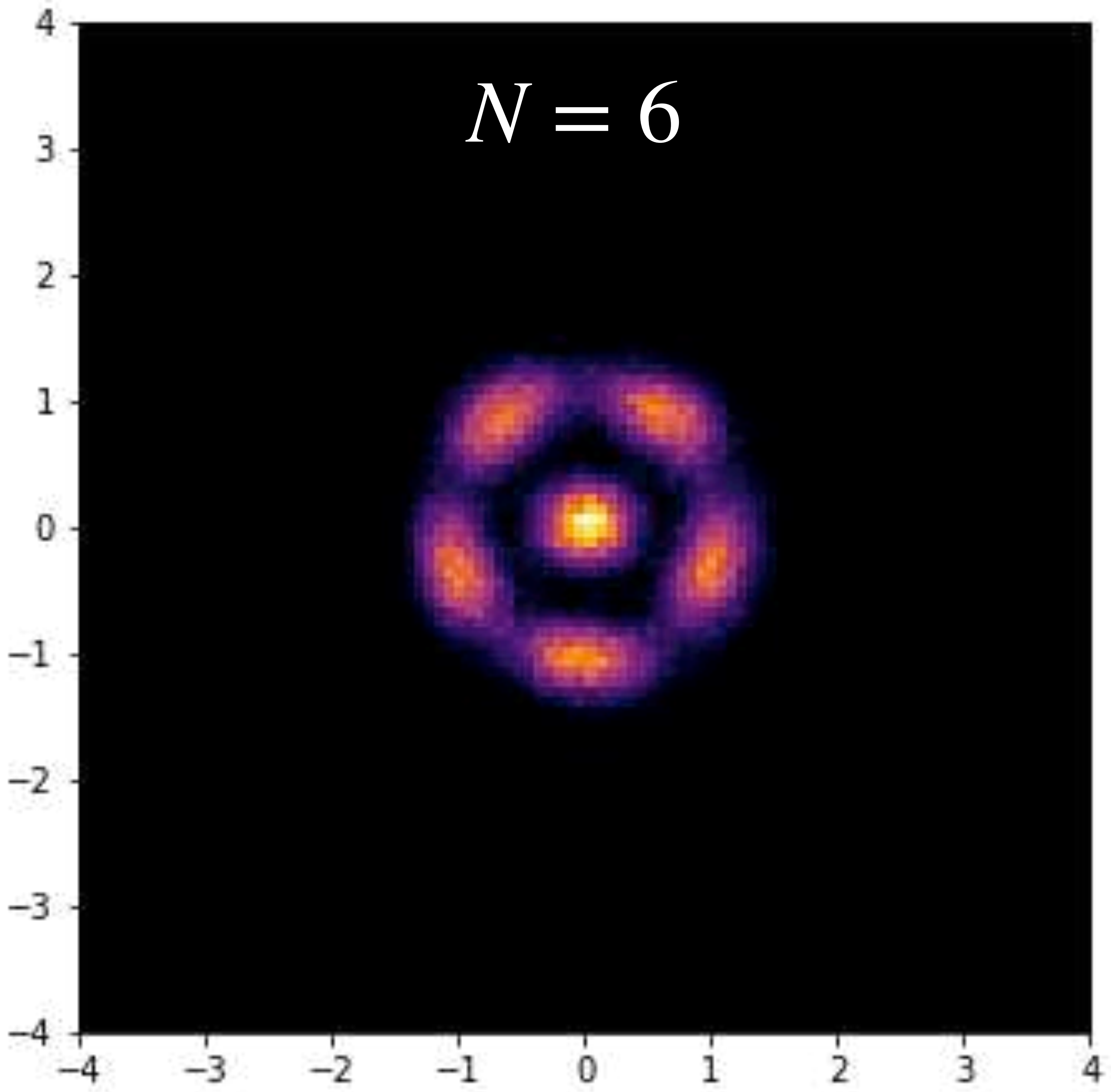


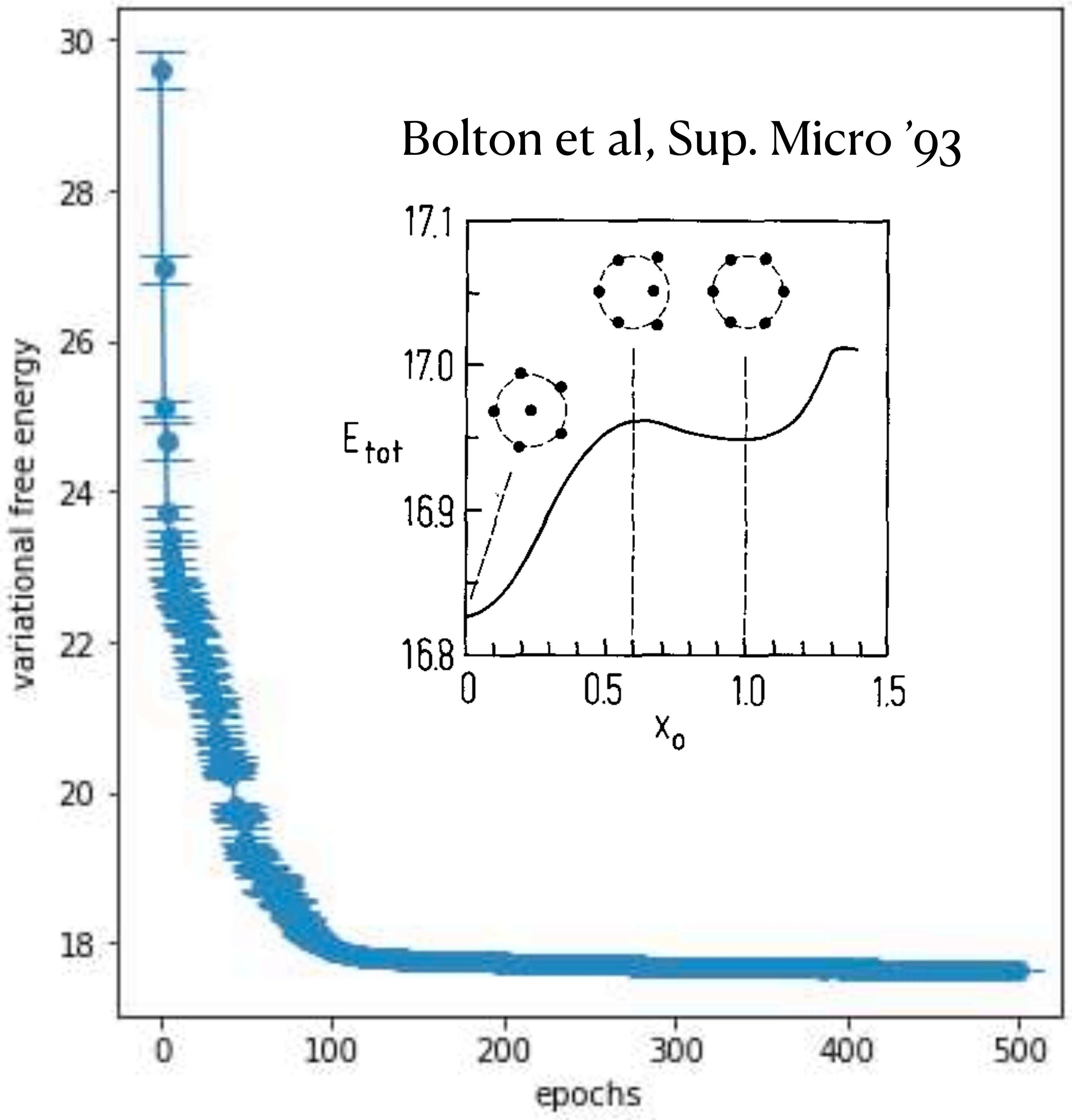
Nonlinear & adaptive generalizations of wavelets

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

Demo: Classical Coulomb gas in a harmonic trap

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





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

Continuous normalizing flows

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

Consider infinitesimal change-of-variables Chen et al 1806.07366

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

$\varepsilon \to 0$

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

Continuous normalizing flows

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

Consider infinitesimal change-of-variables Chen et al 1806.07366

$$\boldsymbol{X} = \boldsymbol{Z} + \varepsilon \boldsymbol{v}$$

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

$\varepsilon \to 0$

$t = 0$ $t = 1$

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

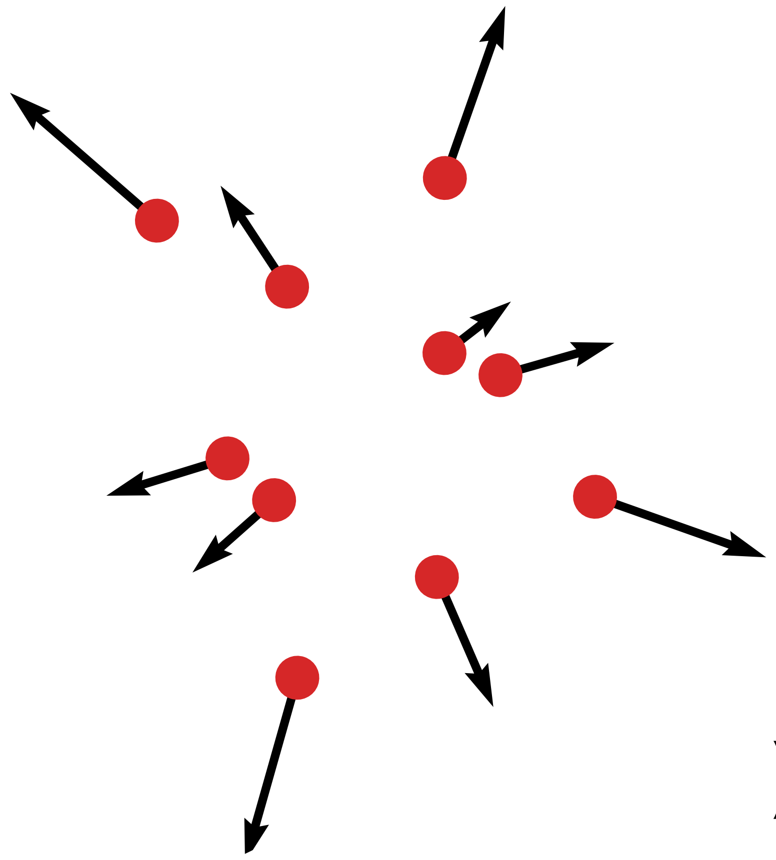
$$\frac{d \ln p(\boldsymbol{X}, t)}{dt} = -\nabla \cdot \boldsymbol{v}$$

Fluid physics behind flows

Zhang, E, LW 1809.10188
wangleiphy/MongeAmpereFlow

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

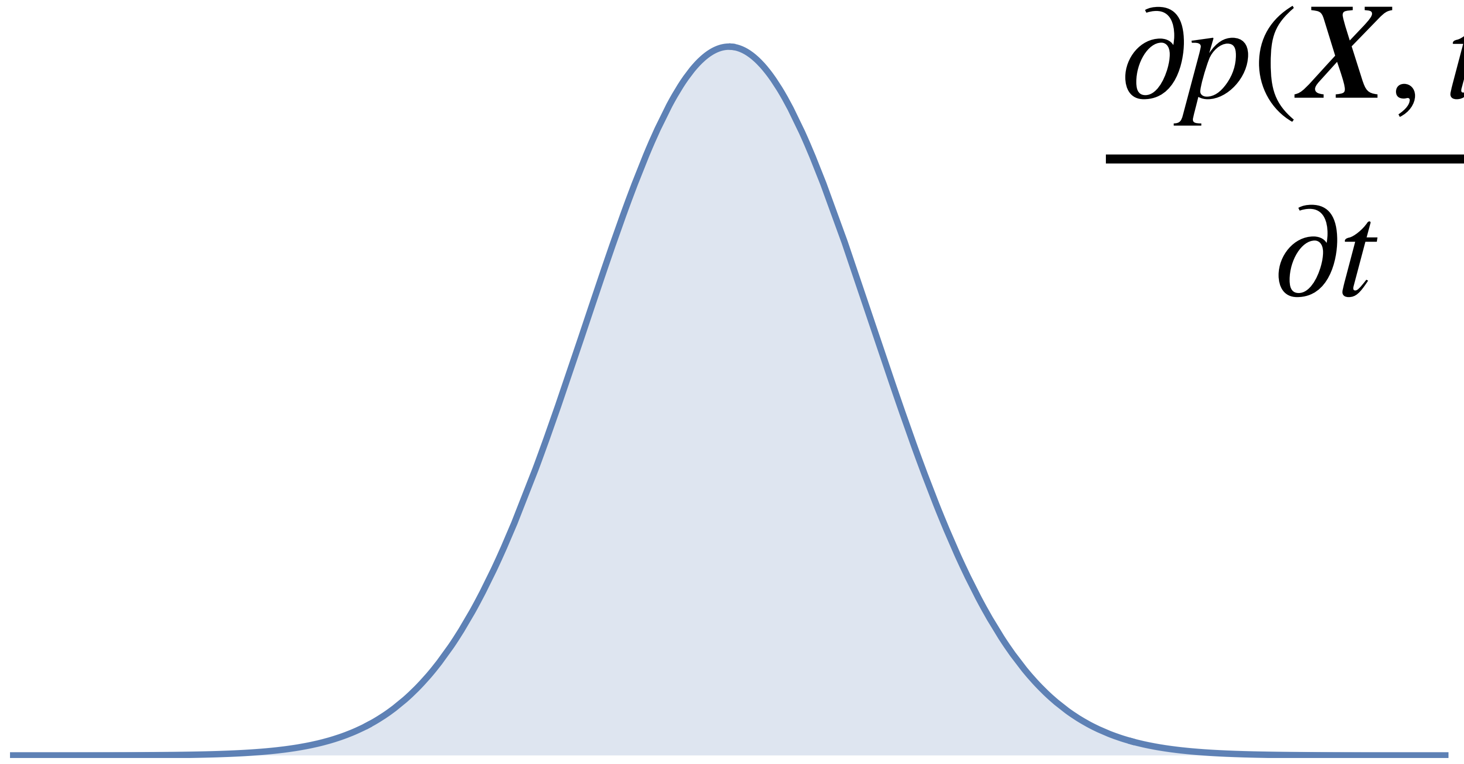
$$\frac{d \ln p(\boldsymbol{X}, t)}{dt} = - \nabla \cdot \boldsymbol{v}$$



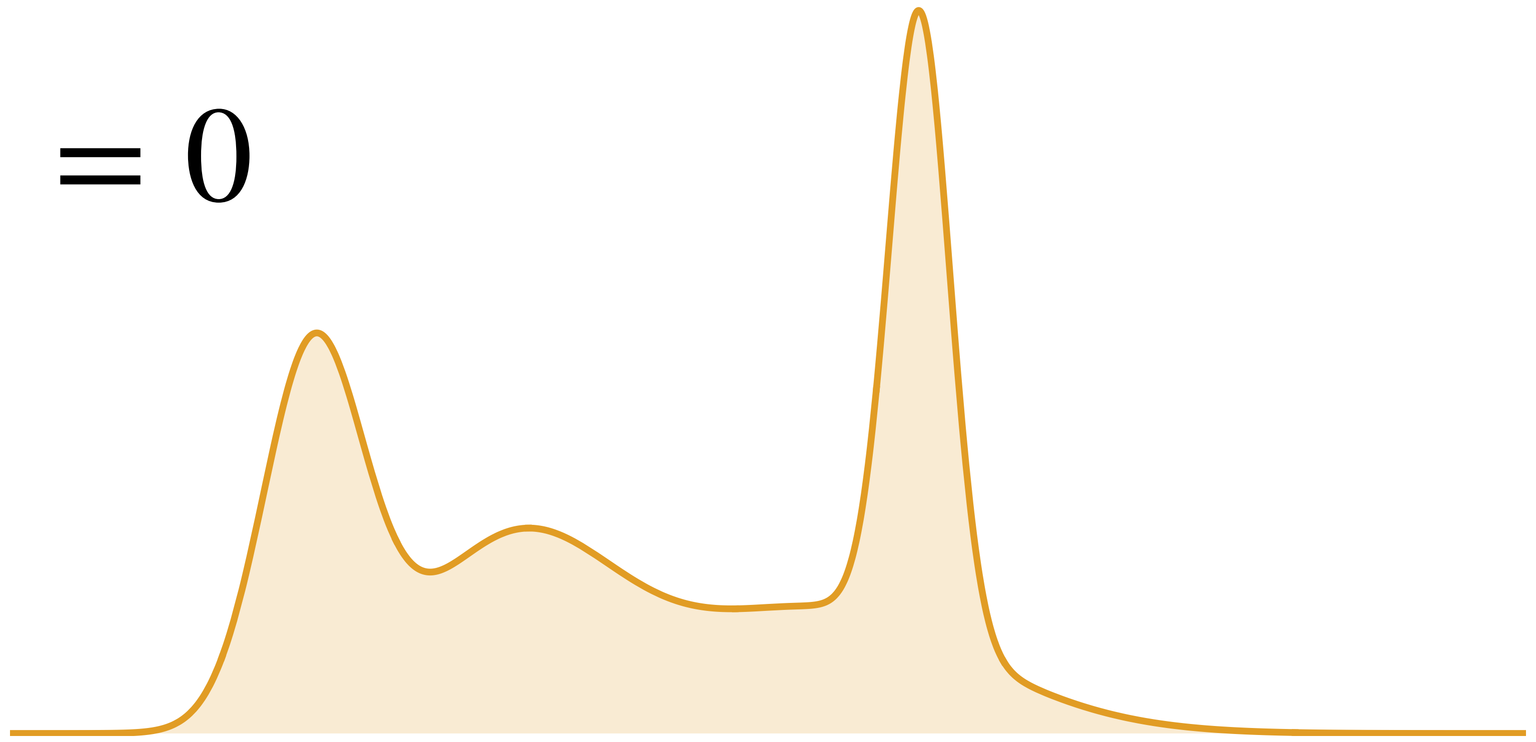
$$\frac{d}{dt} = \frac{\partial}{\partial t} + \boldsymbol{v} \cdot \nabla$$ “material derivative”

Lagrangian v.s. Euler approach to fluid mechanics

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



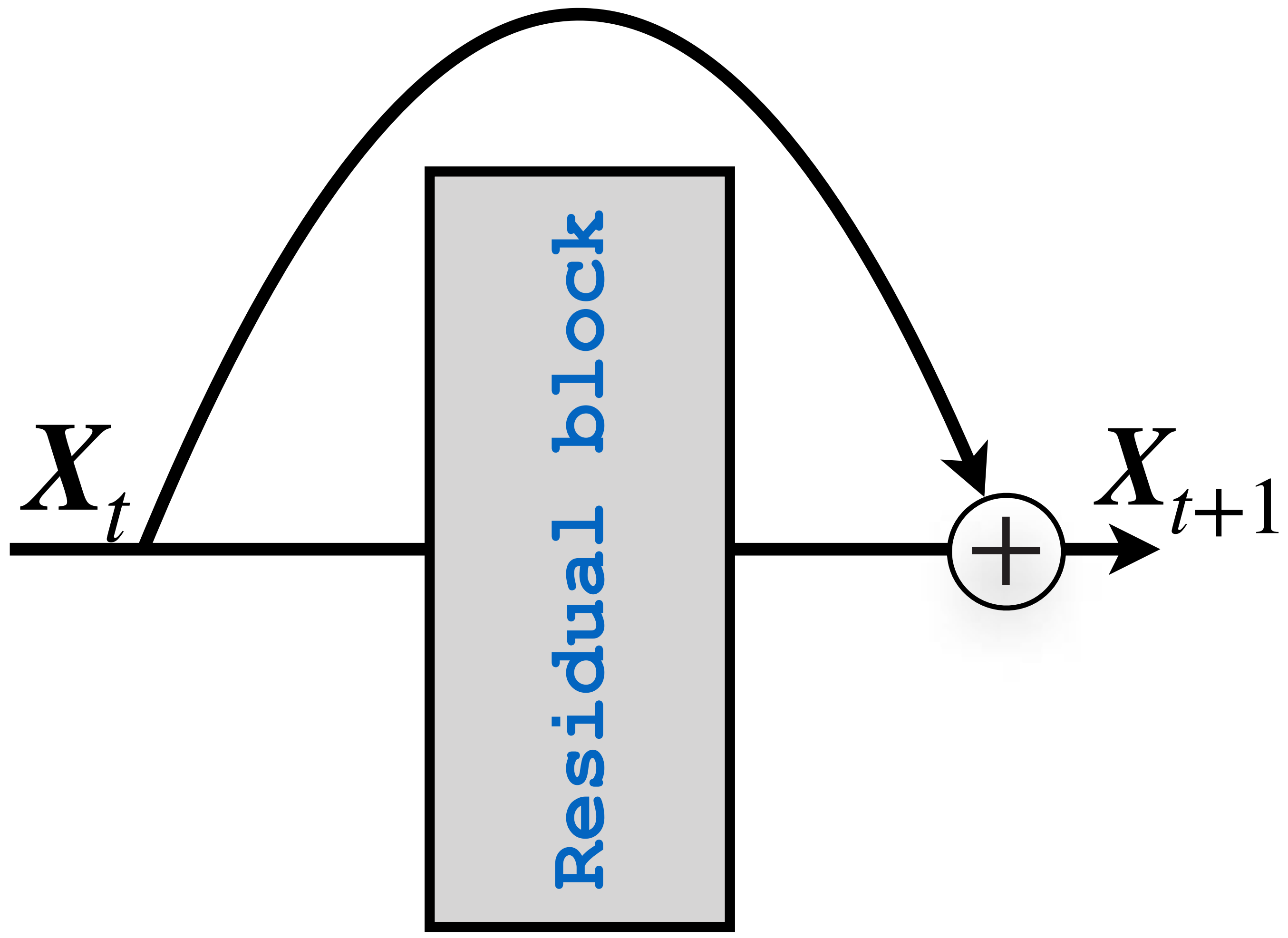
Simple density

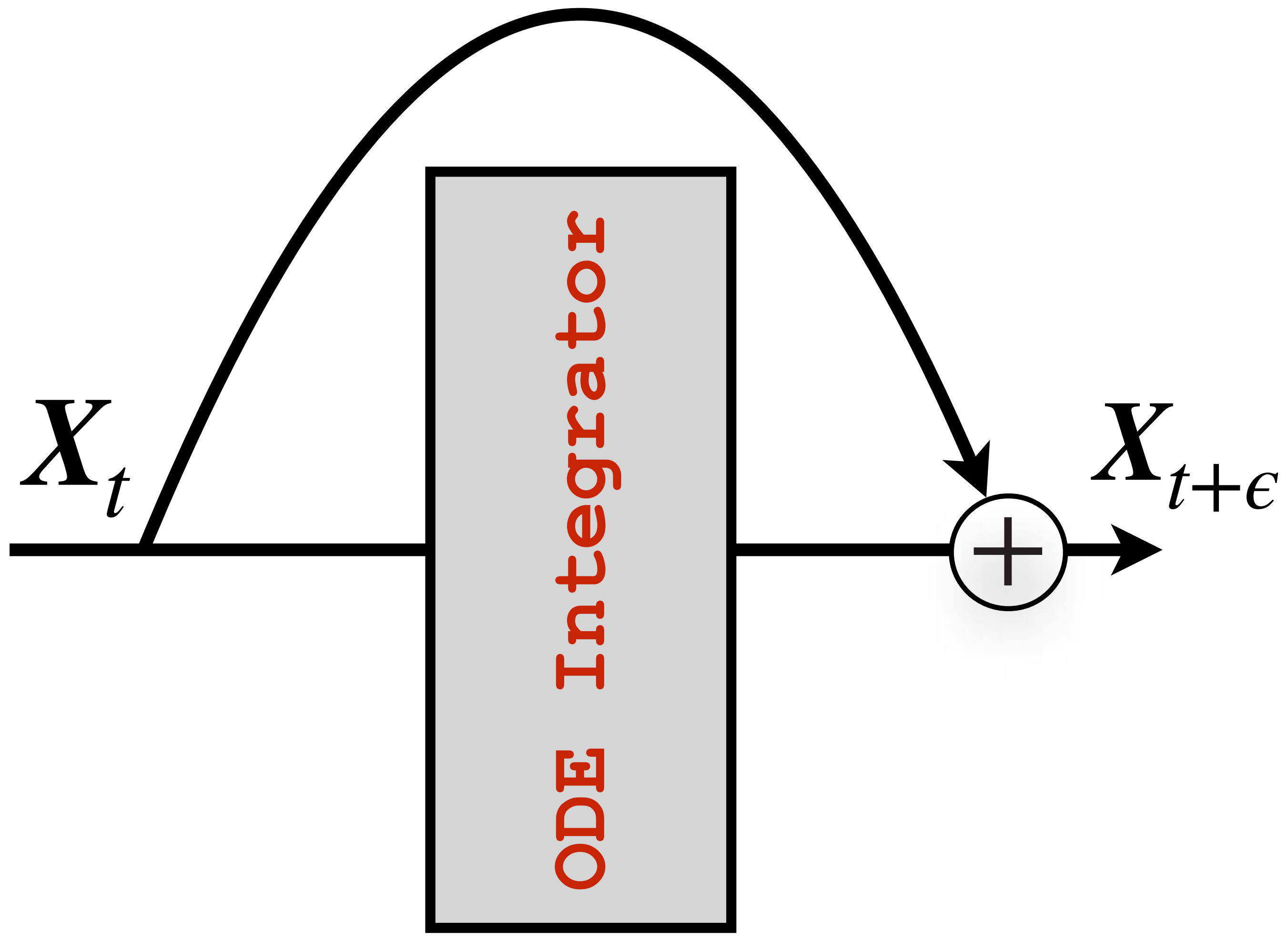


Complex density

Neural Ordinary Differential Equations

Residual network **ODE integration**





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

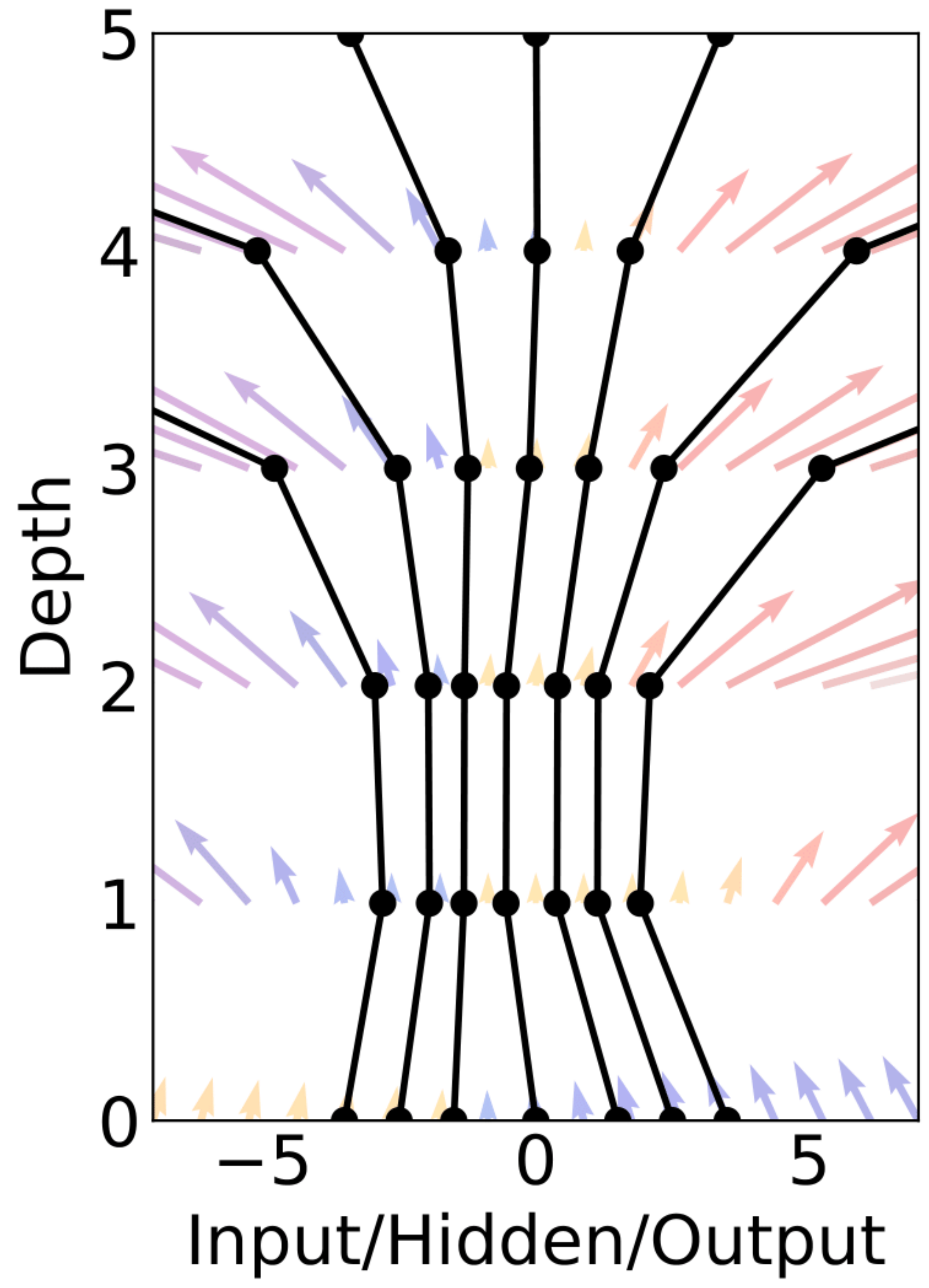
$$dX/dt = v(X)$$

Chen et al, 1806.07366

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

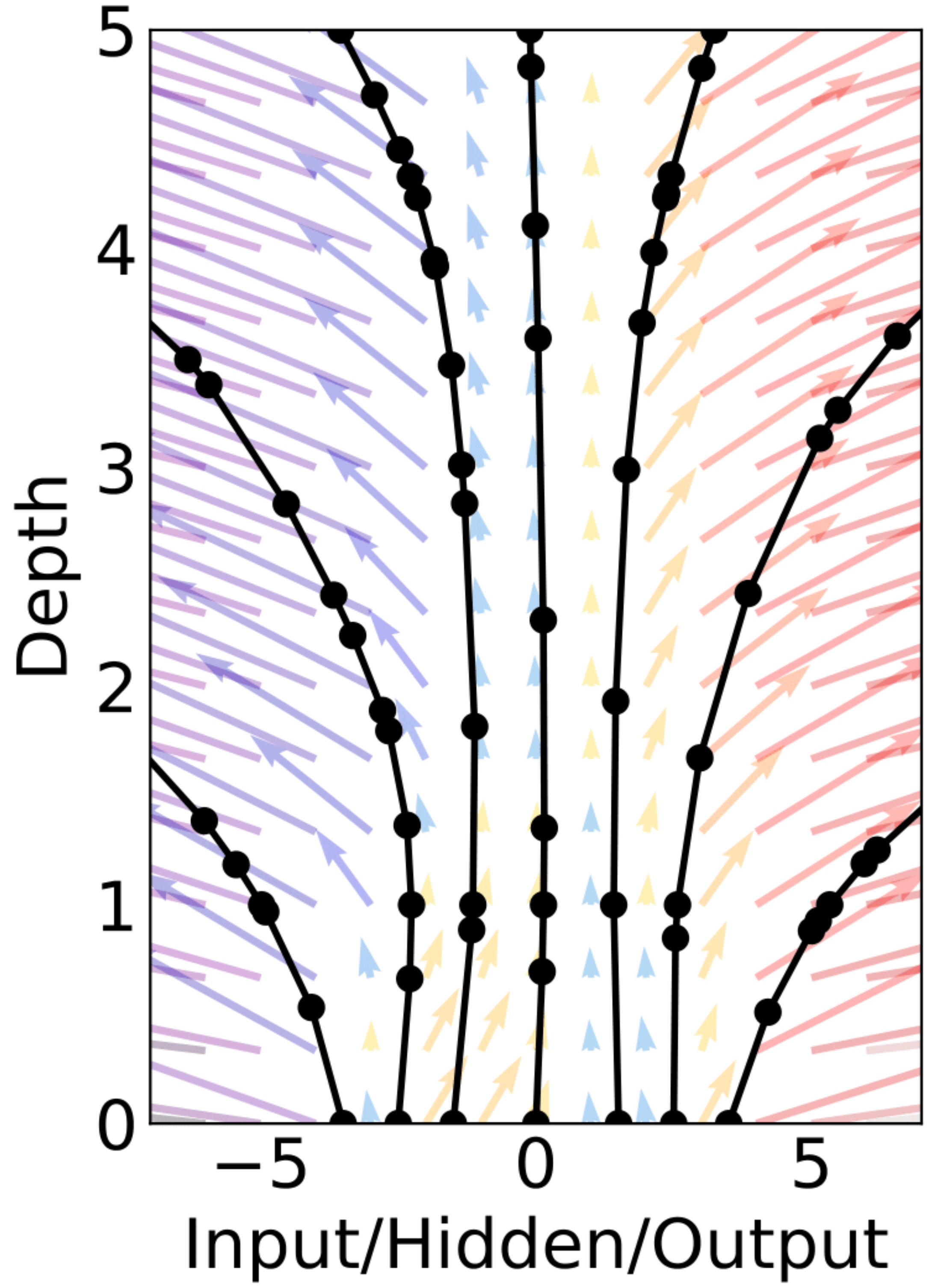
Neural Ordinary Differential Equations

Residual network



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

ODE integration



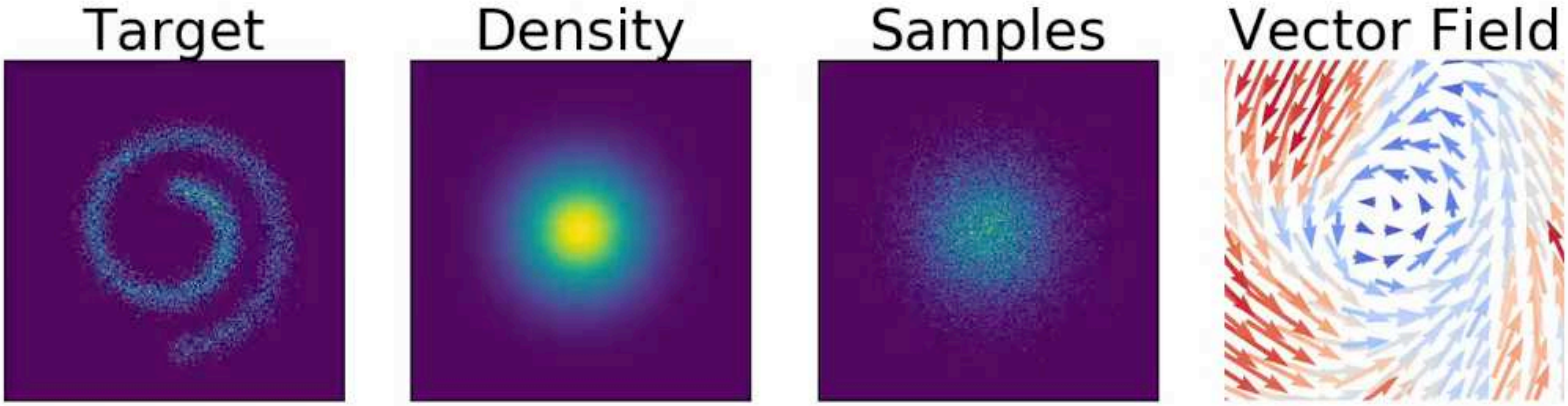
$$dX/dt = v(X)$$

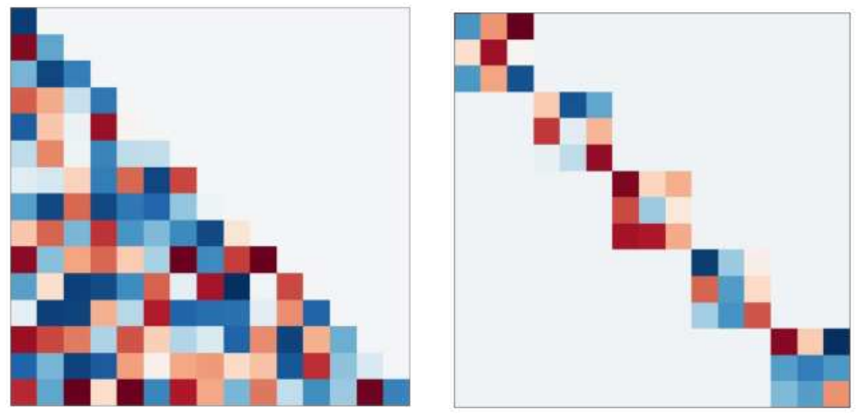
Harbor el al 1705.03341
Lu et al 1710.10121,
E Commun. Math. Stat 17'...

Chen et al, 1806.07366

Continuous normalizing flows implemented with NeuralODE

Chen et al, 1806.07366, Grathwohl et al 1810.01367

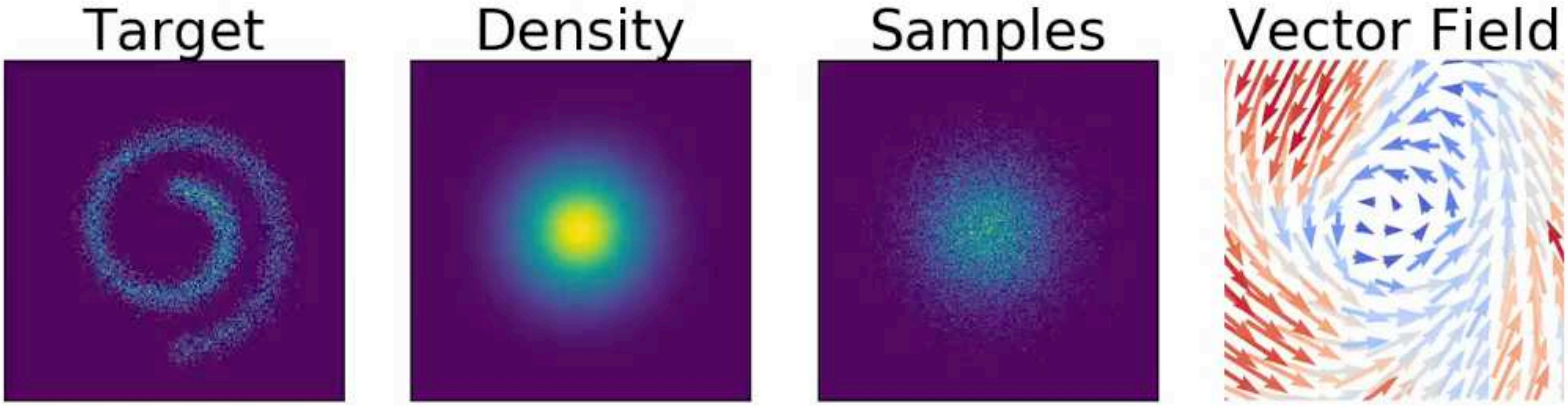


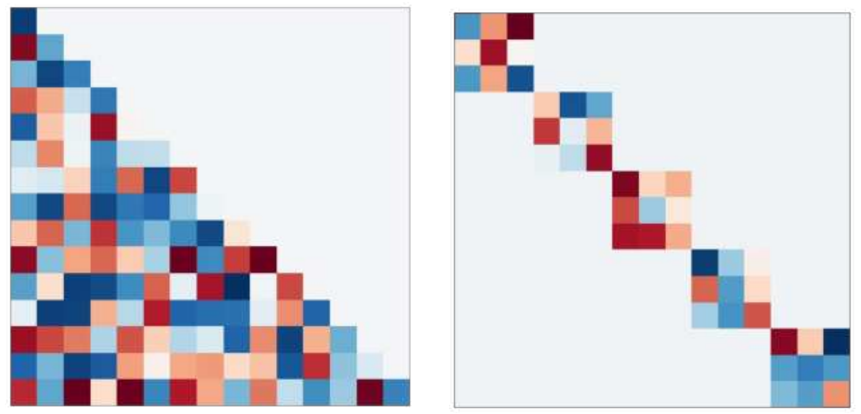


Continuous normalizing flow have no structural constraints on the transformation Jacobian

Continuous normalizing flows implemented with NeuralODE

Chen et al, 1806.07366, Grathwohl et al 1810.01367





Continuous normalizing flow have no structural constraints on the transformation Jacobian

The two use cases

Zhang, E, LW, 1809.10188

Maximum likelihood estimation

$$\mathcal{N}(\boldsymbol{x}) \xleftarrow{\int_T^0 \dot{\boldsymbol{x}}\, dt}$$

"learn from data"

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

Variational free energy

$$\mathcal{N}(\boldsymbol{x}) \xrightarrow{\int_0^T \dot{\boldsymbol{x}}\, dt} \frac{e^{-E(\boldsymbol{x})}}{Z}$$

"learn from Energy"

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

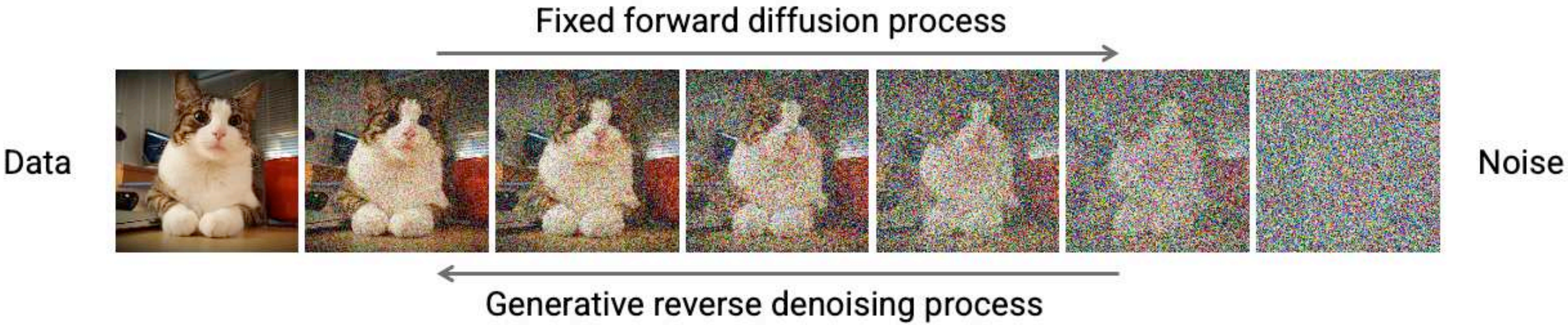
From flow to diffusion model

Continuity equation

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

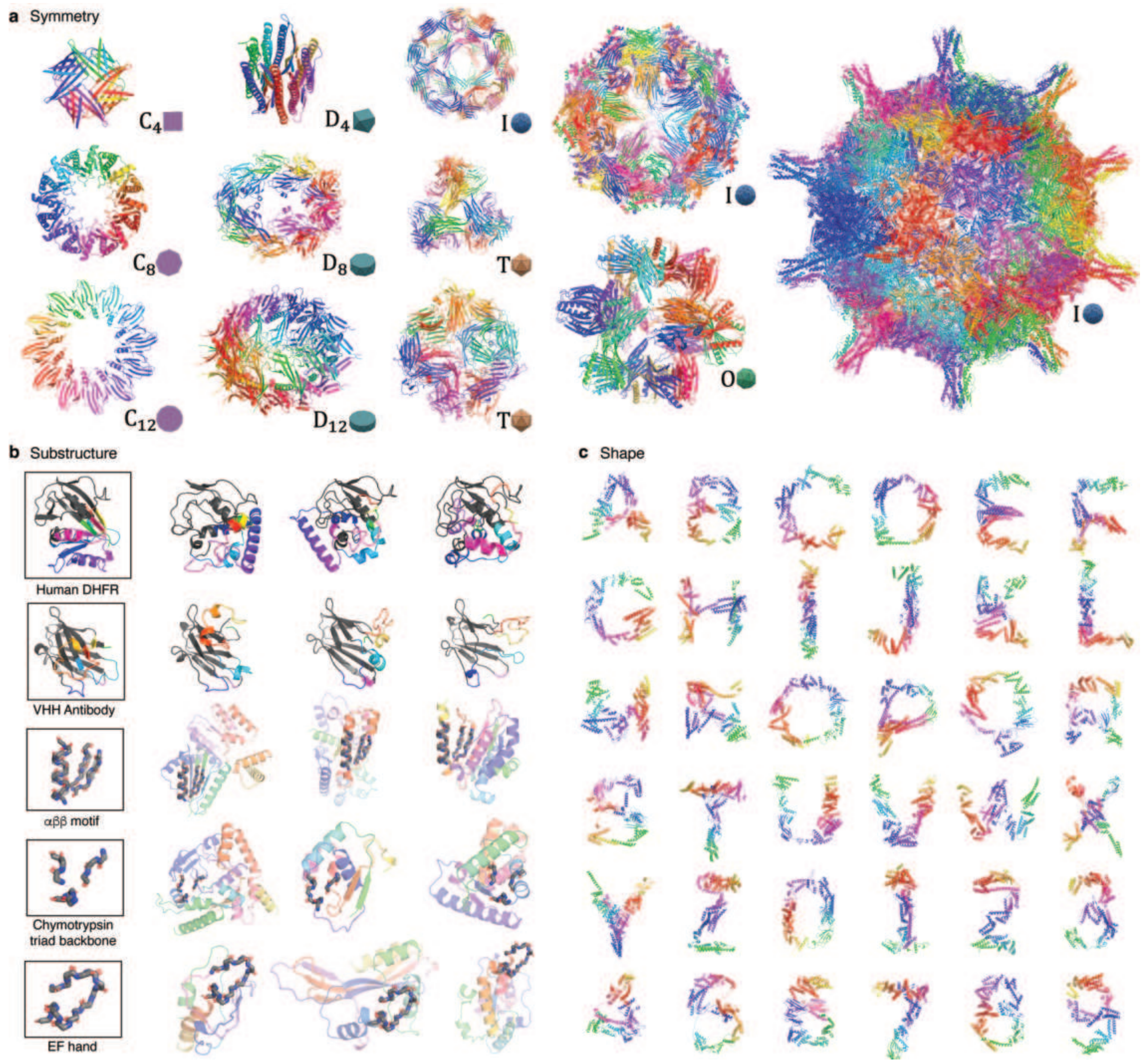
Fokker-Planck equation

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

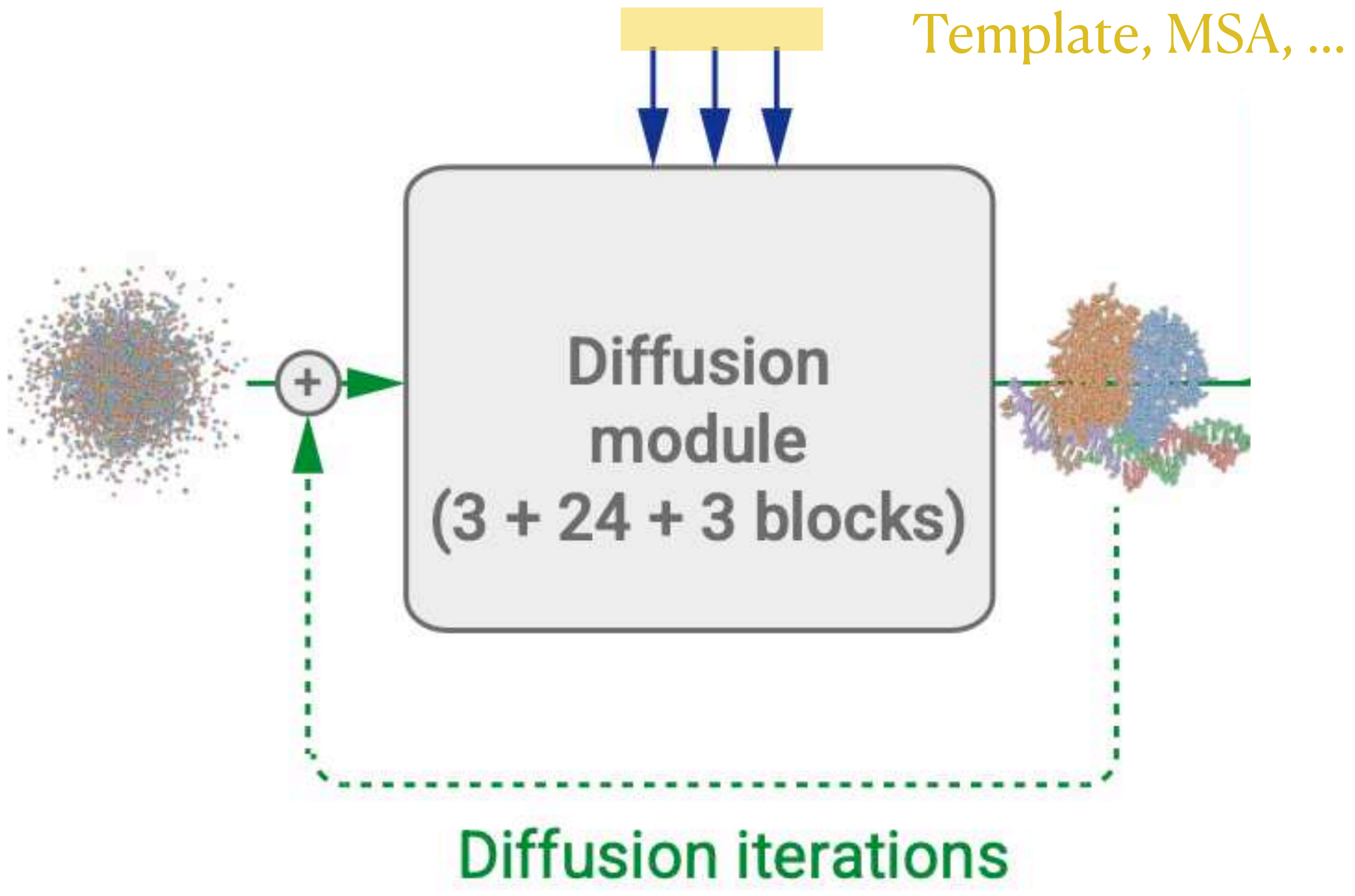


Vincent 2011, Sohl-Dickstein et al, 1503.03585, Song et al, 1907.05600 Ho et al, 2006.11239...

Diffusion models for protein structure prediction and design



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



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

Do they understand physics?

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

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

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

—Richard Feynman

Do they understand physics?

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

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

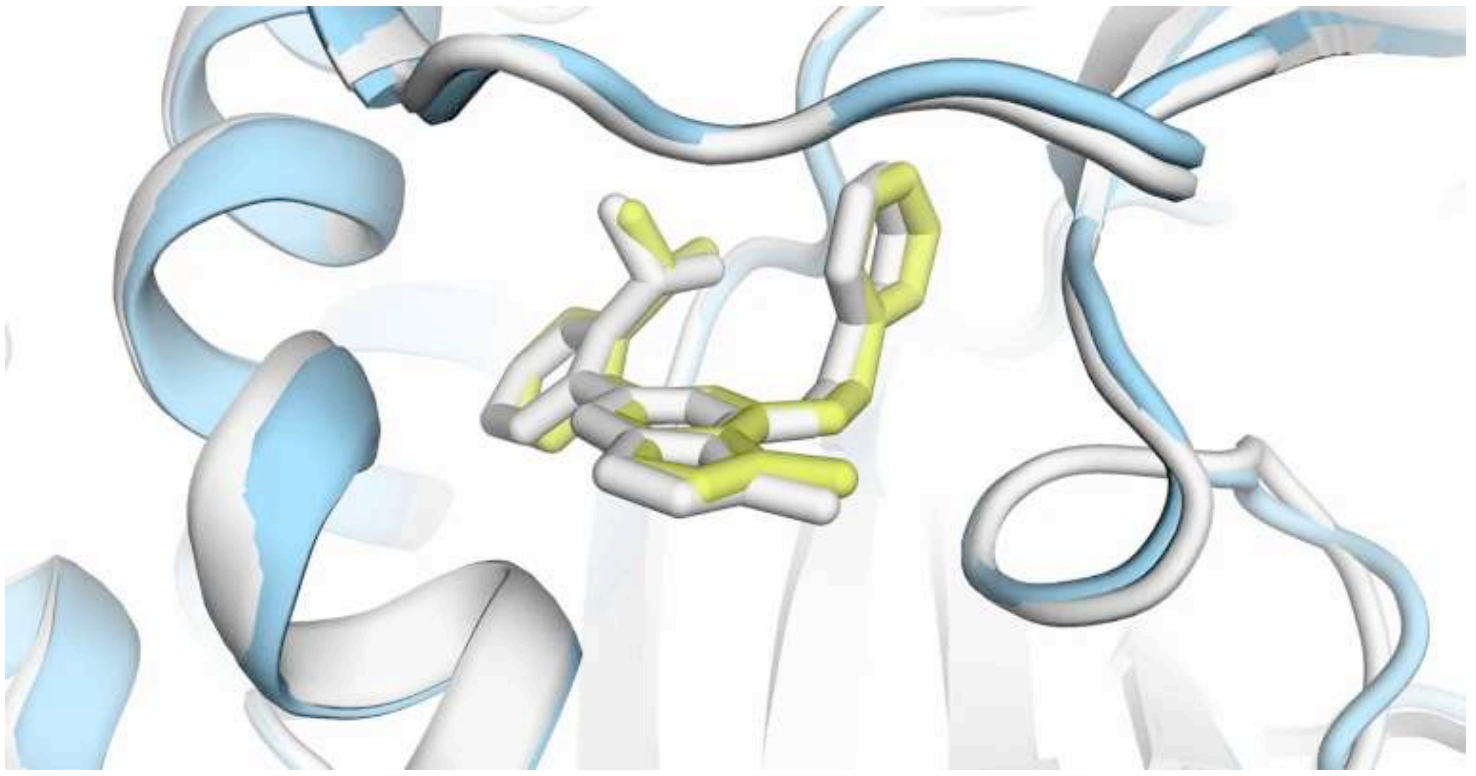
"What I can not create, I do not understand"

—Richard Feynman

Fold by intuition vs fold by equation

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

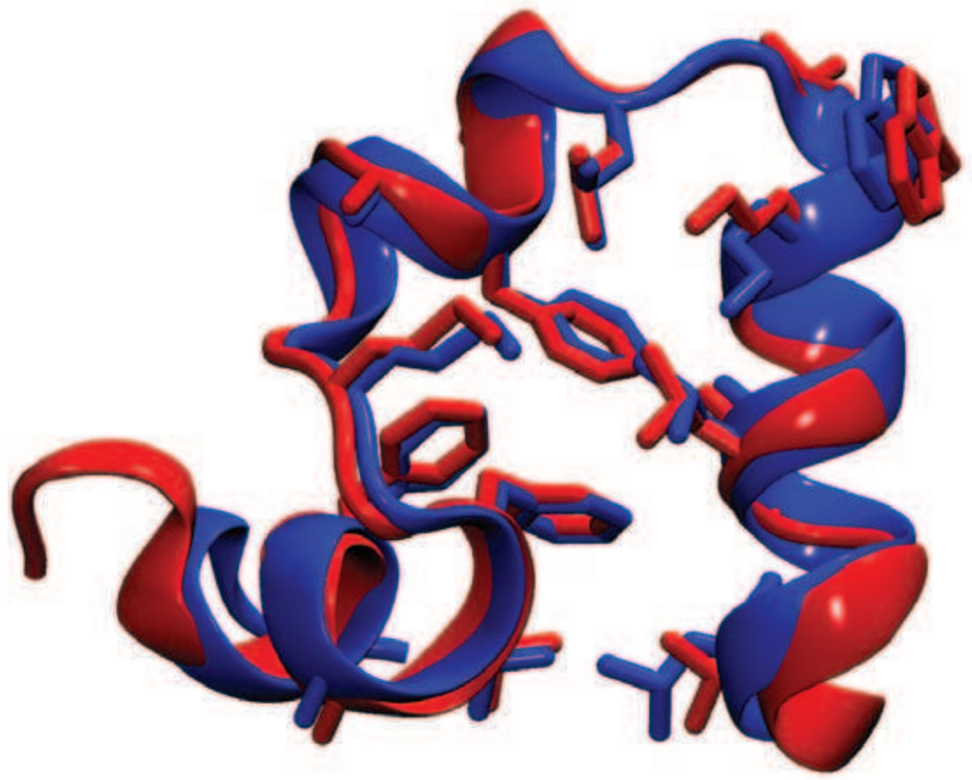
Data-driven generation AlphaFold3



Abramson et al, Nature 2024

The diffusion model may generate right conformations via unphysical pathways

Physics-based molecular dynamics



Shaw et al, Science 2010

Physical force fields may face difficulties in sampling rough energy landscapes

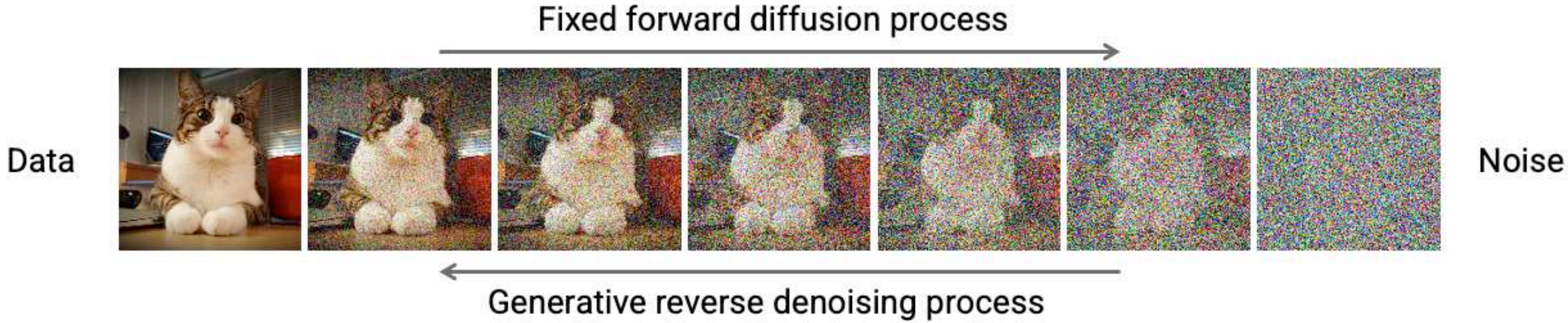
From flow to diffusion model, and back

Continuity equation

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

Fokker-Planck equation

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



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

A tale of three equations

Langevin equation (SDE)

$$\boldsymbol{x}_{t+dt} = \boldsymbol{x}_t + \boldsymbol{f}dt + \sqrt{2dt}\,\mathcal{N}(0,I)$$

Fokker-Planck equation (PDE)

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

"Particle method" (ODE)

$$\frac{d\boldsymbol{x}}{dt} = \boldsymbol{f} - \nabla \ln p(\boldsymbol{x},t) \equiv \boldsymbol{v}$$

(Another way to reverse the diffusion is via the reverse-time SDE Anderson 1982)

Maoutsa et al, 2006.00702
Song et al, 2011.13456

A tale of three equations

Langevin equation (SDE)

$$\boldsymbol{x}_{t+dt} = \boldsymbol{x}_t + \boldsymbol{f}dt + \sqrt{2dt}\,\mathcal{N}(0,I)$$

Fokker-Planck equation (PDE)

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

"Particle method" (ODE)

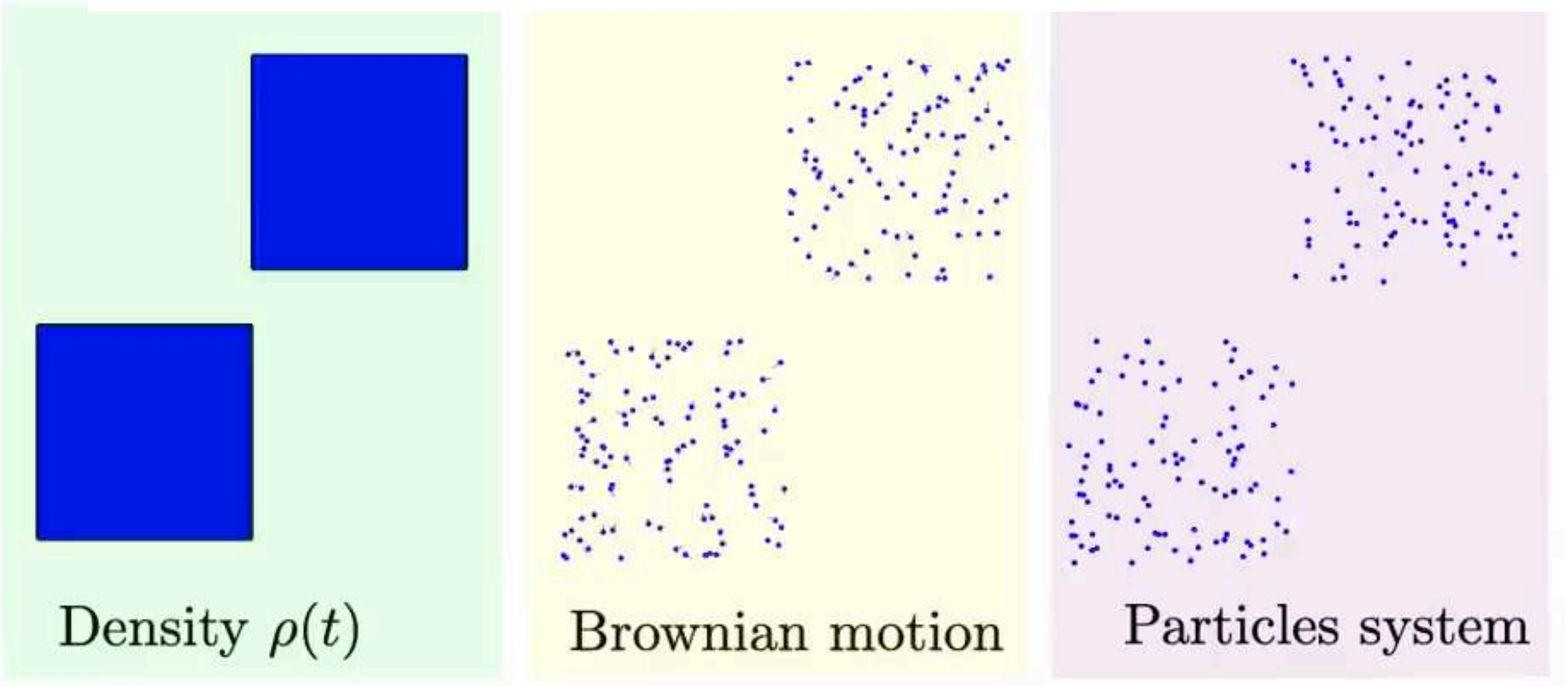
$$\frac{d\boldsymbol{x}}{dt} = \boldsymbol{f} - \nabla \ln p(\boldsymbol{x}, t) \equiv \boldsymbol{v}$$

(Another way to reverse the diffusion is via the reverse-time SDE Anderson 1982)

Maoutsa et al, 2006.00702
Song et al, 2011.13456

A tale of three equations

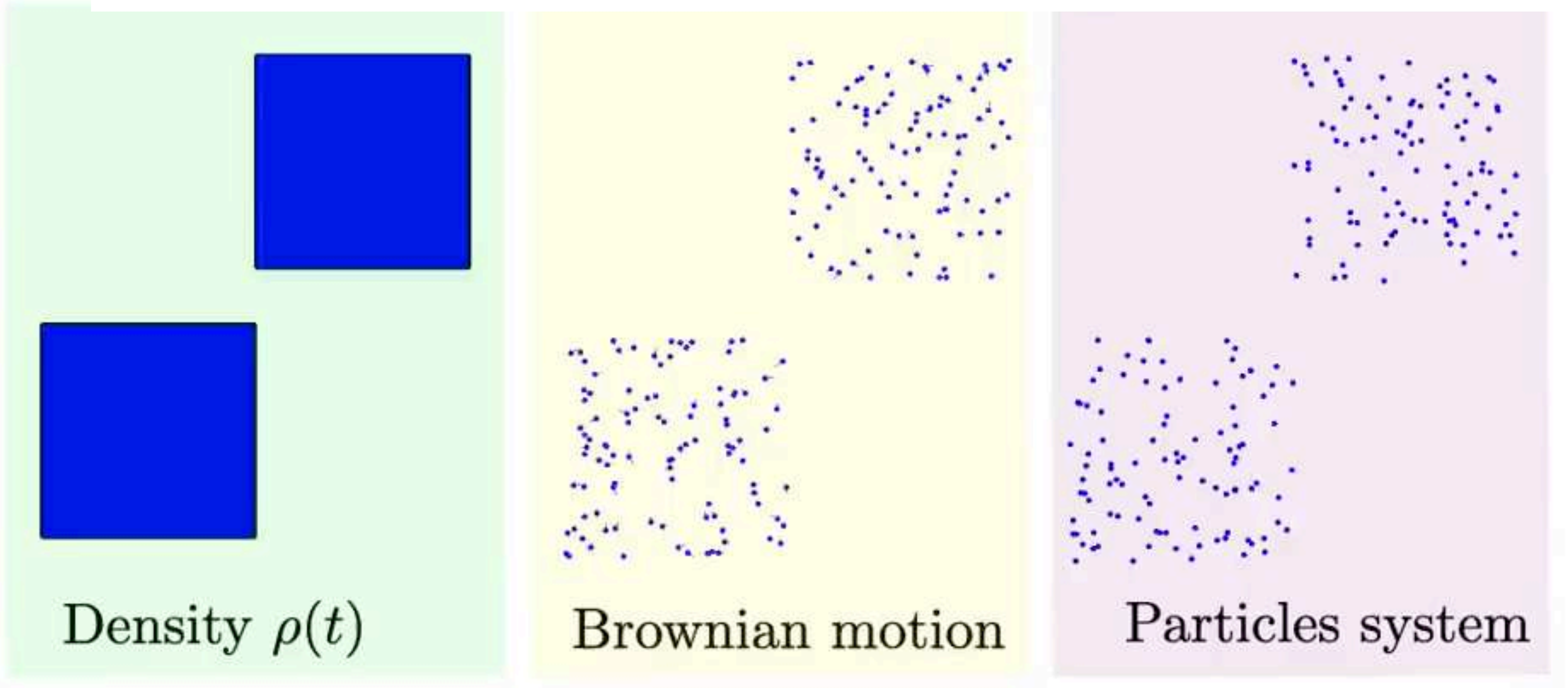
PDE	SDE	ODE
Fokker-Planck	Diffusion model	Flow model



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

A tale of three equations

PDE	SDE	ODE
Fokker-Planck	Diffusion model	Flow model



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

Lessons from diffusion models

Continuous normalizing flow still has GREAT potential!

Going beyond maximum likelihood estimation training (even if we can)

Break the loss into small pieces, sample them (kind of layer-wise training)

https://blog.alexalemi.com/diffusion.html

The conditional trick (originated from denoising score matching Vincent 2011)

$$\min_{\boldsymbol{\theta}} \underbrace{\mathbb{E}_{t\sim\mathcal{U}(0,T)}}_{\text{diffusion time } t} \underbrace{\mathbb{E}_{\mathbf{x}_0\sim q_0(\mathbf{x}_0)}}_{\text{data sample } \mathbf{x}_0} \underbrace{\mathbb{E}_{\mathbf{x}_t\sim q_t(\mathbf{x}_t|\mathbf{x}_0)}}_{\text{diffused data sample } \mathbf{x}_t} ||\underbrace{\mathbf{s}_{\boldsymbol{\theta}}(\mathbf{x}_t, t)}_{\text{neural network}} - \underbrace{\nabla_{\mathbf{x}_t} \log q_t(\mathbf{x}_t|\mathbf{x}_0)}_{\text{score of diffused data sample}}||_2^2$$

https://cvpr2022-tutorial-diffusion-models.github.io/

Lessons from diffusion models

Continuous normalizing flow still has GREAT potential!

Going beyond maximum likelihood estimation training (even if we can)

Break the loss into small pieces, sample them (kind of layer-wise training)

https://blog.alexalemi.com/diffusion.html

The conditional trick (originated from denoising score matching Vincent 2011)

$$\min_{\boldsymbol{\theta}} \underbrace{\mathbb{E}_{t \sim \mathcal{U}(0,T)}}_{\text{diffusion time } t} \underbrace{\mathbb{E}_{\mathbf{x}_t \sim q_t(\mathbf{x}_t)}}_{\text{diffused data } \mathbf{x}_t} ||\underbrace{\mathbf{s}_{\boldsymbol{\theta}}(\mathbf{x}_t, t)}_{\text{neural network}} - \underbrace{\nabla_{\mathbf{x}_t} \log q_t(\mathbf{x}_t)}_{\text{score of diffused data (marginal)}}||_2^2$$

https://cvpr2022-tutorial-diffusion-models.github.io/

Lessons from diffusion models

Continuous normalizing flow still has GREAT potential!

Going beyond maximum likelihood estimation training (even if we can)

Break the loss into small pieces, sample them (kind of layer-wise training)

https://blog.alexalemi.com/diffusion.html

The conditional trick (originated from denoising score matching Vincent 2011)

$$\mathbb{E}_{x\sim q(x)} |s(x) - \nabla_x \ln q(x)|^2 = \mathbb{E}_{x\sim q(x|x_0)}\mathbb{E}_{x_0\sim q_0(x_0)} |s(x) - \nabla_x \ln q(x|x_0)|^2 + \text{const}\,.$$

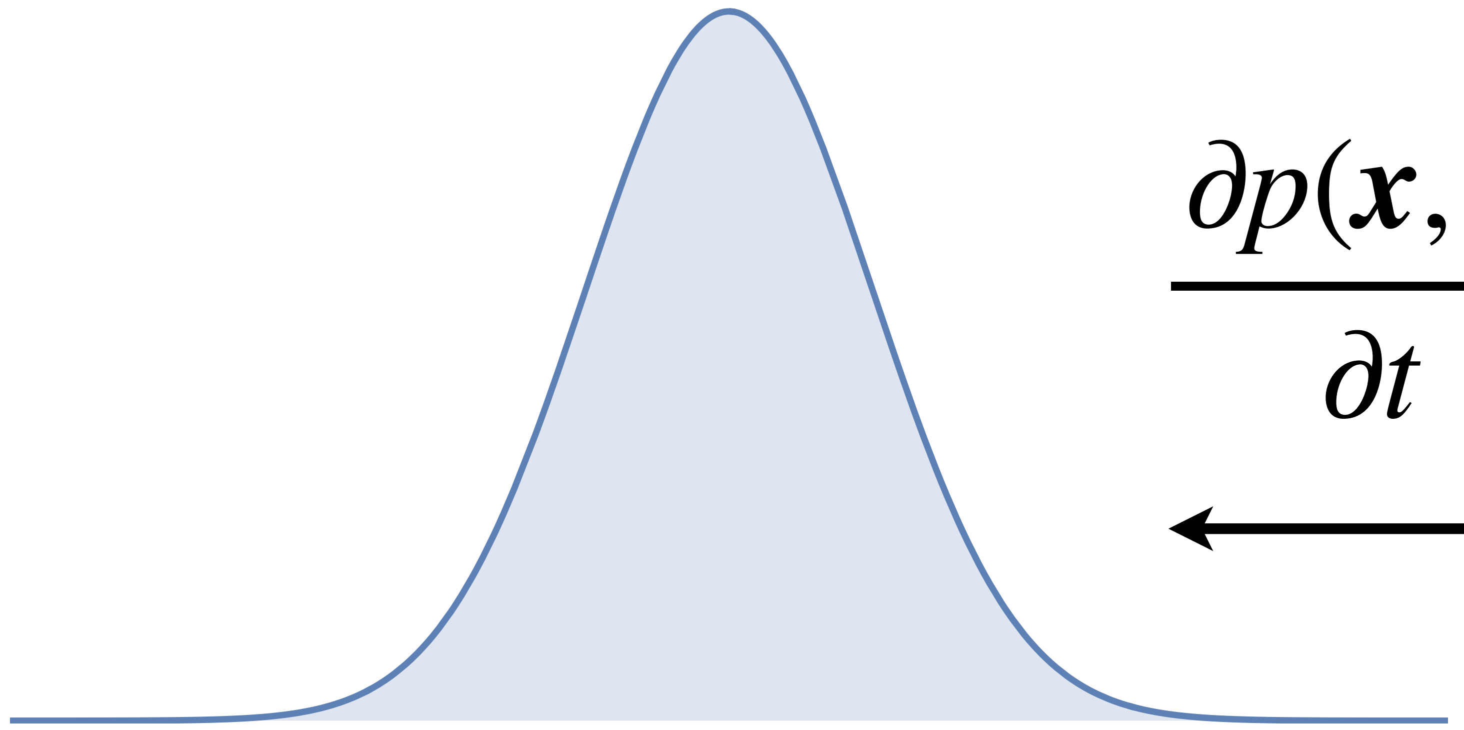
$$q(x) = \int q(x|x_0) q_0(x_0) dx_0$$

Vincent 2011

Flow matching

$p(\boldsymbol{x},1) = \mathcal{N}(0,I)$

base distribution



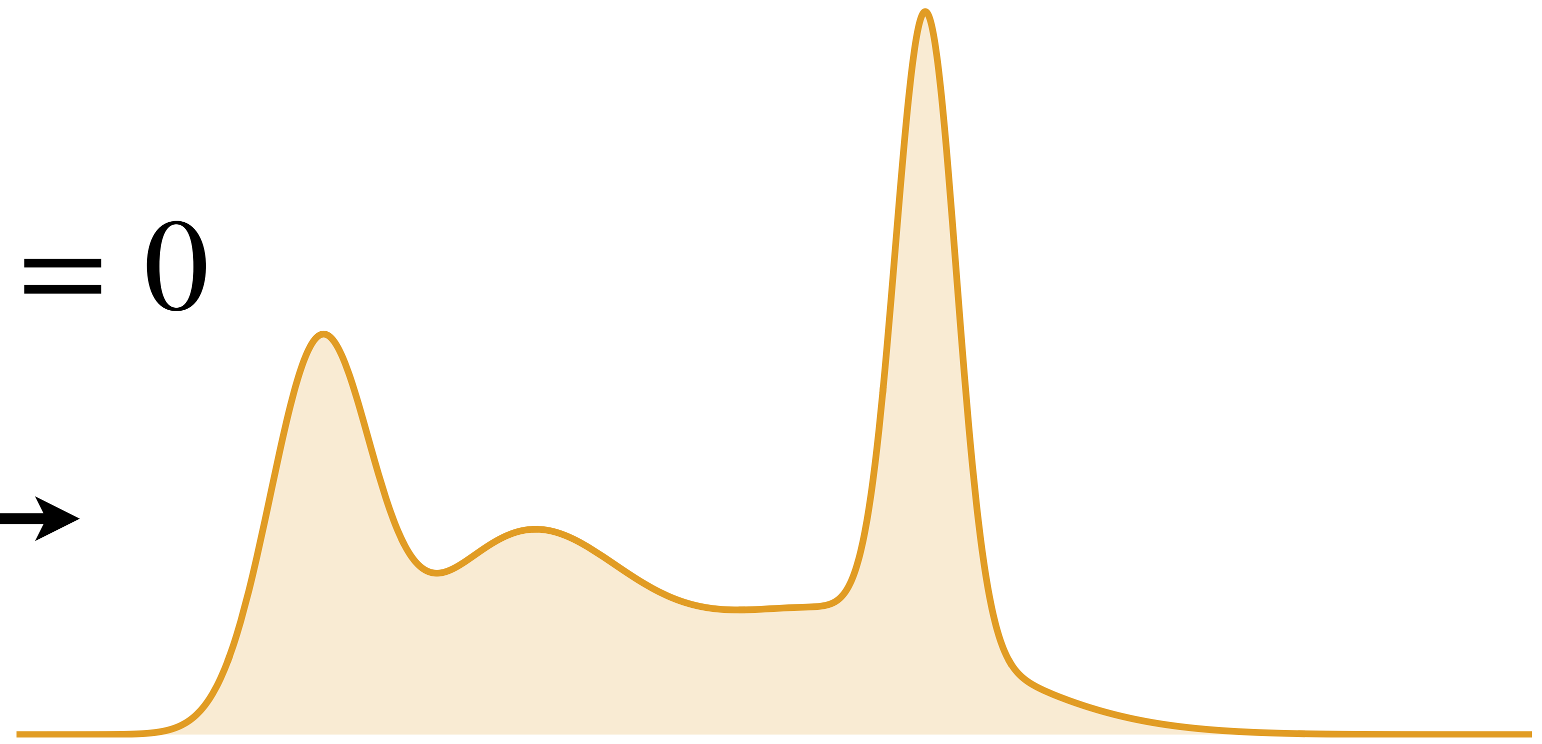
ground truth

velocity field

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

$p(\boldsymbol{x},0) = q(\boldsymbol{x})$

data distribution



$$\mathscr{L} = \mathbb{E}_{t\sim\mathscr{U}(0,1)}\mathbb{E}_{\boldsymbol{x}\sim p(\boldsymbol{x},t)}\left|\boldsymbol{v}_\theta(\boldsymbol{x},t) - \boldsymbol{u}(\boldsymbol{x},t)\right|^2$$

Liu et al 2209.03003, Albergo et al, 2209.15571, Lipman et al, 2210.02747

https://neurips.cc/virtual/2024/tutorial/99531

The “conditional” trick

Given a conditional continuity equation

$$\frac{\partial p(\boldsymbol{x} \mid \boldsymbol{x}_0, t)}{\partial t} + \nabla \cdot \left[p(\boldsymbol{x} \mid \boldsymbol{x}_0, t) \boldsymbol{u}(\boldsymbol{x} \mid \boldsymbol{x}_0, t) \right] = 0$$

Then, up to a constant, we have

$$\mathscr{L} = \mathbb{E}_{t \sim \mathscr{U}(0,1)} \mathbb{E}_{\boldsymbol{x}_0 \sim q(\boldsymbol{x}_0)} \mathbb{E}_{\boldsymbol{x} \sim p(\boldsymbol{x} \mid \boldsymbol{x}_0, t)} \left| \boldsymbol{v}_\theta(\boldsymbol{x}, t) - \boldsymbol{u}(\boldsymbol{x} \mid \boldsymbol{x}_0, t) \right|^2$$

We can learn the ground truth velocity by regressing on the conditional velocity

Claim: $\mathscr{L}_{\mathrm{CFM}} = \mathscr{L}_{\mathrm{FM}} + \mathrm{const}\,.$

where $$\mathscr{L}_{\mathrm{FM}} = \mathbb{E}_{t\sim\mathscr{U}(0,1)}\mathbb{E}_{\boldsymbol{x}\sim p(\boldsymbol{x},t)}\left|\boldsymbol{v}_\theta(\boldsymbol{x},t) - \boldsymbol{u}(\boldsymbol{x},t)\right|^2$$

$$\mathscr{L}_{\mathrm{CFM}} = \mathbb{E}_{t\sim\mathscr{U}(0,1)}\mathbb{E}_{\boldsymbol{x}_0\sim q(\boldsymbol{x}_0)}\mathbb{E}_{\boldsymbol{x}\sim p(\boldsymbol{x}|\boldsymbol{x}_0,t)}\left|\boldsymbol{v}_\theta(\boldsymbol{x},t) - \boldsymbol{u}(\boldsymbol{x}\,|\,\boldsymbol{x}_0,t)\right|^2$$

$$p(\boldsymbol{x},t) = \int p(\boldsymbol{x}\,|\,\boldsymbol{x}_0,t)\,q(\boldsymbol{x}_0)d\boldsymbol{x}_0 \qquad p(\boldsymbol{x},t)\boldsymbol{u}(\boldsymbol{x},t) = \int p(\boldsymbol{x}\,|\,\boldsymbol{x}_0,t)\boldsymbol{u}(\boldsymbol{x}\,|\,\boldsymbol{x}_0,t)\,q(\boldsymbol{x}_0)d\boldsymbol{x}_0$$

Proof:

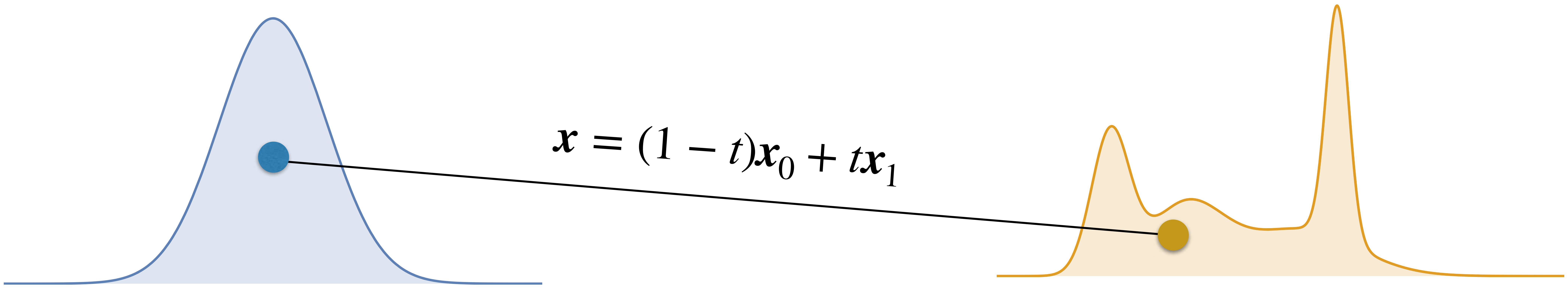
$$\mathbb{E}_{\boldsymbol{x}_0\sim q(\boldsymbol{x}_0)}\mathbb{E}_{\boldsymbol{x}\sim p(\boldsymbol{x}|\boldsymbol{x}_0,t)}\left|\boldsymbol{v}_\theta\right|^2 = \int d\boldsymbol{x}_0\int d\boldsymbol{x}q(\boldsymbol{x}_0)p(\boldsymbol{x}\,|\,\boldsymbol{x}_0,t)\,|\,\boldsymbol{v}_\theta|^2 = \int d\boldsymbol{x}p(\boldsymbol{x},t)\,|\,\boldsymbol{v}_\theta|^2 = \mathbb{E}_{\boldsymbol{x}\sim p(\boldsymbol{x},t)}\left|\boldsymbol{v}_\theta\right|^2$$

$$\mathbb{E}_{\boldsymbol{x}_0\sim q(\boldsymbol{x}_0)}\mathbb{E}_{\boldsymbol{x}\sim p(\boldsymbol{x}|\boldsymbol{x}_0,t)}\left[\boldsymbol{v}_\theta\cdot\boldsymbol{u}(\boldsymbol{x}\,|\,\boldsymbol{x}_0,t)\right] = \int d\boldsymbol{x}_0\int d\boldsymbol{x}q(\boldsymbol{x}_0)p(\boldsymbol{x}\,|\,\boldsymbol{x}_0,t)\left[\boldsymbol{v}_\theta\cdot\boldsymbol{u}(\boldsymbol{x}\,|\,\boldsymbol{x}_0,t)\right]$$

$$= \int d\boldsymbol{x}p(\boldsymbol{x},t)\boldsymbol{v}_\theta\cdot\boldsymbol{u}(\boldsymbol{x},t) = \mathbb{E}_{\boldsymbol{x}\sim p(\boldsymbol{x},t)}\left[\boldsymbol{v}_\theta\cdot\boldsymbol{u}(\boldsymbol{x},t)\right]$$

Examples of flow matching

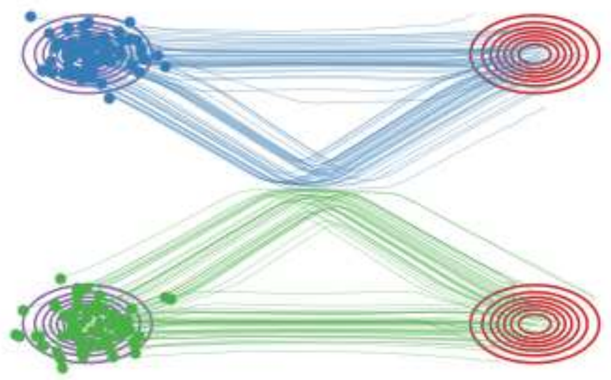
$$p(\boldsymbol{x} \,|\, \boldsymbol{x}_0, t) = \mathcal{N}\left((1-t)\boldsymbol{x}_0, t^2\right) \qquad \boldsymbol{u}(\boldsymbol{x} \,|\, \boldsymbol{x}_0, t) = \frac{d\boldsymbol{x}}{dt} = \boldsymbol{x}_1 - \boldsymbol{x}_0$$



$$\boldsymbol{x} = (1-t)\boldsymbol{x}_0 + t\boldsymbol{x}_1$$

$$\boldsymbol{x}_1 \sim \mathcal{N}(0,I) \qquad \boldsymbol{x}_0 \sim q(\boldsymbol{x})$$

$$\mathcal{L} = \mathbb{E}_{t\sim\mathcal{U}(0,1)}\mathbb{E}_{\boldsymbol{x}_0\sim q(\boldsymbol{x}_0)}\mathbb{E}_{\boldsymbol{x}_1\sim\mathcal{N}(0,I)}\left|\boldsymbol{v}_\theta(\boldsymbol{x},t) - (\boldsymbol{x}_1 - \boldsymbol{x}_0)\right|^2$$

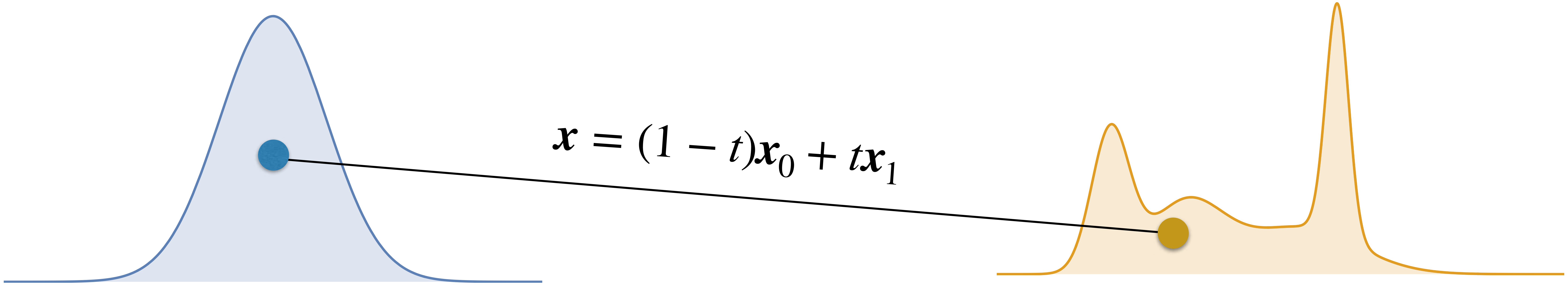


Causalizing linear interpolation with rectified flow 2209.03003

https://www.cs.utexas.edu/~lqiang/rectflow/html/intro.html

Examples of flow matching

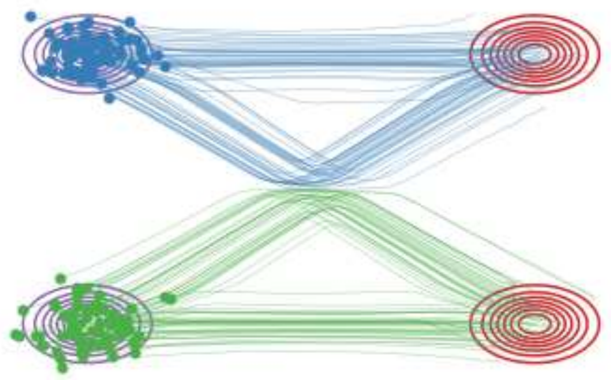
$$p(\boldsymbol{x} \,|\, \boldsymbol{x}_0, t) = \mathcal{N}\left((1-t)\boldsymbol{x}_0, t^2\right) \qquad \boldsymbol{u}(\boldsymbol{x} \,|\, \boldsymbol{x}_0, t) = \frac{d\boldsymbol{x}}{dt} = \boldsymbol{x}_1 - \boldsymbol{x}_0$$



$$\boldsymbol{x} = (1-t)\boldsymbol{x}_0 + t\boldsymbol{x}_1$$

$\boldsymbol{x}_1 \sim \mathcal{N}(0, I)$ $\qquad$ $\boldsymbol{x}_0 \sim q(\boldsymbol{x})$

$$\mathcal{L} = \mathbb{E}_{t \sim \mathcal{U}(0,1)} \mathbb{E}_{\boldsymbol{x}_0 \sim q(\boldsymbol{x}_0)} \mathbb{E}_{\boldsymbol{x}_1 \sim \mathcal{N}(0,I)} \left| \boldsymbol{v}_\theta(\boldsymbol{x}, t) - (\boldsymbol{x}_1 - \boldsymbol{x}_0) \right|^2$$



Causalizing linear interpolation with rectified flow 2209.03003

https://www.cs.utexas.edu/~lqiang/rectflow/html/intro.html

Flow matching is all you need!

This framework contains various diffusion models as special cases

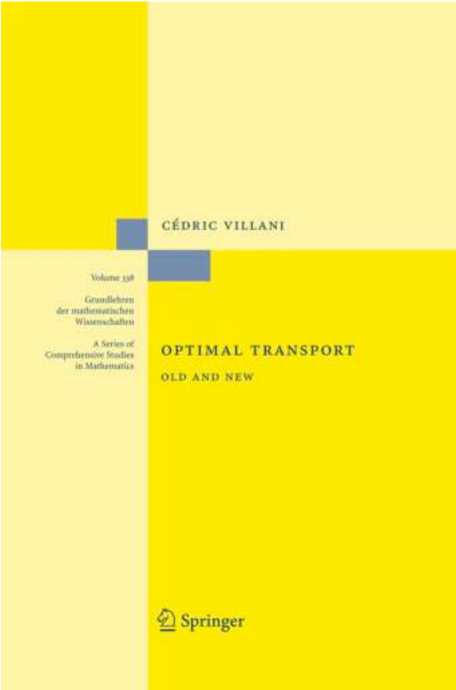
The base distribution does not have to be Gaussian

Fast generation with rectified transportation path (Liu et al 2209.03003)

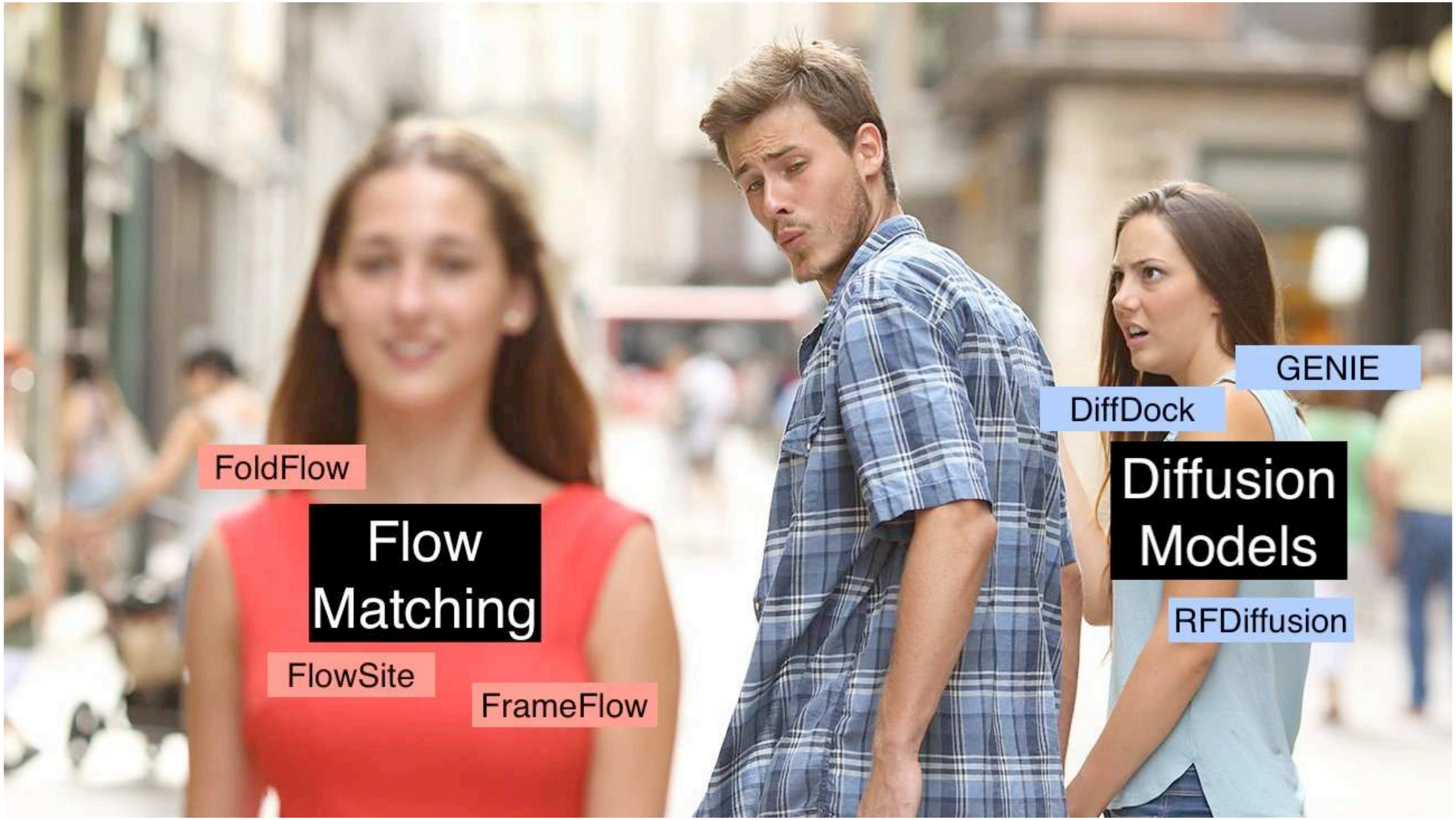
400x speedup compared to continuous normalizing flow (Albergo et al, 2209.15571)

Surpasses diffusion model on Imagenet in likelihood and sample quality (Lipman et al, 2210.02747)

Generalization to flow on Riemannian manifolds (Chen et al, 2302.03660)



Part II Optimal transport and Riemannian geometry

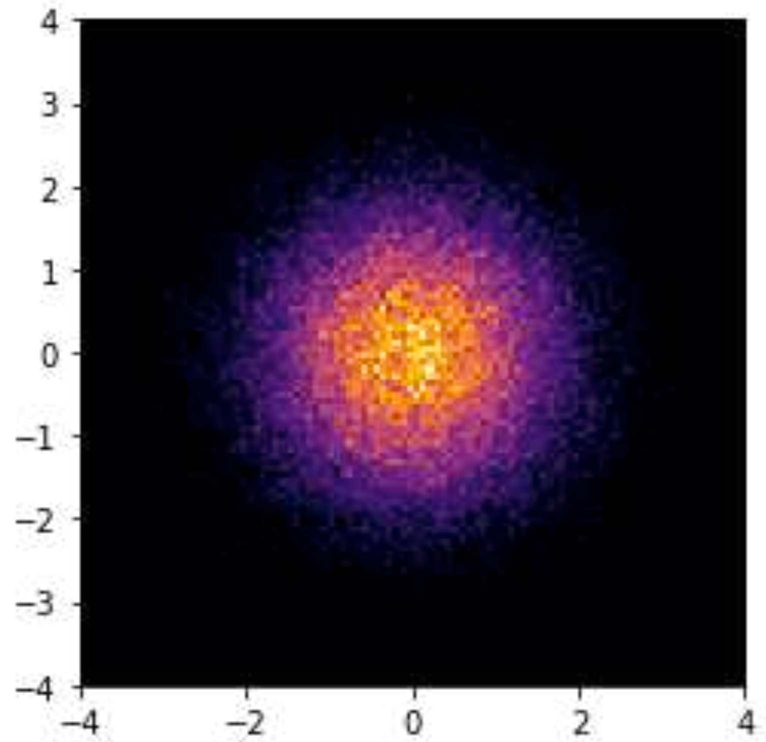


https://twitter.com/michael_galkin/status/1711845455817261409

Demo: bounding free energy of classical Coulomb gas

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

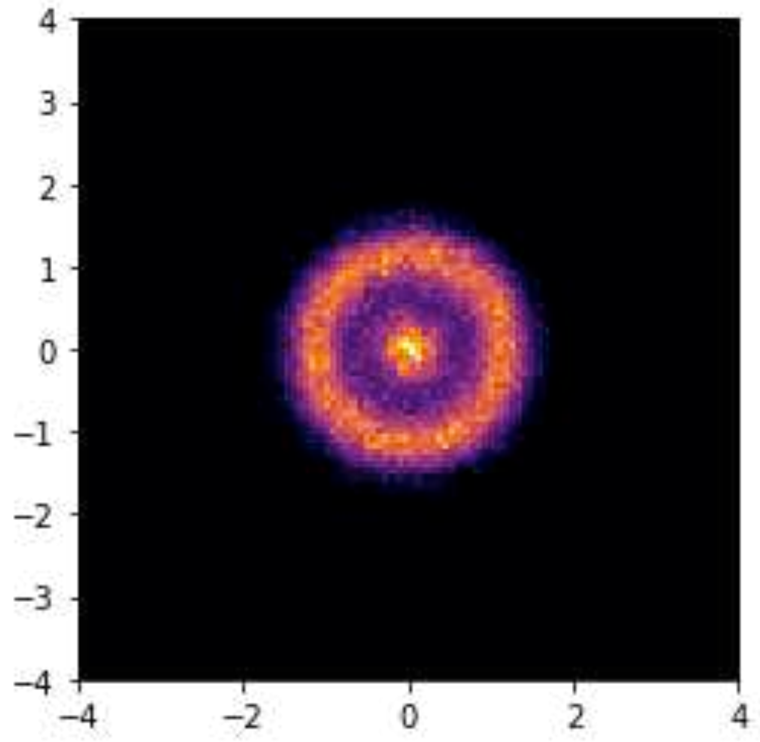
$$Z = \mathbb{E}_{\boldsymbol{x}\sim q(\boldsymbol{x})}\left[e^{-\beta E(\boldsymbol{x}) - \ln q(\boldsymbol{x})}\right] \quad \ln q(\boldsymbol{x}) = \ln \mathcal{N}(0,I) - \int_0^1 \nabla\cdot\boldsymbol{v}dt$$



Base density

Gaussian samples

Interpolate samples to estimate free energy differences

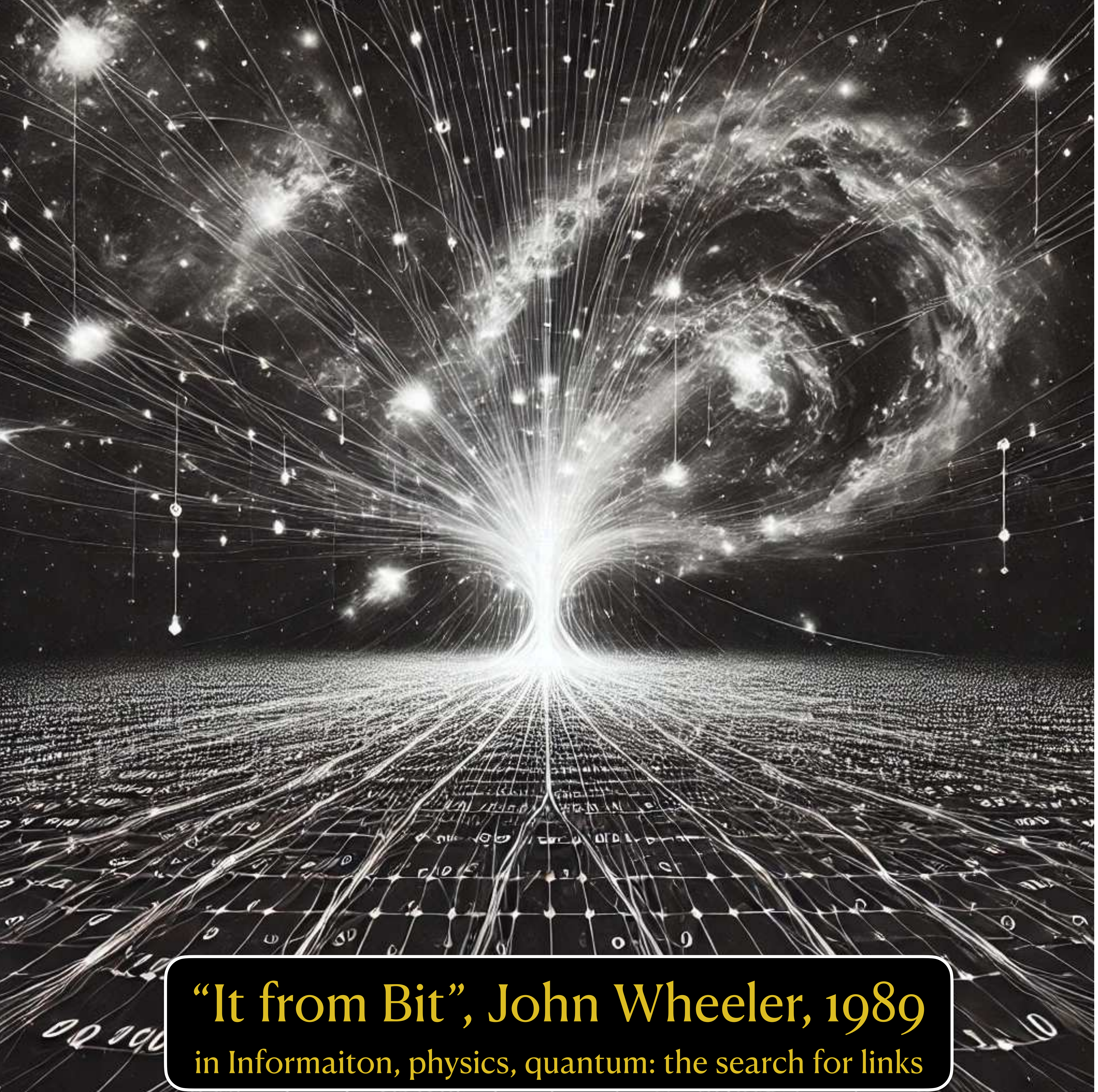


Target density

Monte Carlo samples

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

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



"It from Bit", John Wheeler, 1989

in Informaiton, physics, quantum: the search for links

Generative AI for It

Hamiltonian dynamics as symplectic gradient flow

Hamiltonian equations

$$\begin{cases} \dot{p} = -\frac{\partial H}{\partial q} \\ \dot{q} = +\frac{\partial H}{\partial p} \end{cases}$$

Phase space variables

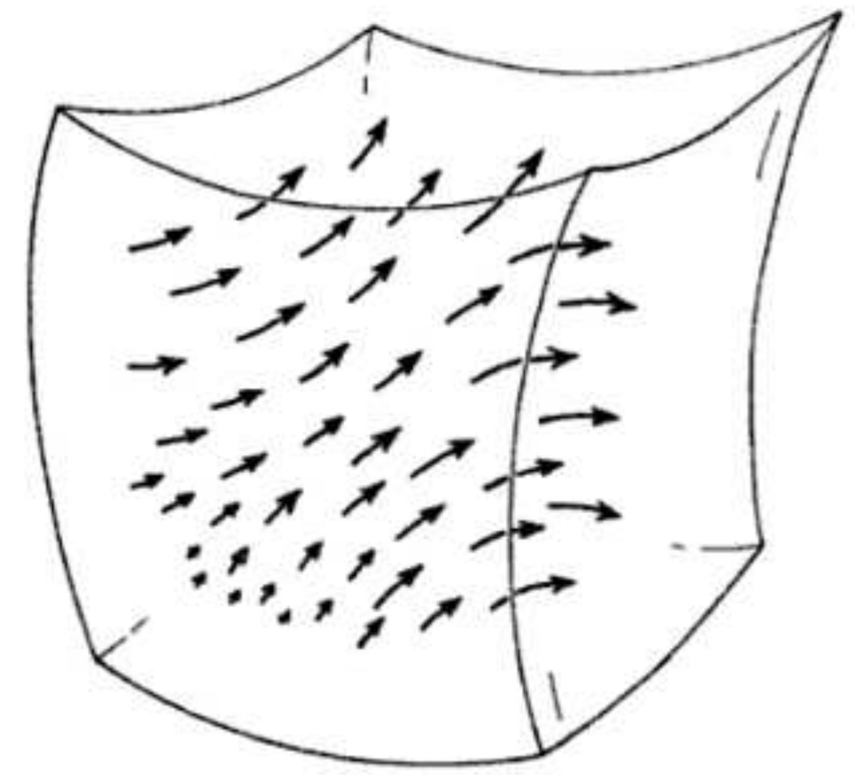
$$\boldsymbol{x} = (p, q)$$

Symplectic metric

$$J = \begin{pmatrix} & I \\ -I & \end{pmatrix}$$

Symplectic gradient flow

$$\dot{\boldsymbol{x}} = \nabla_{\boldsymbol{x}} H(\boldsymbol{x}) J$$



Hamiltonian dynamics as symplectic gradient flow

Hamiltonian ec **pace va** **c gradient flow**

$= (p, q$ $\nabla_x H(\boldsymbol{x}) J$

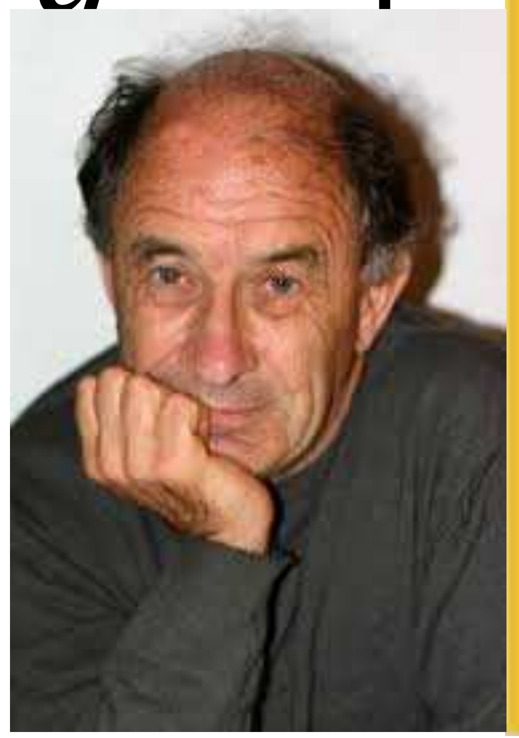
$\dot{p} = -$

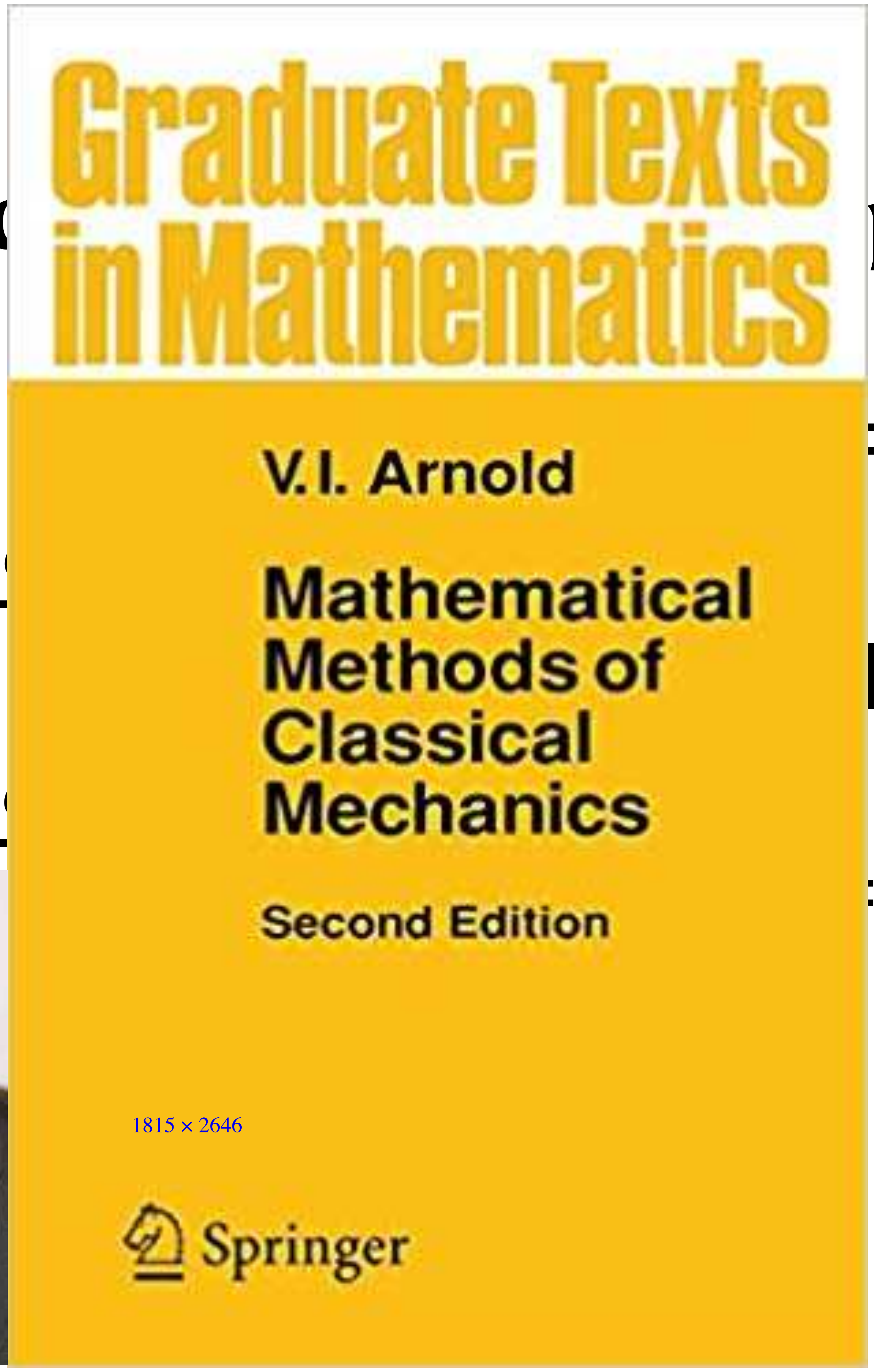
lectic m

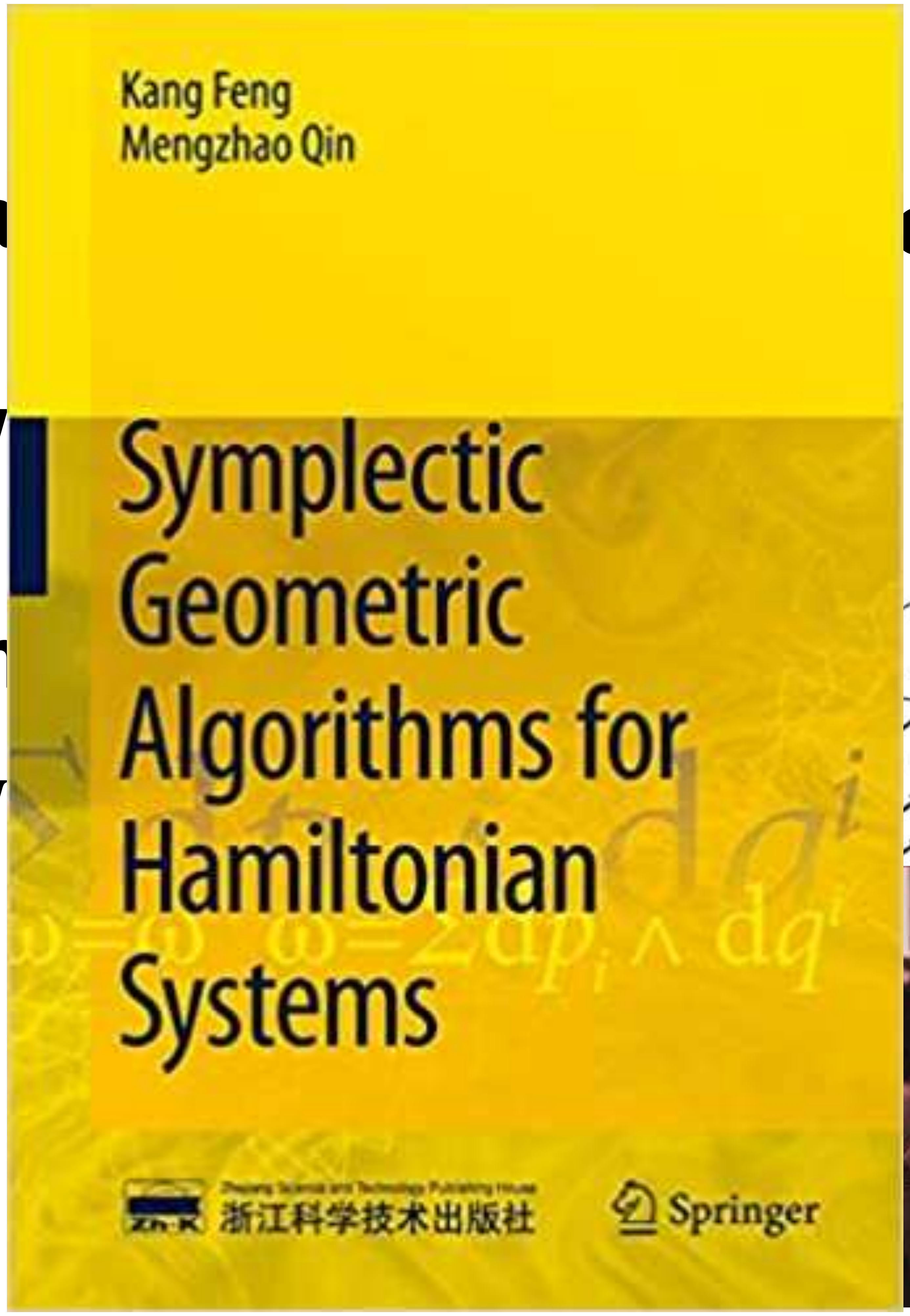
$\dot{q} = +$

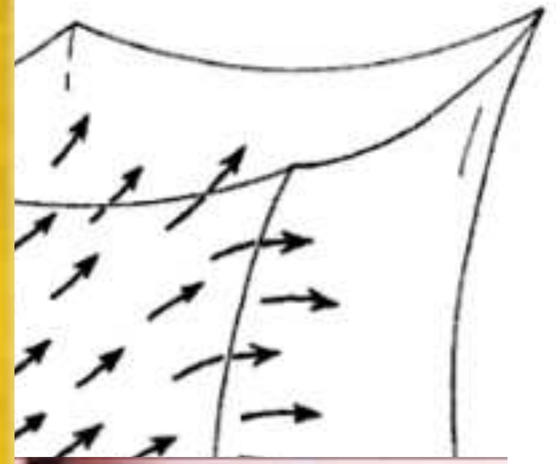
$($ I

$-I$

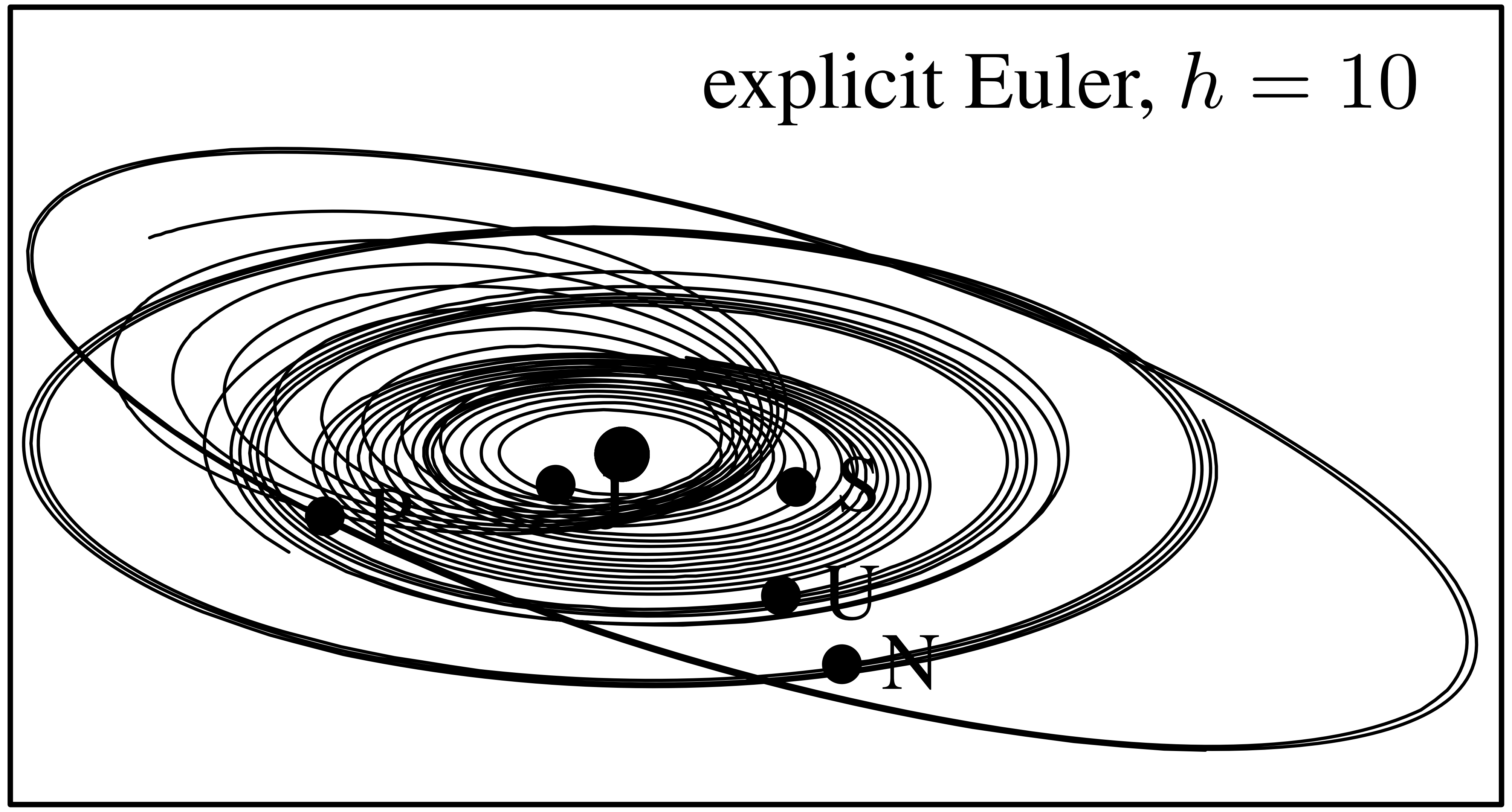


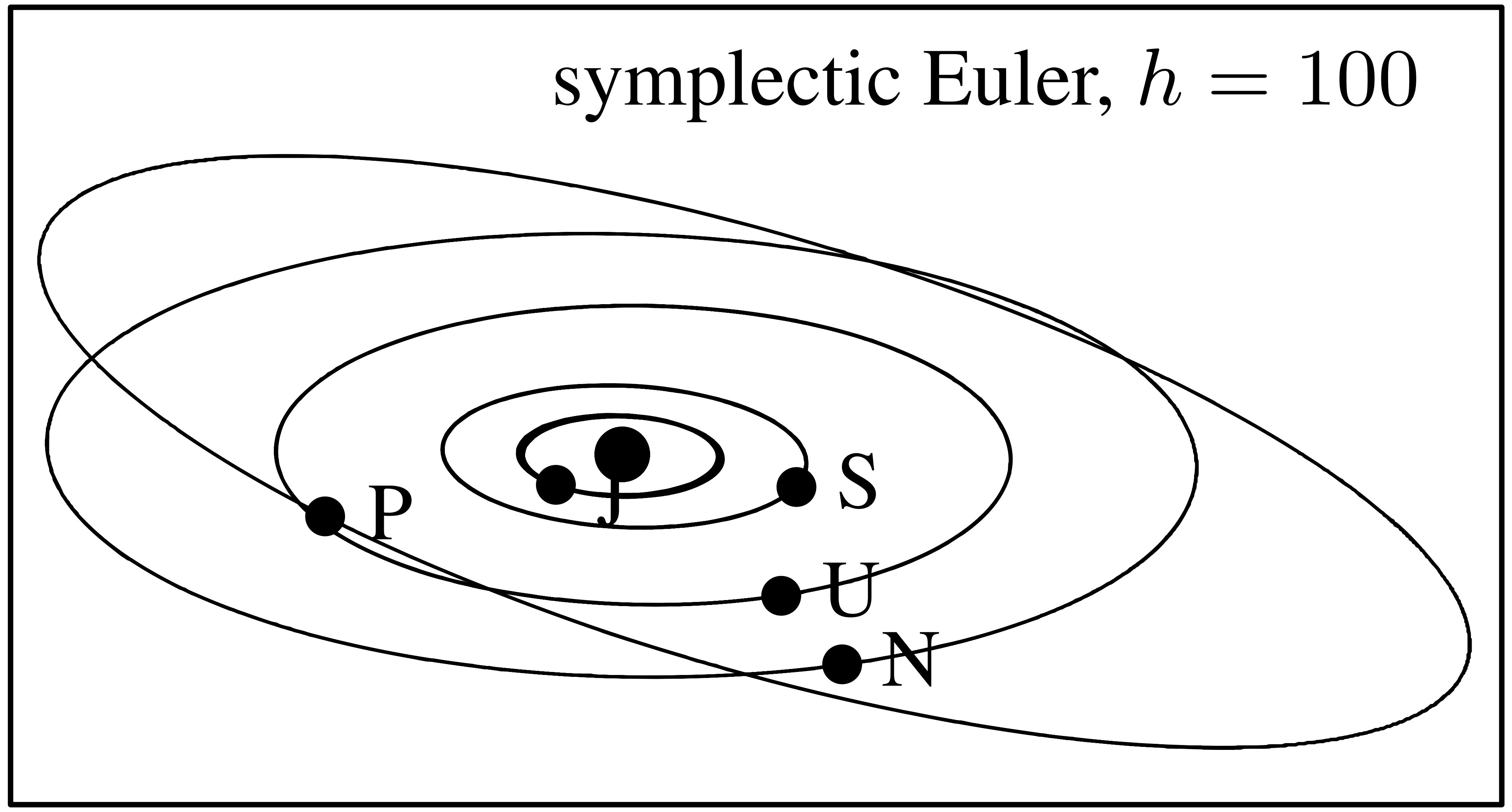






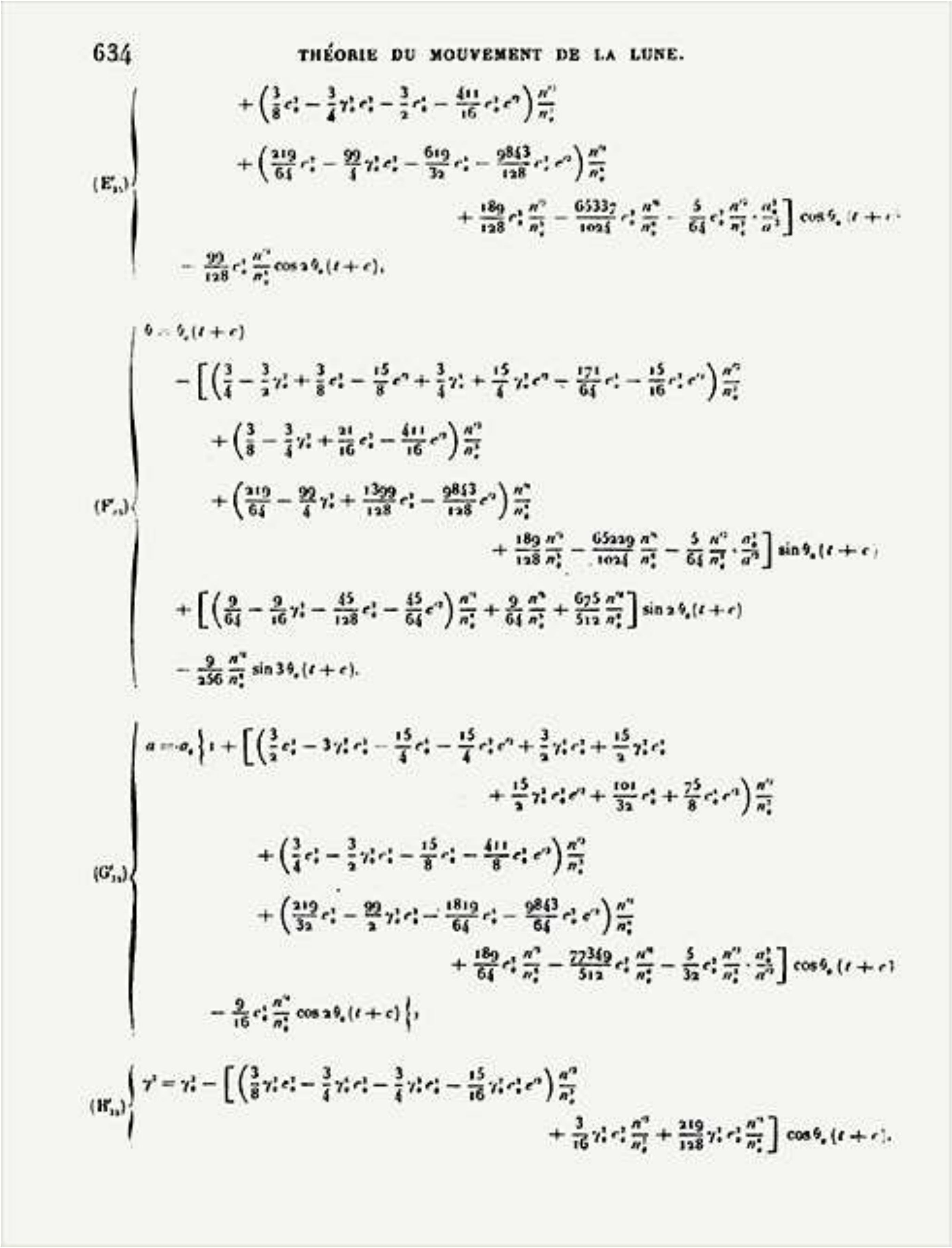
Symplectic Integrators

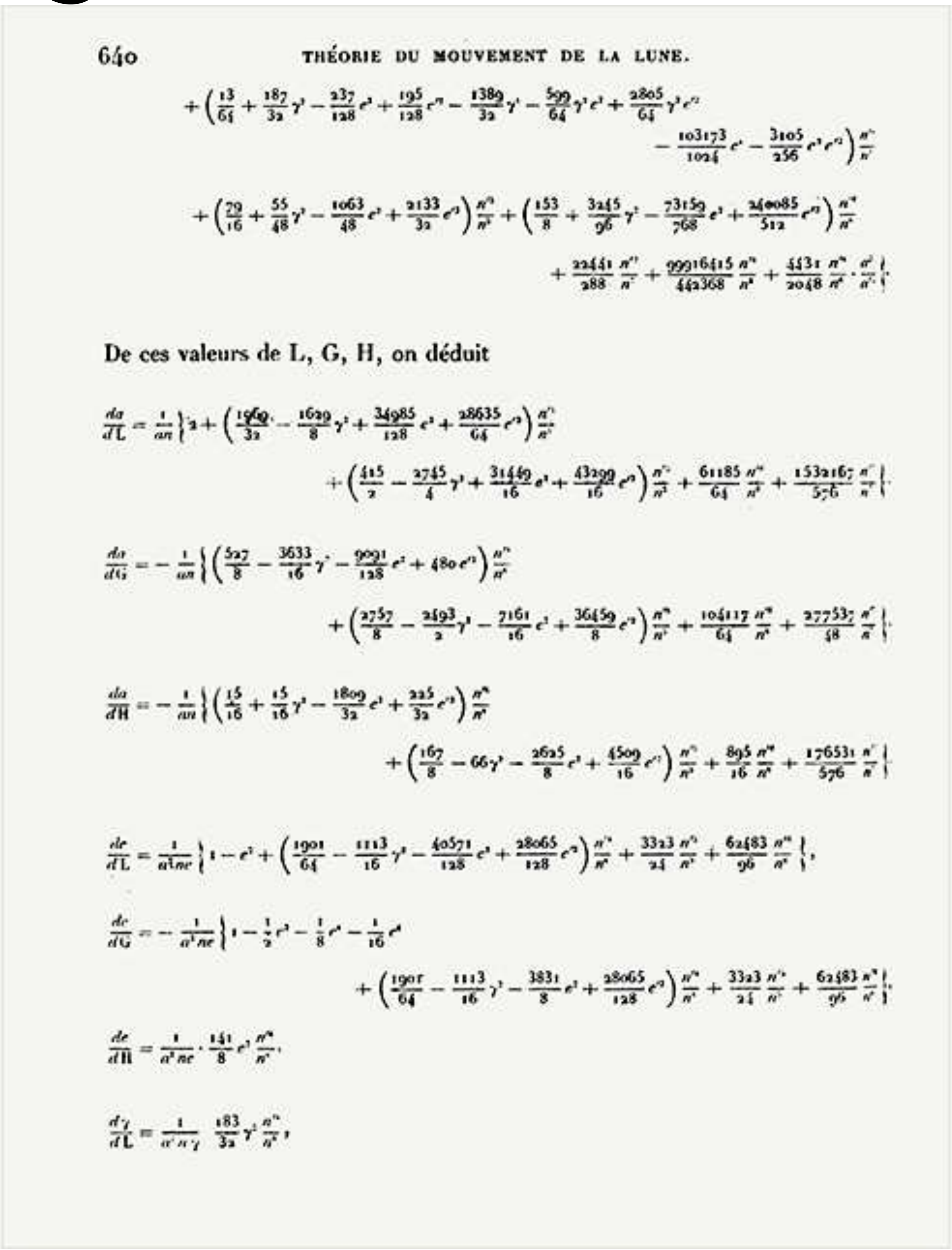




from Hairer et al, Geometric Numerical Integration

Canonical transformation for Moon-Earth-Sun 3-body problem


634 THÉORIE DU MOUVEMENT DE LA LUNE.


640 THÉORIE DU MOUVEMENT DE LA LUNE.

De ces valeurs de L, G, H, on déduit

Charles Delaunay

↑

More than 1800 pages of this, ~20 years of efforts (1846-1867)

How to find useful canonical transformations for more complex systems?

Canonical transformations and deep learning

Change of variables

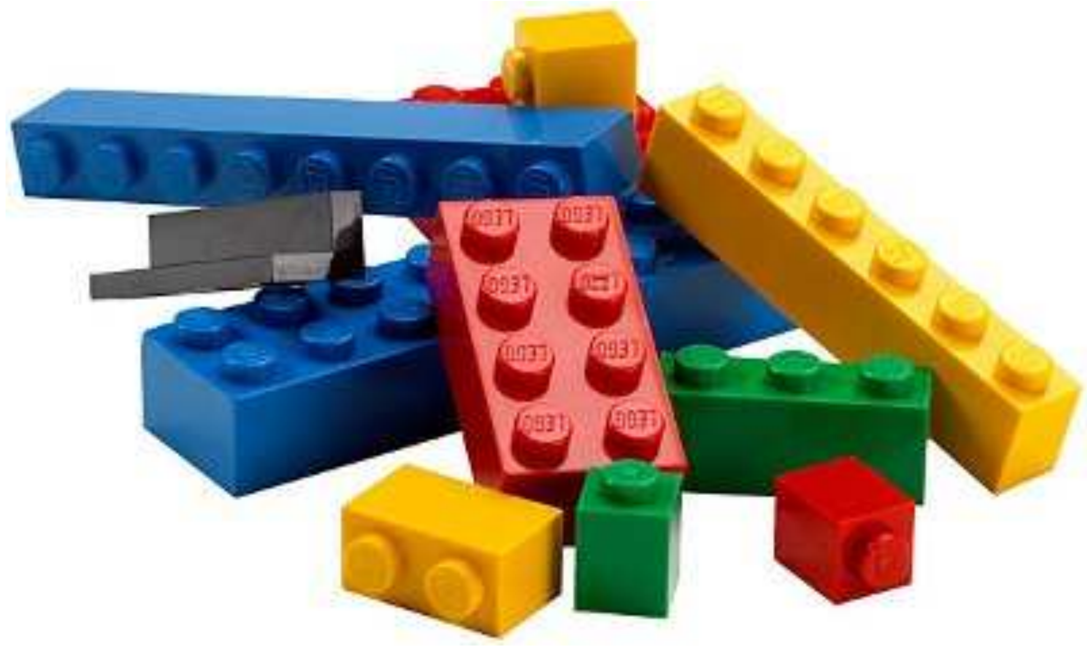
$$\boldsymbol{X} = (p, q) \longleftrightarrow \boldsymbol{Z} = (P, Q)$$

$$(\nabla_X \boldsymbol{Z}) \begin{pmatrix} & I \\ -I & \end{pmatrix} (\nabla_X \boldsymbol{Z})^T = \begin{pmatrix} & I \\ -I & \end{pmatrix}$$

symplectic condition

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

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



Compose symplectic layers to form a deep neural network
and learn them either from data or variationally without data

Canonical transformations and generative models

Canonical transformation deforms phase space density $p(\boldsymbol{X}) = e^{-\beta H(\boldsymbol{X})}$

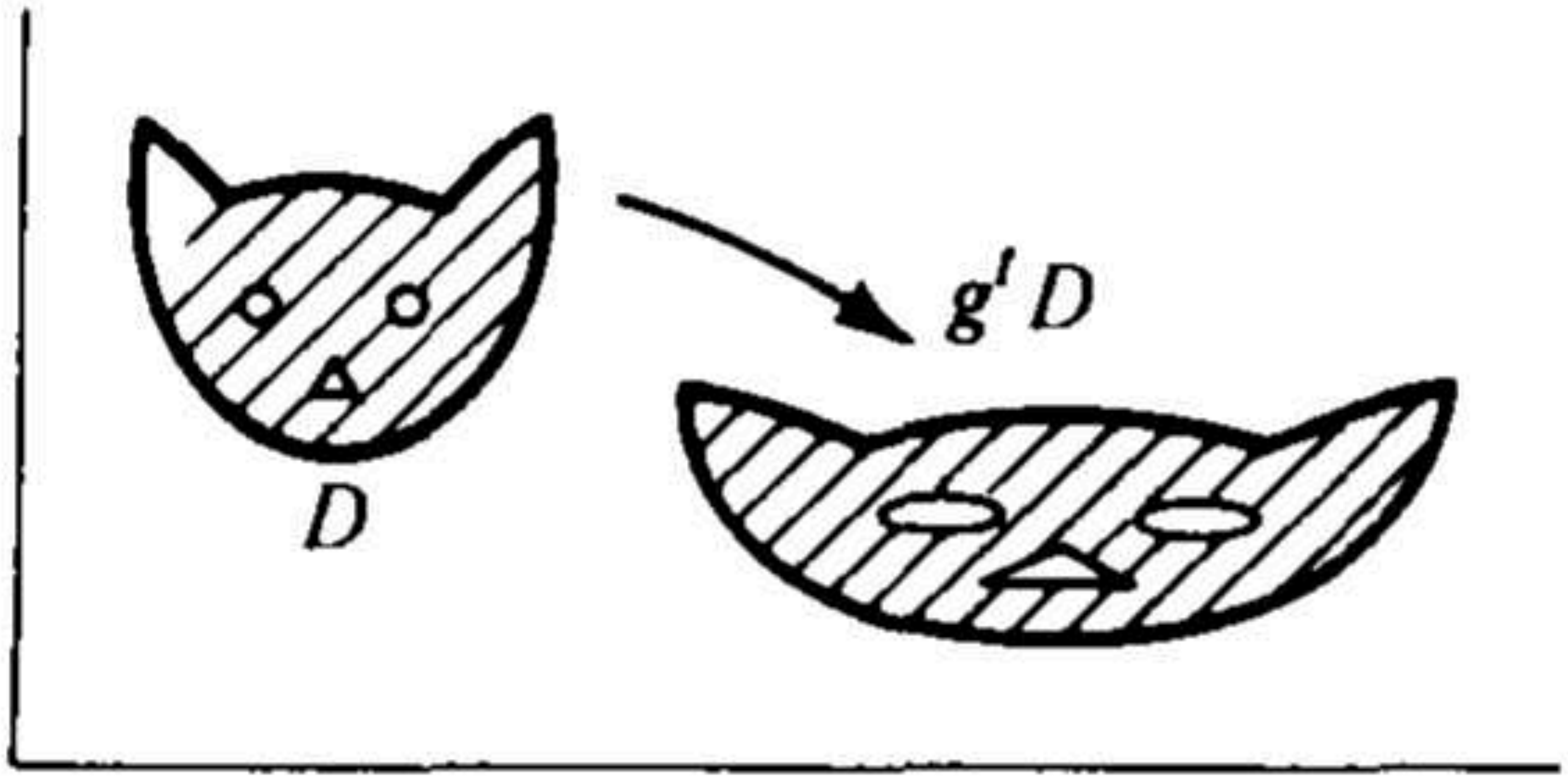


Image from Arnold '78

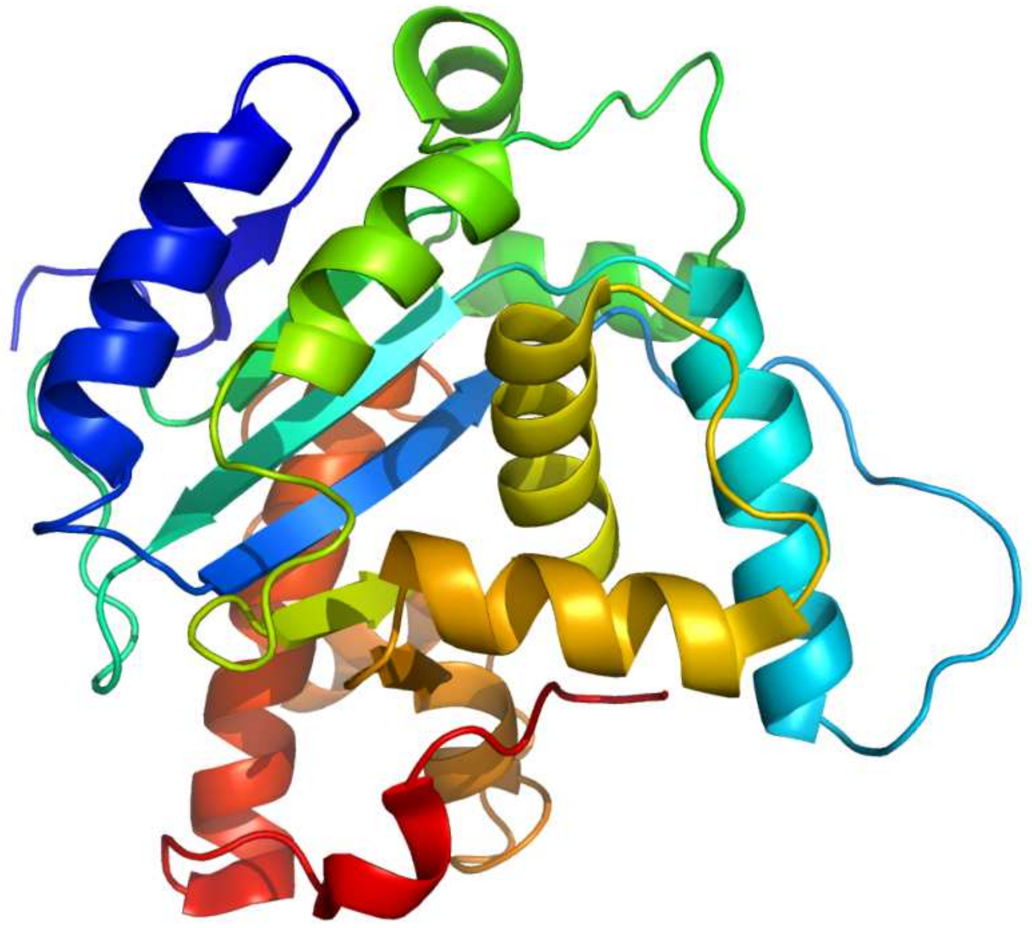
Flow models are great at transforming probability densities

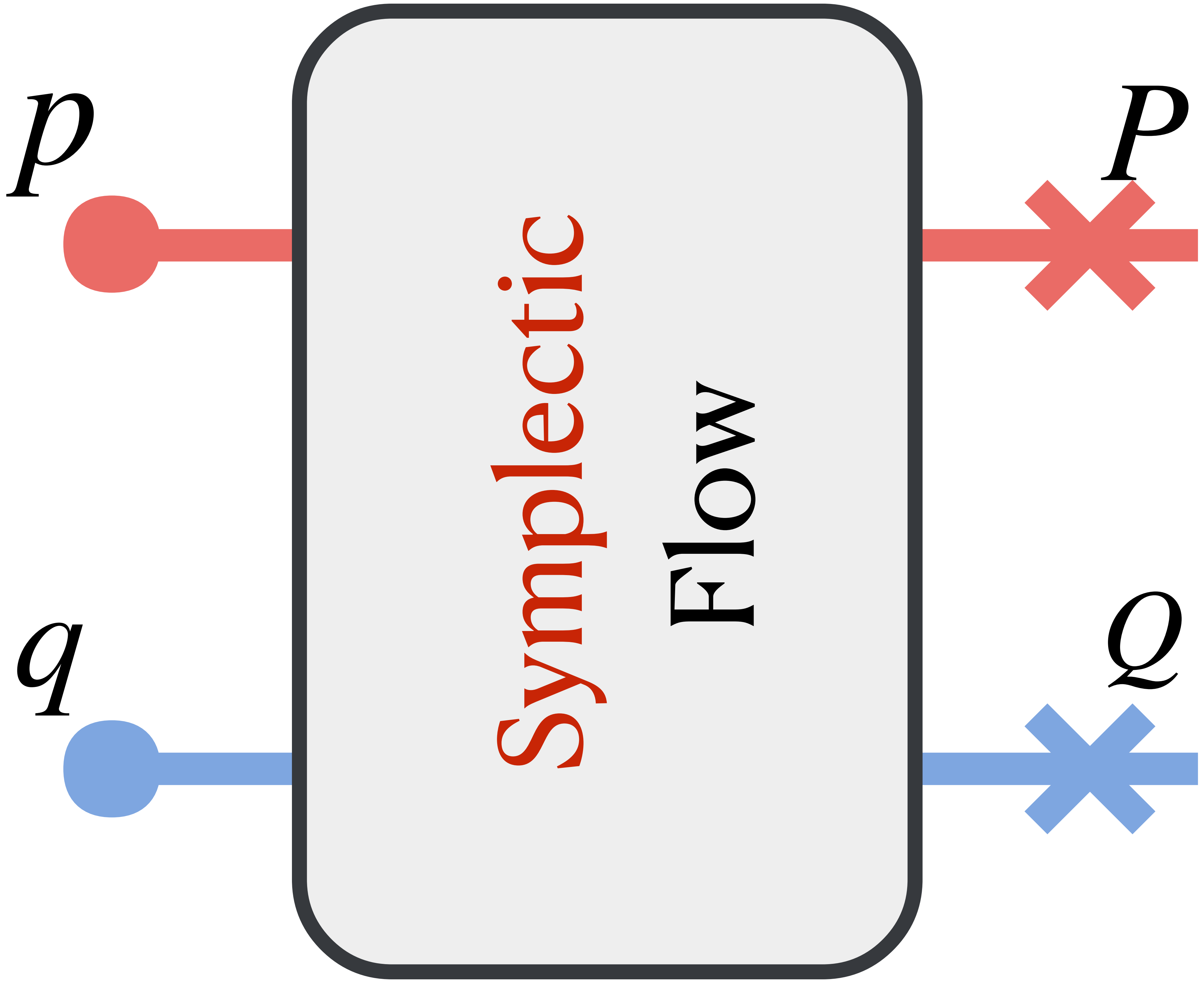
Neural Canonical Transformations

Li, Dong, Zhang, LW, PRX '20 li012589/neuralCT

$H(p, q)$

$$K(P, Q) = \sum_k \frac{P_k^2 + \omega_k^2 Q_k^2}{2}$$





physical space

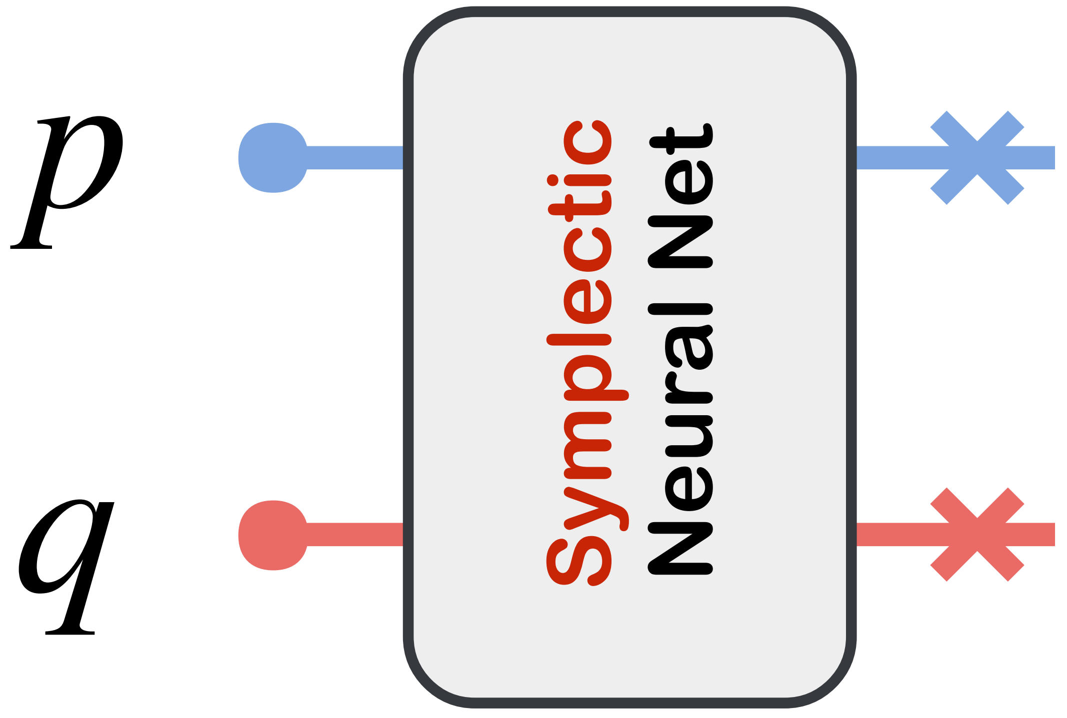
latent space

Learn harmonic frequencies of the base to identify slow collective modes

See Bondesan et al 1906.04645, Ishikawa et al 2103.00372 for investigations on integrability

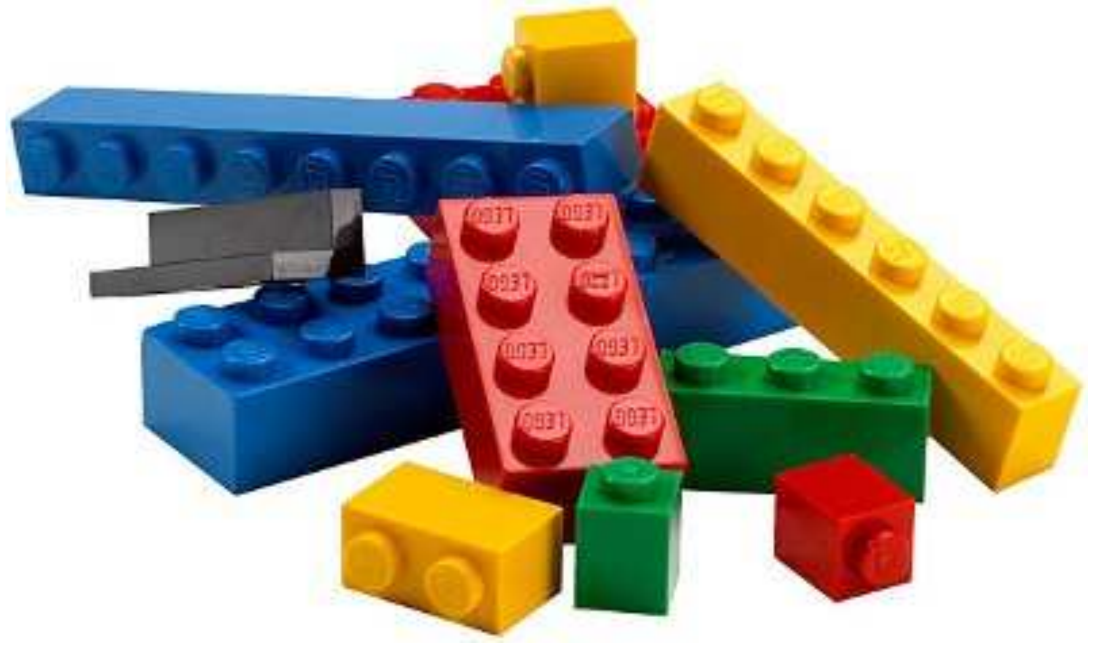
Symplectic primitives

- **Neural point transformations**



$$P = p\left(\nabla_Q q\right)$$

$$Q = f(q)$$ arbitrary invertible neural net

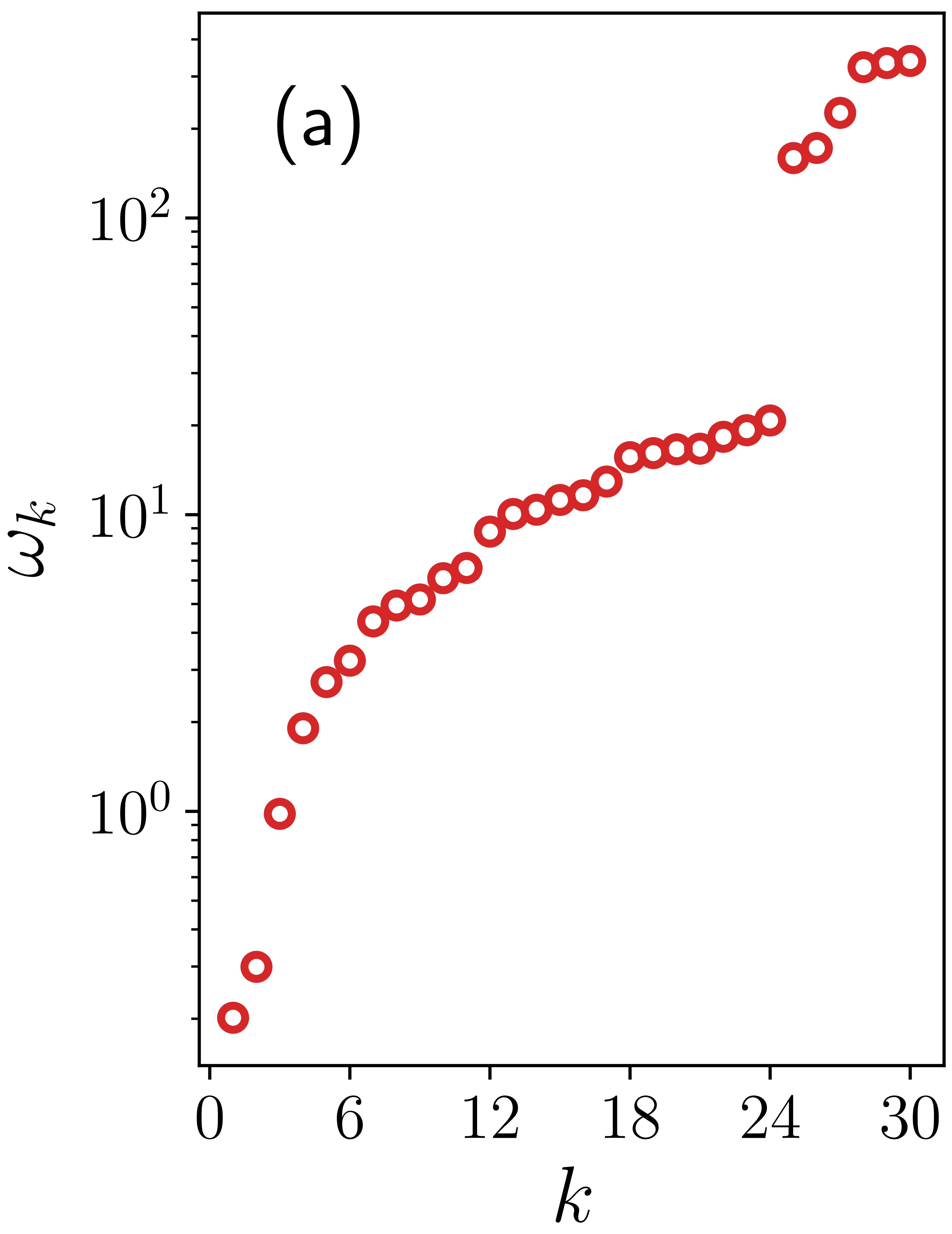


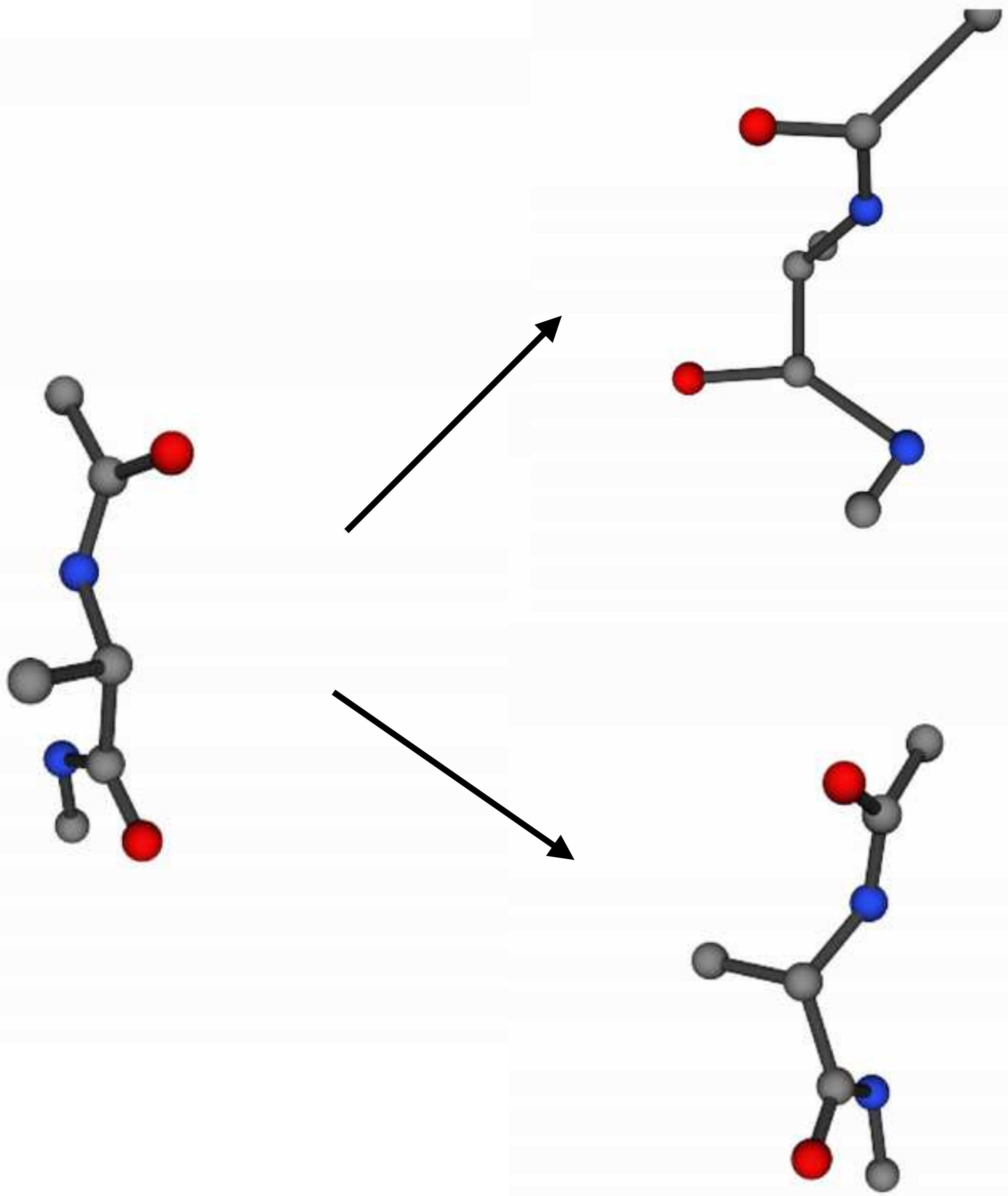
- Linear transformation: Symplectic Lie group Sp($2n$)

- Continuous-time flow: Symplectic generating functions via Hamiltonian dynaics See also Bondesan, Lamacraft, 1906.04645

Neural ODE, Chen et al, 1806.07366, Monge-Ampère flow, Zhang et al 1809.10188

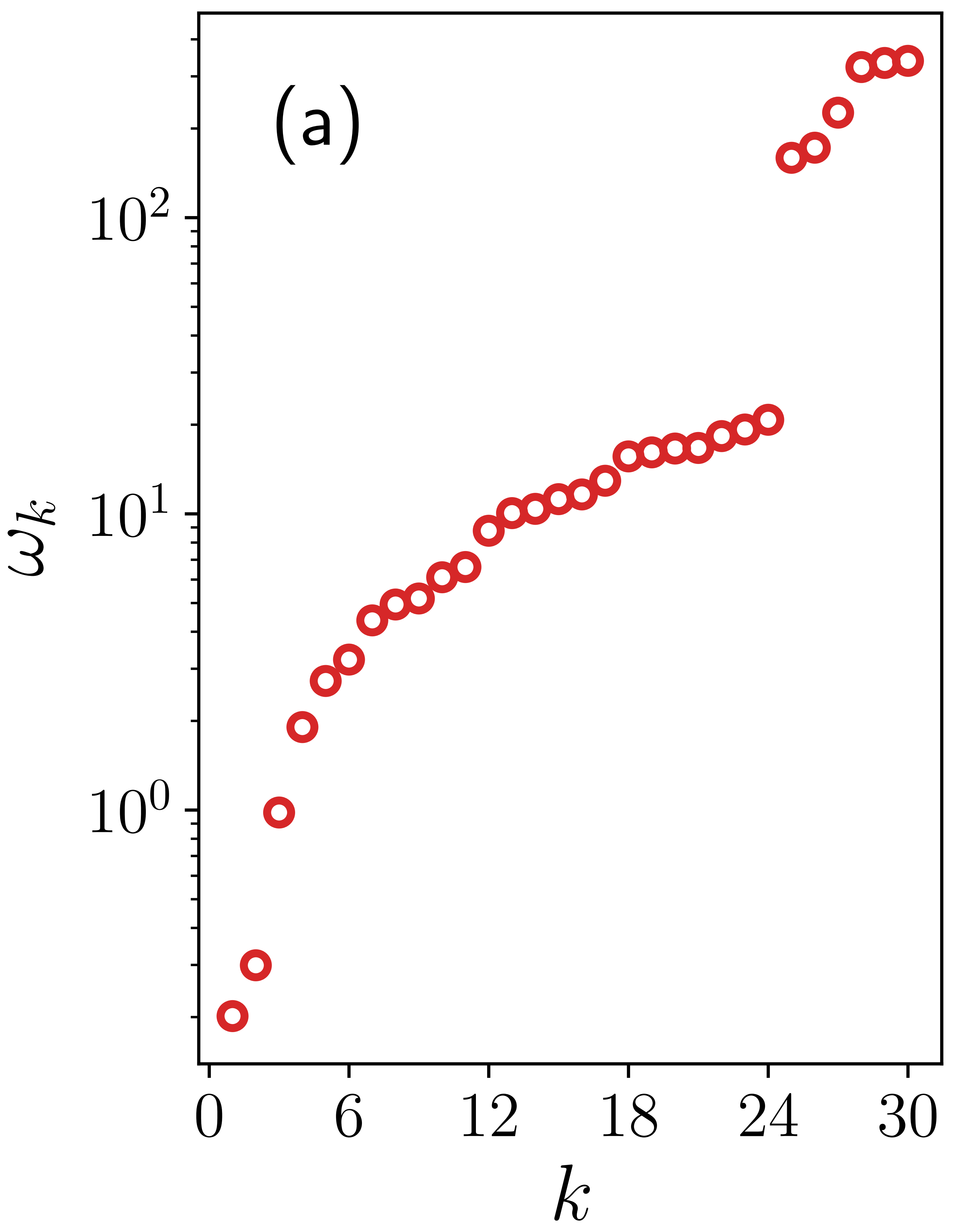
Li, Dong, Zhang, LW, PRX '20

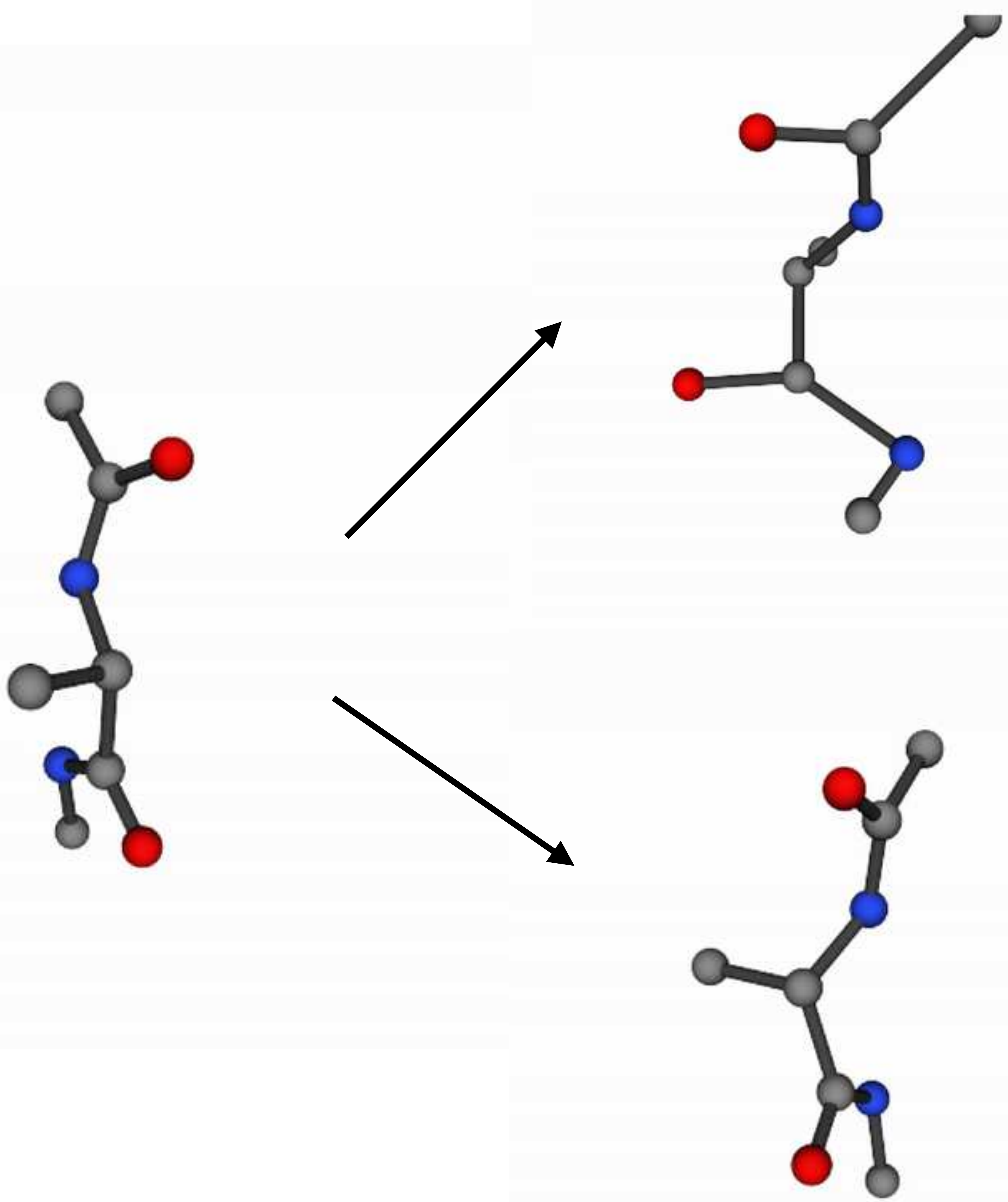




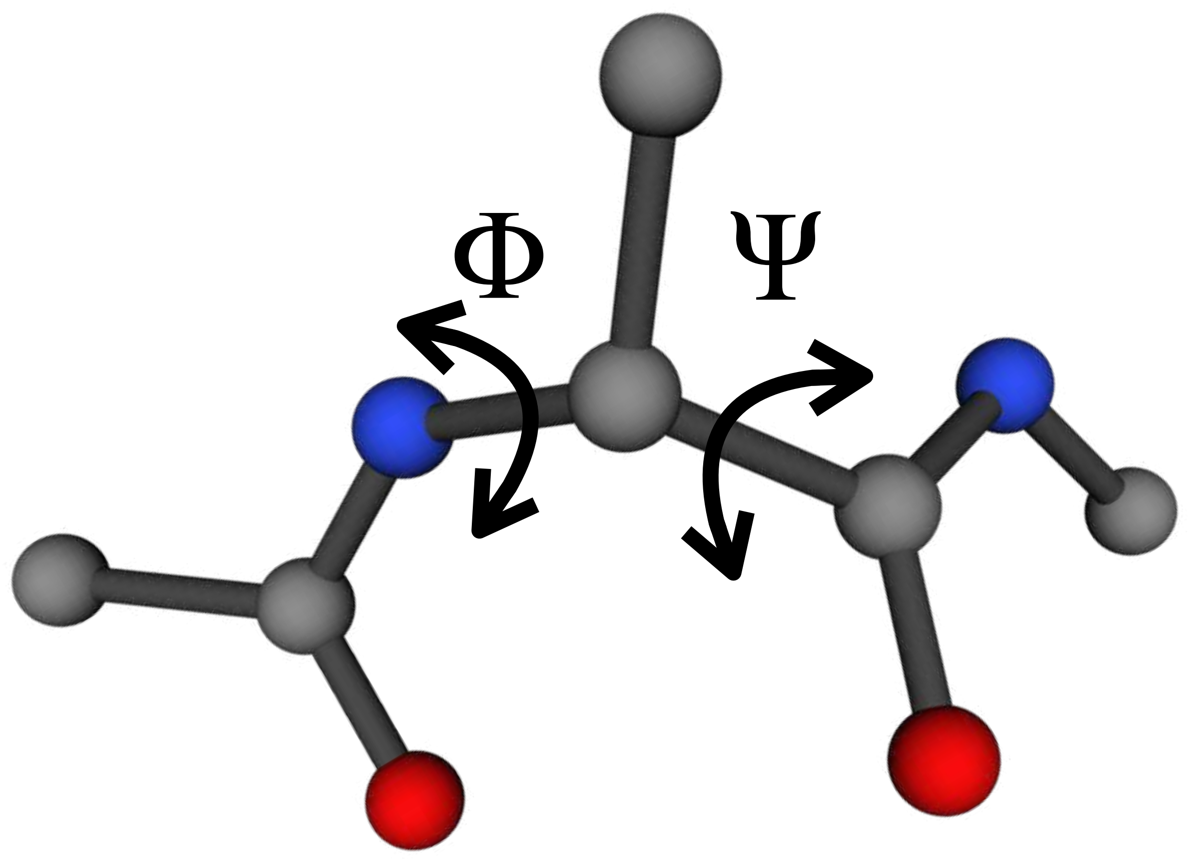
Neural canonical transformation identifies nonlinear slow modes

Li, Dong, Zhang, LW, PRX '20

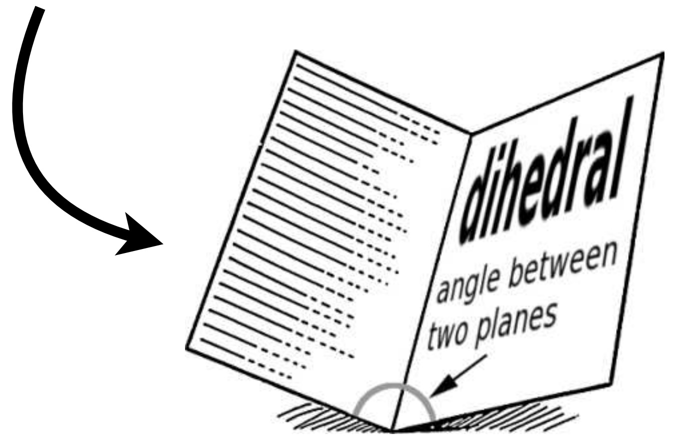


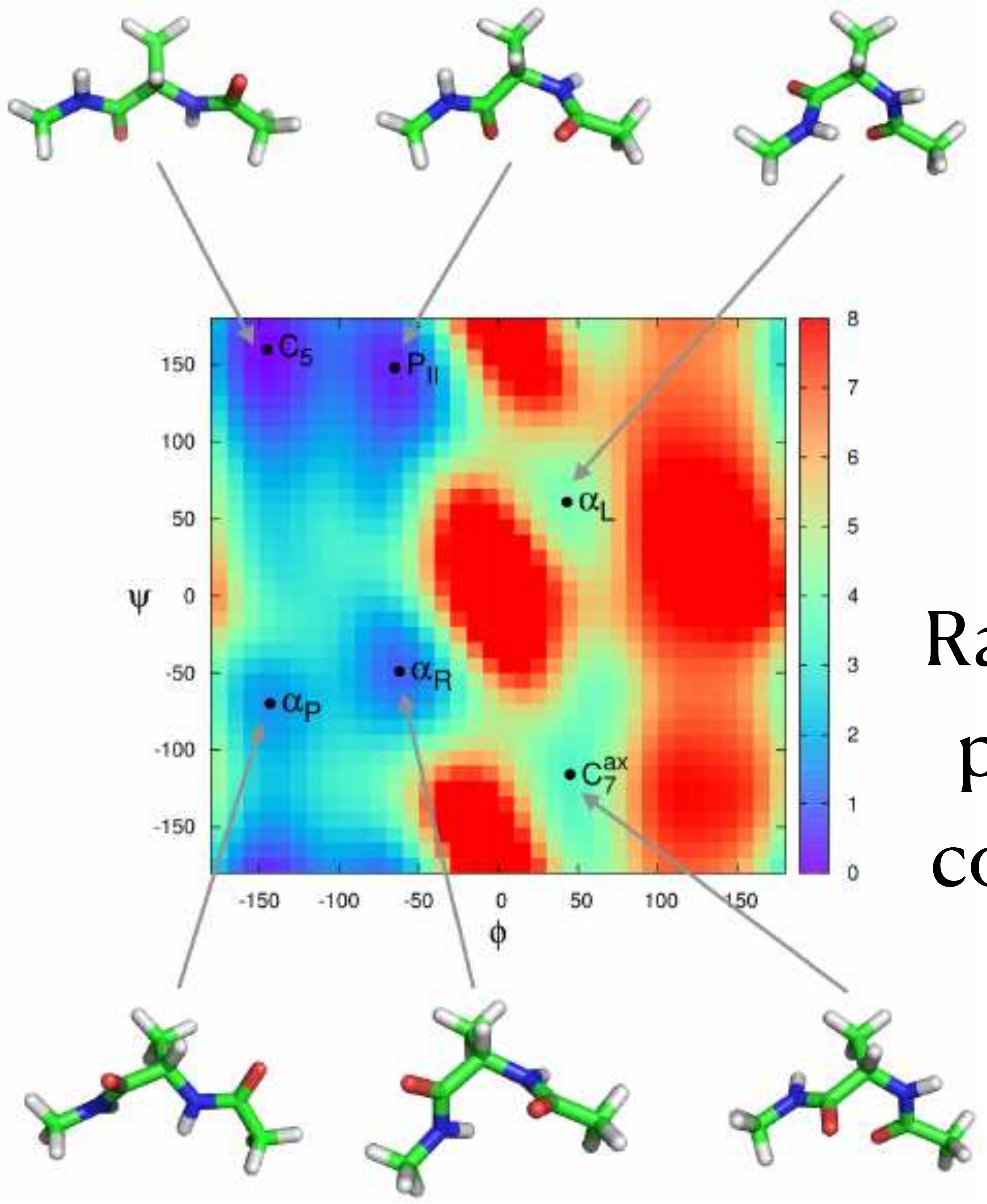


Neural canonical transformation identifies nonlinear slow modes



slow motion of the two torsion angles





Ramachandran plot of stable conformations

Dimensional reduction to slow collective variables useful for control, prediction, enhanced sampling...

check the paper 1910.00024, PRX '20 for more examples & applications

Entering the quantum world

Neural *canonical* transformations

Li, Dong, Zhang, LW, PRX '20

Xie, Zhang, LW, JML '21

classical world

Symplectic transformation

Probability density

$$p$$

Kullback-Leibler divergence

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

quantum world

Unitary transformation

Density matrix

$$\rho$$

Quantum relative entropy

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

Neural *canonical* transformations

Li, Dong, Zhang, LW, PRX '20

Xie, Zhang, LW, JML '21

classical world

Symplectic transformation

Probability density

$$p$$

Kullback-Leibler divergence

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

quantum world

Unitary transformation

Density matrix

$$\rho$$

Quantum relative entropy

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

The variational free energy principle

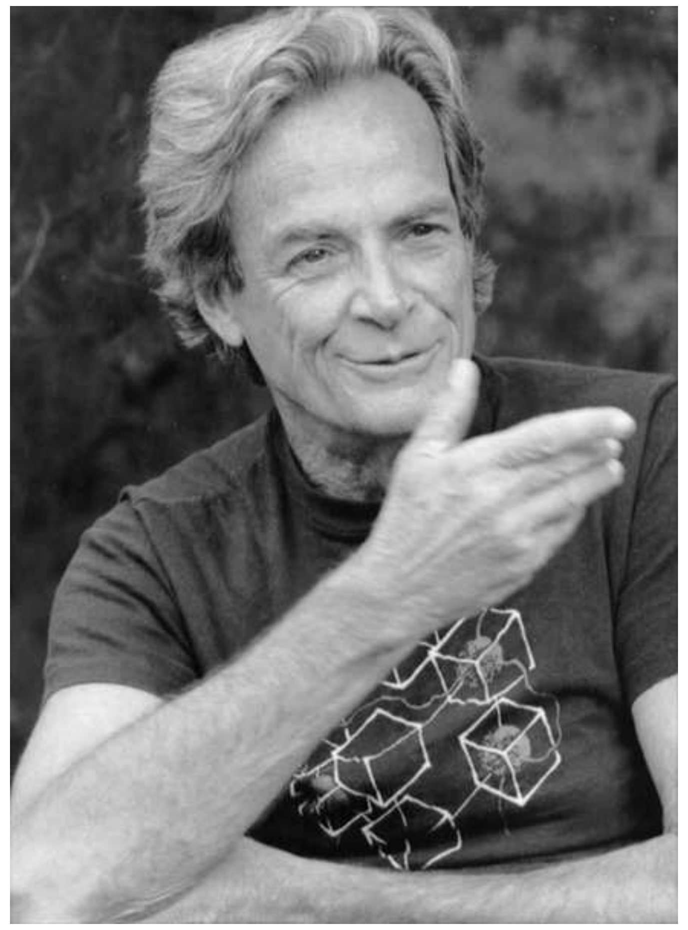
Gibbs–Bogolyubov-Feynman-Delbrück–Molière

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

↓ variational density matrix

↓ energy

↓ entropy



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

Richard P. Feynman

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

ρ ?

Generative models !

1981

14318 citations

Simulating Physics with Computers

Richard P. Feynman

"... it's a wonderful problem, because it doesn't look so easy."

1987

22 citations

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

Richard P. Feynman

"...it is no damn good at all!"

Learn **1 Sensitivity to High Frequencies.**

Express **2 Only Gaussian Trial States**

Sample **3 We Still Have To Do a Functional Integral**

Variational density matrices as generative models

Learnable unitary transformation generated by point transformation

Learnable probabilistic model for ocuupation probability

$\sqrt{\text{flow}}$

JML '22, SciPost Physics'23

$$\rho = \sum_{\boldsymbol{n}} U |\boldsymbol{n}\rangle p_{\boldsymbol{n}} \langle \boldsymbol{n}| U^\dagger$$

VAN

PRL '19

See Cranmer et al 1904.05903
Saleh et al, 2308.16468
Siciliano et al 2407.03802

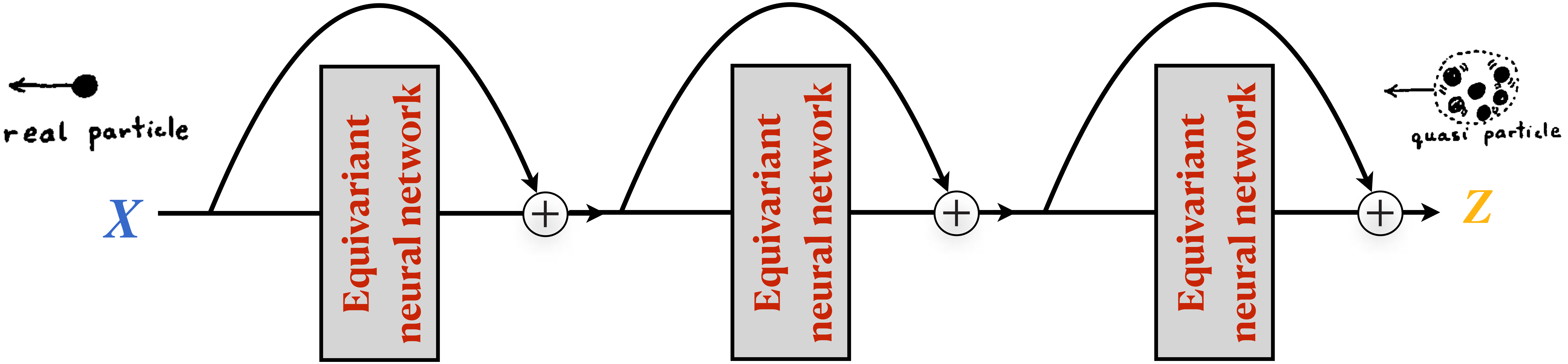
$$\mathrm{Tr}\rho = 1$$
$$\rho \succ 0$$
$$\rho^\dagger = \rho$$

Many-body "base" states e.g. Fermi sea, Hartree-Fock states, harmonic crystal, ...

The physics of $\sqrt{\text{flow}}$

Xie, Zhang, LW, JML '22

$$\langle X | U | \boldsymbol{n} \rangle = \langle Z | \boldsymbol{n} \rangle \cdot \left| \det \left(\frac{\partial Z}{\partial X} \right) \right|^{\frac{1}{2}}$$



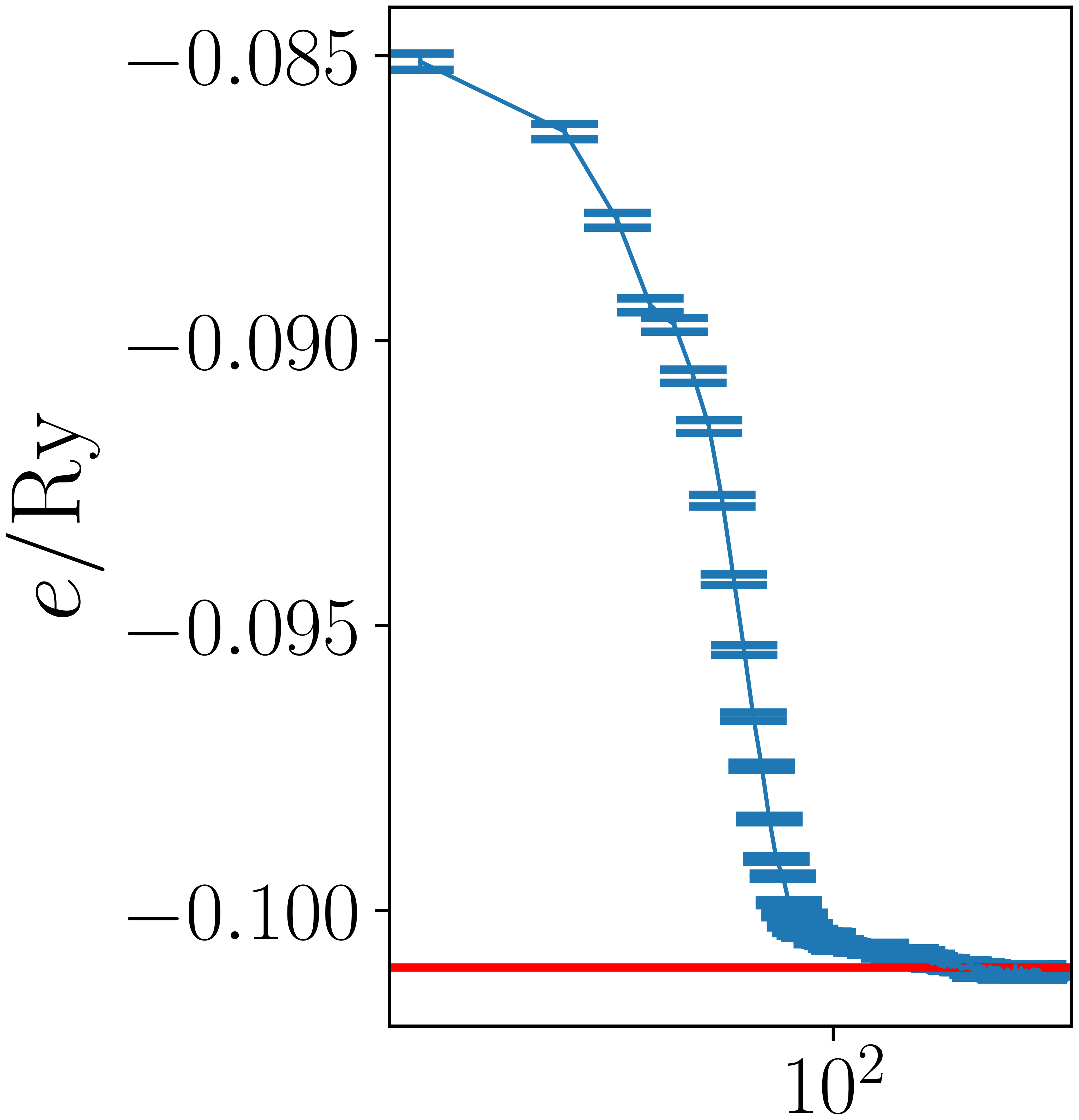
$X \leftrightarrow Z$: unitary backflow between particle and quasiparticle coordinates

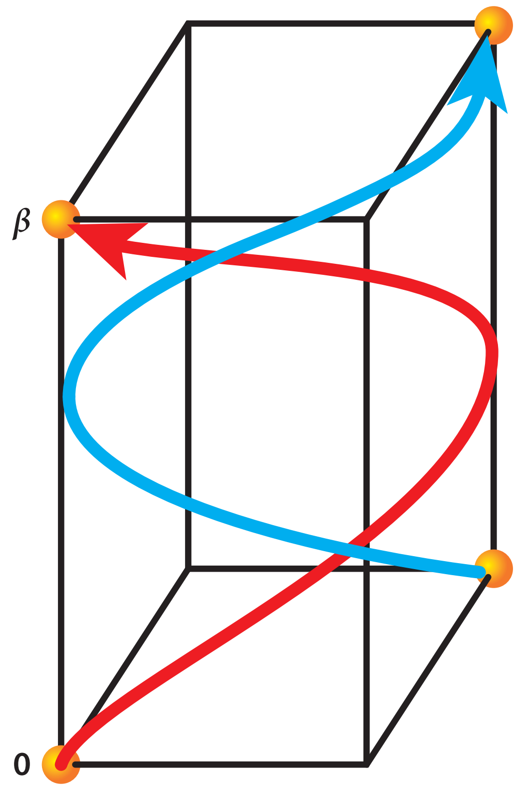
e.g. backflow transformation Feynman & Cohen 1956 $z_i = \boldsymbol{x}_i + \sum_{j \neq i} \eta(|\boldsymbol{x}_i - \boldsymbol{x}_j|)(\boldsymbol{x}_j - \boldsymbol{x}_i)$

Benchmarks on finite-temperature Coulomb gas

Xie, Zhang, LW, SciPost Physics '23

$r_s = 10$, T/T_F=0.0625, N = 33





Severe sign problem

r_s	Θ	$\langle sign \rangle$	E_{tot}^{exact}	E_{tot}
4.0	0.0625	−0.00055(62)	−0.5(1)	−0.1023(7)
10.0	0.0625	−0.002(1)	−0.16(2)	−0.1010(1)

Brown et al, PRL '13 Restricted PIMC

see also Schoof et al PRL '15, Malone et al PRL '16

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

Point Transformations and the Many Body Problem*

M. EGER† AND E. P. GROSS

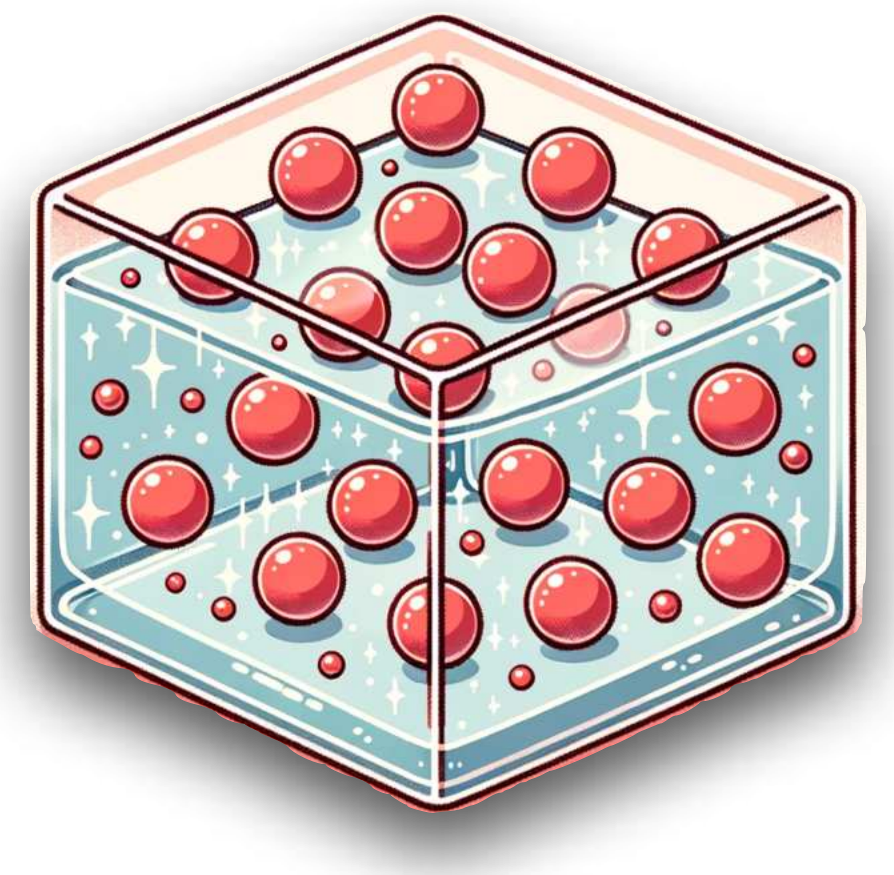
Brandeis University, Waltham, Massachusetts

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

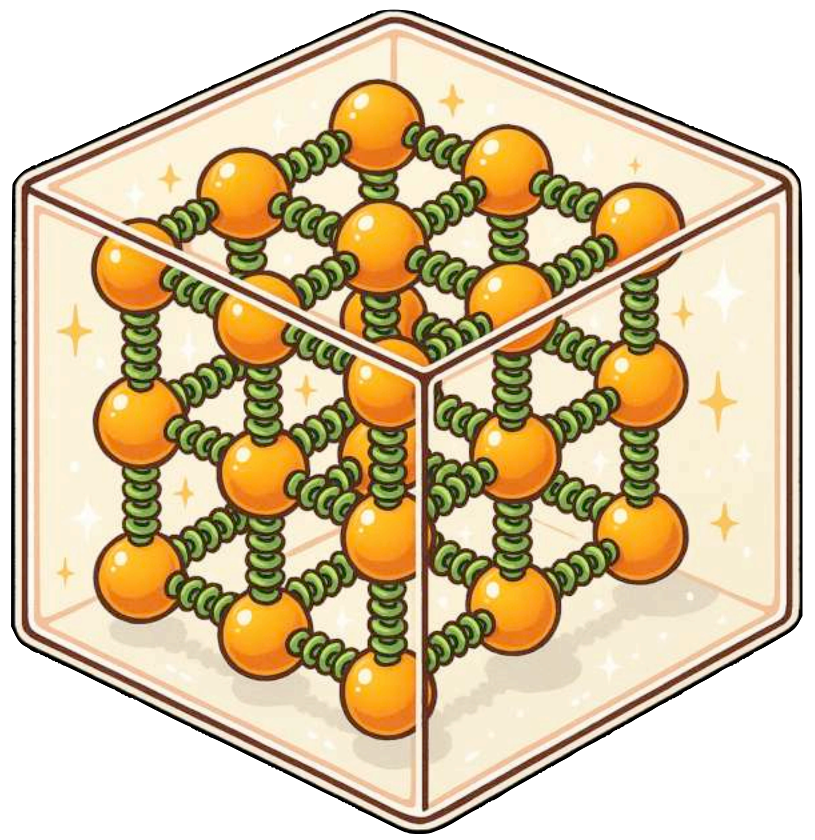
√flow

materializes this dream

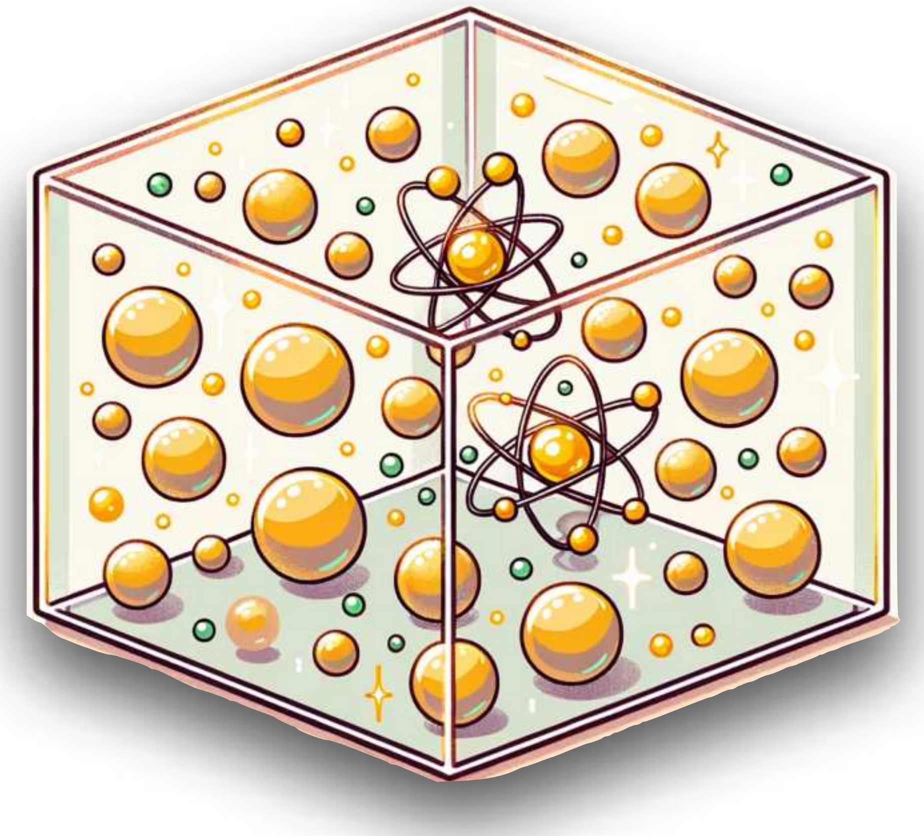
The deep variational free energy approach



Electron liquid
thermodynamics
JML '22 and SciPost Physics '23



Anaharmonic quantum
solid structure
JCP '24 and PRL '25



Dense hydrogen
equaiton of state
PRL '23 and PRB '25

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

Generative AI for It

① $$p(X|y) \propto p(X)p(y|X)$$

Matter inverse design

Exploiting intuitions in data

② $$F[\rho] = E - TS$$

Nature's cost function

Variational free energy is finally practical

Turning physics problems into stochastic optimization

Leverages the deep learning engine