Title: Neural Canonical Transformations

Speakers: Lei Wang

Series: Machine Learning Initiative

Date: January 27, 2023 - 9:00 AM

URL: https://pirsa.org/23010099

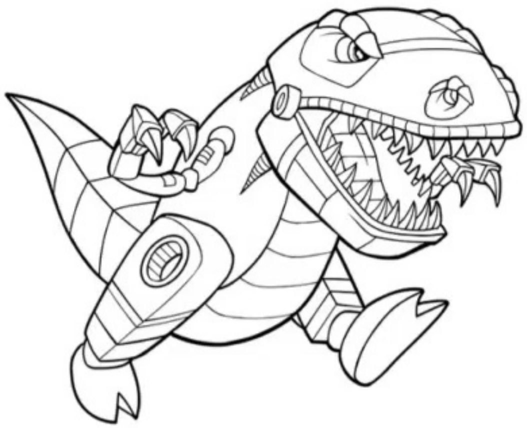
Abstract: Canonical transformations play fundamental roles in simplifying and solving physical systems. However, their design and implementation can be challenging in the many-particle setting. Viewing canonical transformations from the angle of learnable diffeomorphism reveals a fruitful connection to normalizing flows in machine learning. The key issue is then how to impose physical constraints such as symplecticity, unitarity, and permutation equivariance in the flow transformations. In this talk, I will present the design and application of neural canonical transformations for several physical problems. Symplectic flow identifies independent and nonlinear modes of classical Hamiltonians and natural datasets. Fermi flow variationally solves ab initio many-electron problems at finite temperatures.

Refs:

[1] Shuo-Hui Li, Chen-Xiao Dong, Linfeng Zhang, and Lei Wang, Phys. Rev. X 10, 021020 (2020)

[2] Hao Xie, Linfeng Zhang, and Lei Wang, J. Mach. Learn. , 1, 38 (2022)

Zoom link: https://pitp.zoom.us/j/98830940500?pwd=WjdydGY5aS9QQzk5SnI0TE1xMkwrdz09



Neural Canonical Transformations

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

1910.00024, PRX '20
2105.08644, JML '22
2201.03156

li012589/neuralCT
FermiFlow/fermiflow
FermiFlow/CoulombGas

Classical Hamiltonian dynamics

Hamiltonian equations

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

Phase space variables

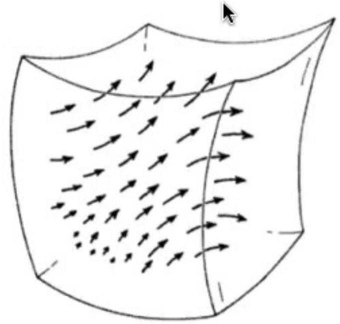
$$\boldsymbol{x} = (p, q)$$

Symplectic metric

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

Symplectic gradient flow

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



Classical Hamiltonian dynamics

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

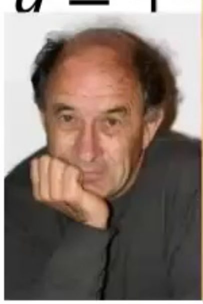
$$= (p, q$$

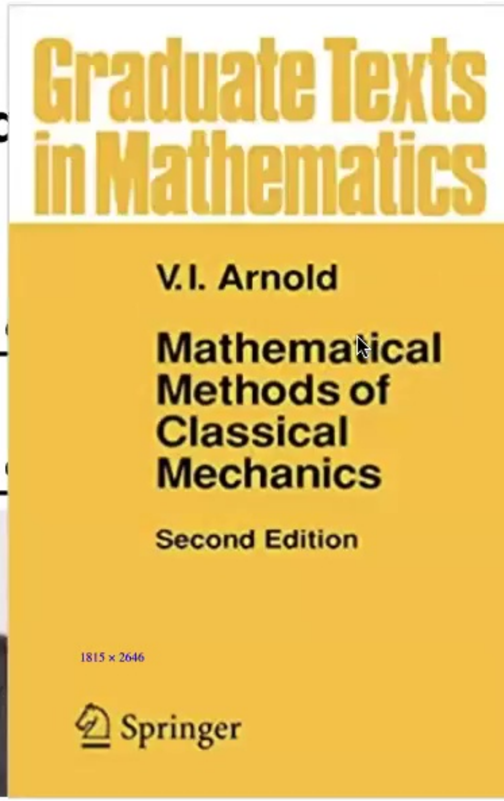
$$\nabla_x H(\boldsymbol{x}) J$$

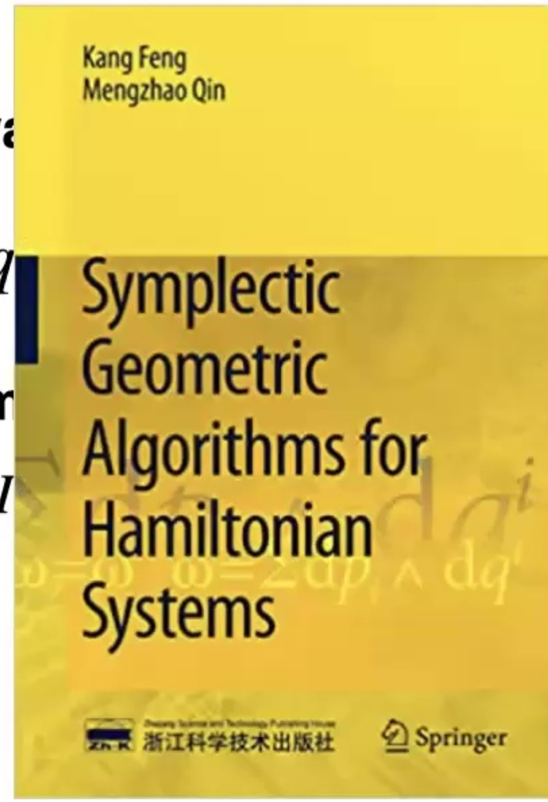
$$\begin{cases} \dot{p} = - \\ \dot{q} = + \end{cases}$$

lectic m

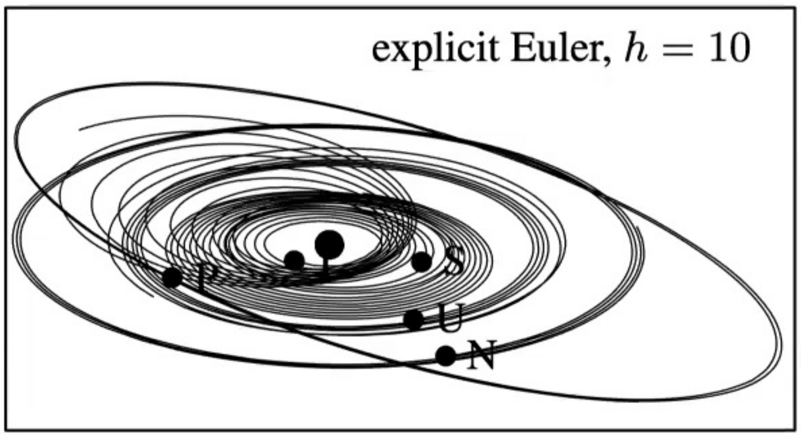
$$\begin{pmatrix} & I \\ -I & \end{pmatrix}$$

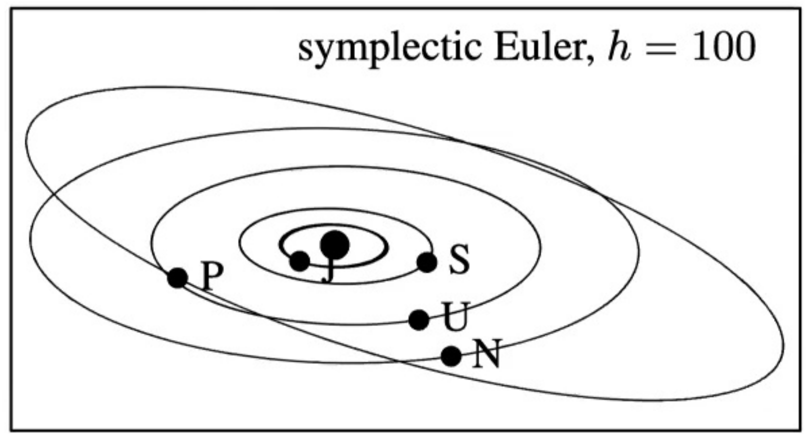






Symplectic Integrators





from Hairer et al, Geometric Numerical Integration

Canonical Transformations

Change of variables

$\boldsymbol{x} = (p, q) \longleftrightarrow \boldsymbol{z} = (P, Q)$

which satisfies $\left(\nabla_{\boldsymbol{x}} \boldsymbol{z}\right) J \left(\nabla_{\boldsymbol{x}} \boldsymbol{z}\right)^T = J$ symplectic condition

Canonical Transformations

Change of variables

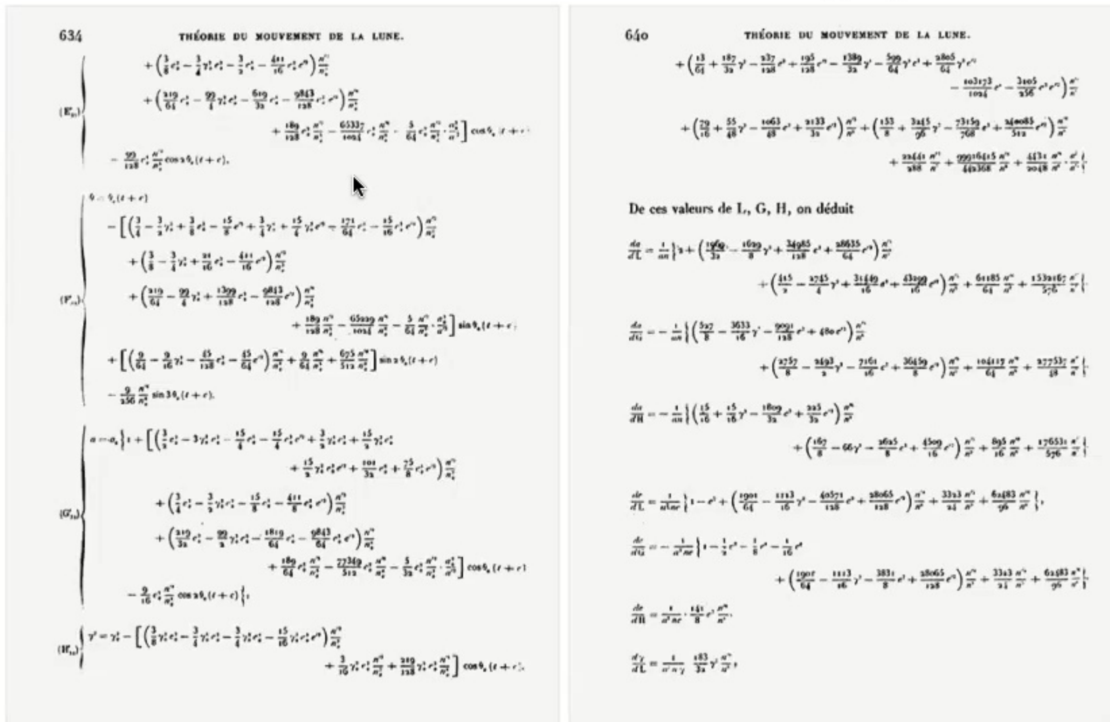
$$\boldsymbol{x} = (p, q) \longleftrightarrow \boldsymbol{z} = (P, Q)$$

which satisfies $\left(\nabla_{\boldsymbol{x}} \boldsymbol{z}\right) J \left(\nabla_{\boldsymbol{x}} \boldsymbol{z}\right)^T = J$ symplectic condition

one has $\dot{\boldsymbol{z}} = \nabla_{\boldsymbol{z}} K(\boldsymbol{z}) J$ where $K(\boldsymbol{z}) = H \circ \boldsymbol{x}(\boldsymbol{z})$

Preserves Hamiltonian dynamics in the "latent phase space"

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 canonical transformations for more complex systems?

A probabilisitic perspective

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

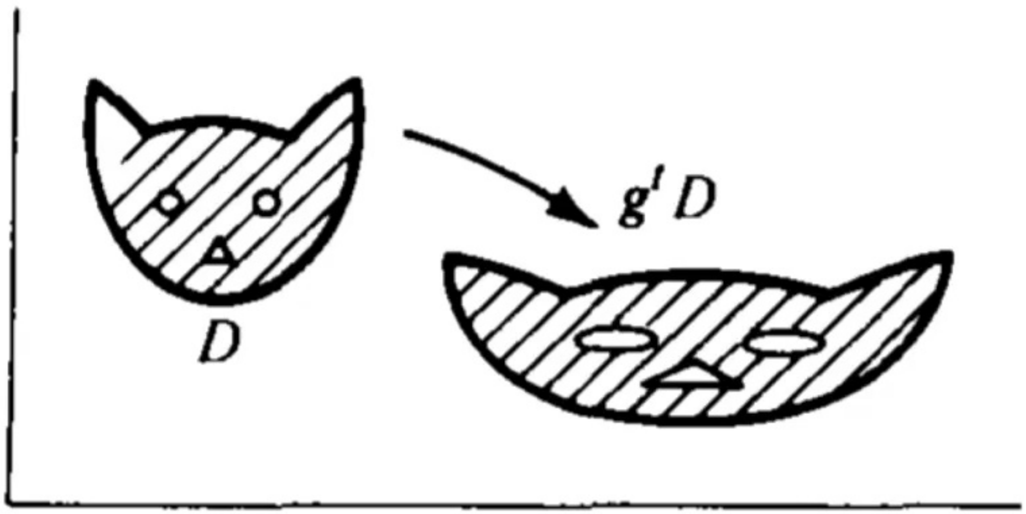
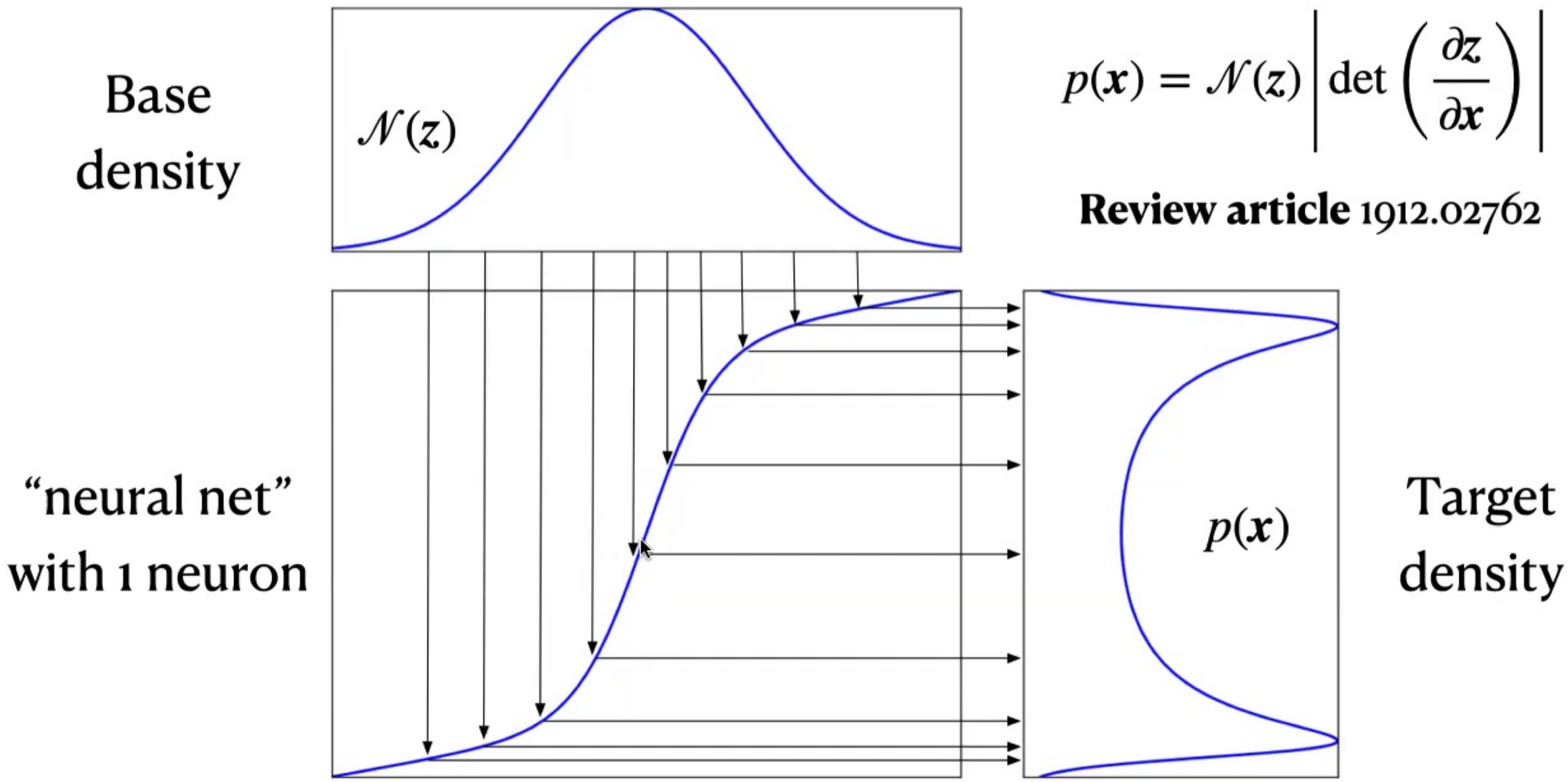


Image from Arnold '78

Neural networks are good at transforming probability densities

Normalizing flow in a nutshell



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

Review article 1912.02762

Continuous normalizing flow

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

Consider infinitesimal change-of-variables Chen et al 1806.07366

$$\boldsymbol{x} = z + \varepsilon \boldsymbol{v}$$

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

$\varepsilon \to 0$

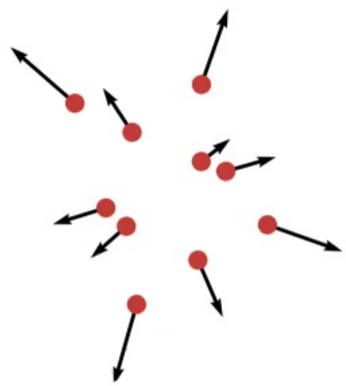
$$\frac{d\boldsymbol{x}}{dt} = \boldsymbol{v}$$

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

Fluid physics behind flows

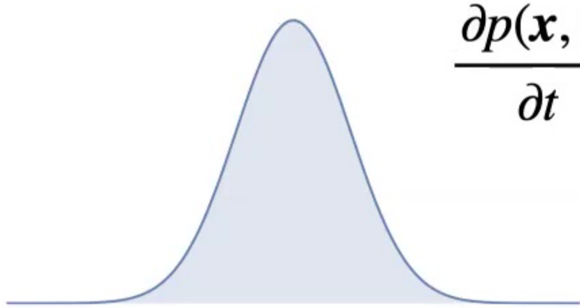
Zhang, E, LW 1809.10188

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

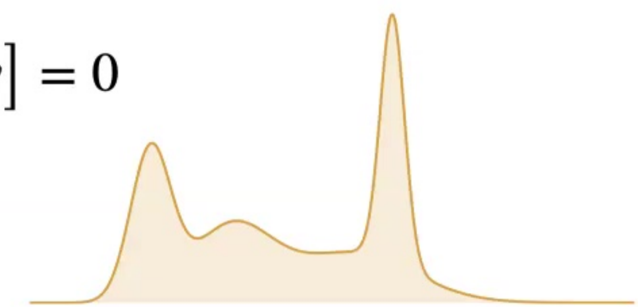


$$\frac{d}{dt} = \frac{\partial}{\partial t} + \boldsymbol{v} \cdot \nabla$$ "material derivative"

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



Simple density

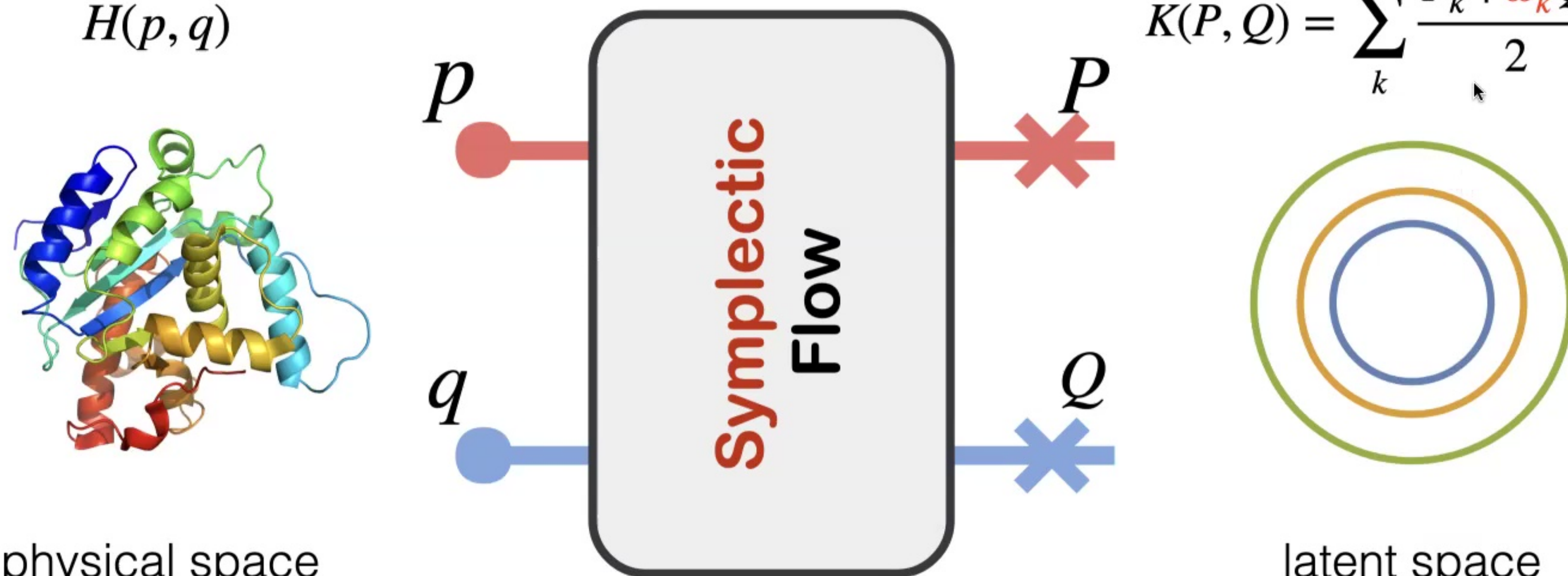


Complex density

Neural Canonical Transformations

$H(p, q)$

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



physical space

latent space

Learn the network parameter and the latent harmonic frequency

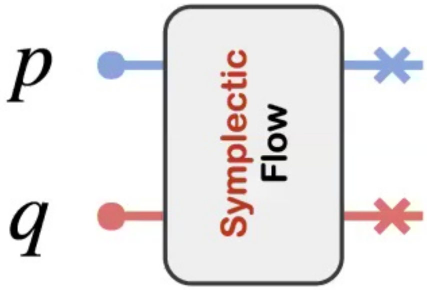
Symplectic primitives

- Linear transformation: Symplectic Lie algebra
- Continuous-time flow: Symplectic generating functions

Symplectic integrator as a “layer” in deep net

Harbor el al 1705.03341
Lu et al 1710.10121
E, Commun. Math. Stat 17
Chen et al 1806.07366
Zhu et al, 2004.13830
...

- **Neural point transformation**



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

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

Modular design of the symplectic network

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

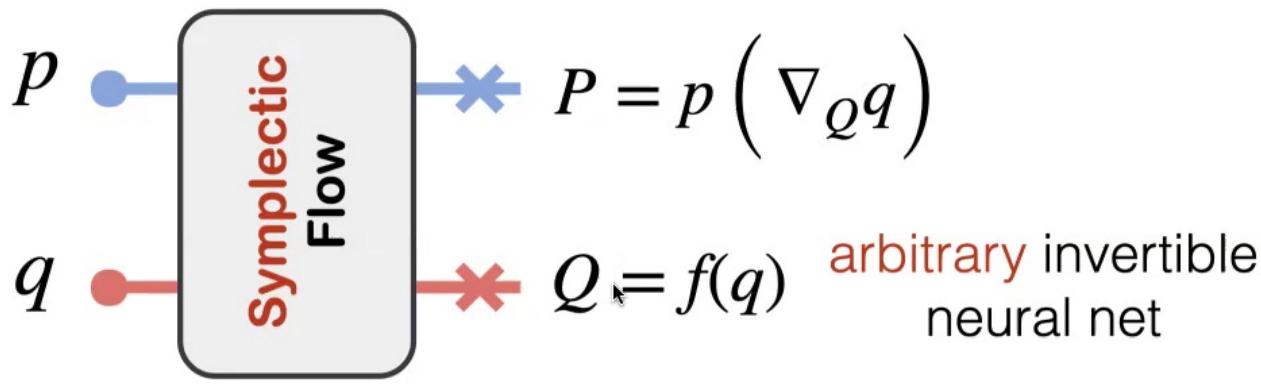
$$\mathcal{T} = \mathcal{T}_1 \circ \mathcal{T}_2 \circ \mathcal{T}_3 \circ \cdots$$

$$\left(\nabla_x z\right) J \left(\nabla_x z\right)^T = J$$

symplectic group

Compose symplectic primitives to form a deep neural network

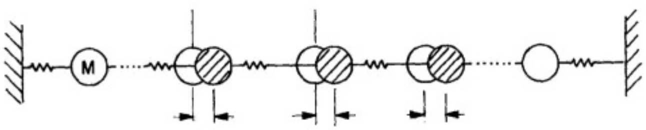
Neural point transformation

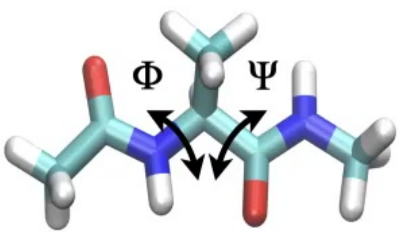



```
Q = forward(params, q)
_, vjp = jax.vjp(lambda Q: inverse(params, Q), Q)
P = vjp(p)[0]
```

How is this going to be useful ?

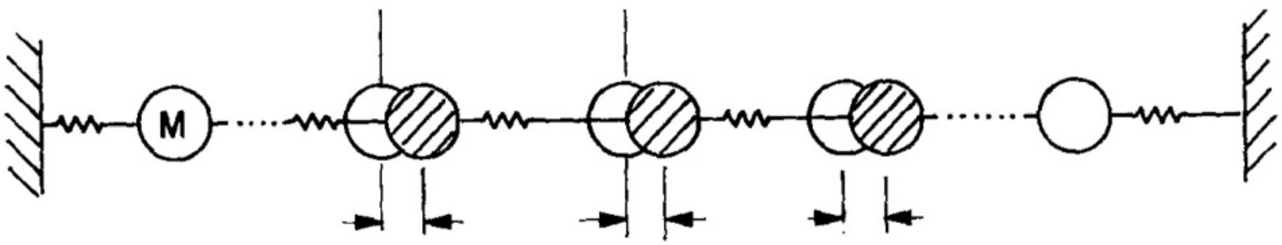
Let's play with examples!





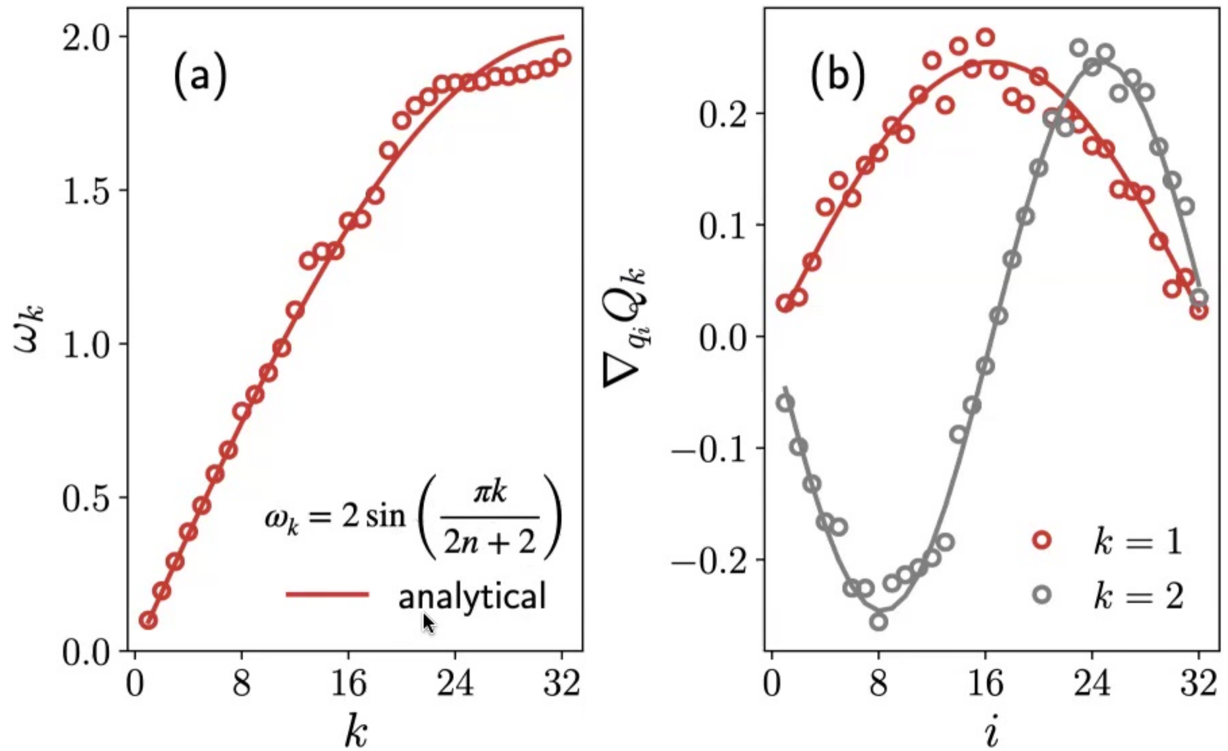
Warm up: Harmonic Chain

$$H = \frac{1}{2}\sum_{i=1}^{n}\left[p_i^2 + (q_i - q_{i-1})^2\right]$$



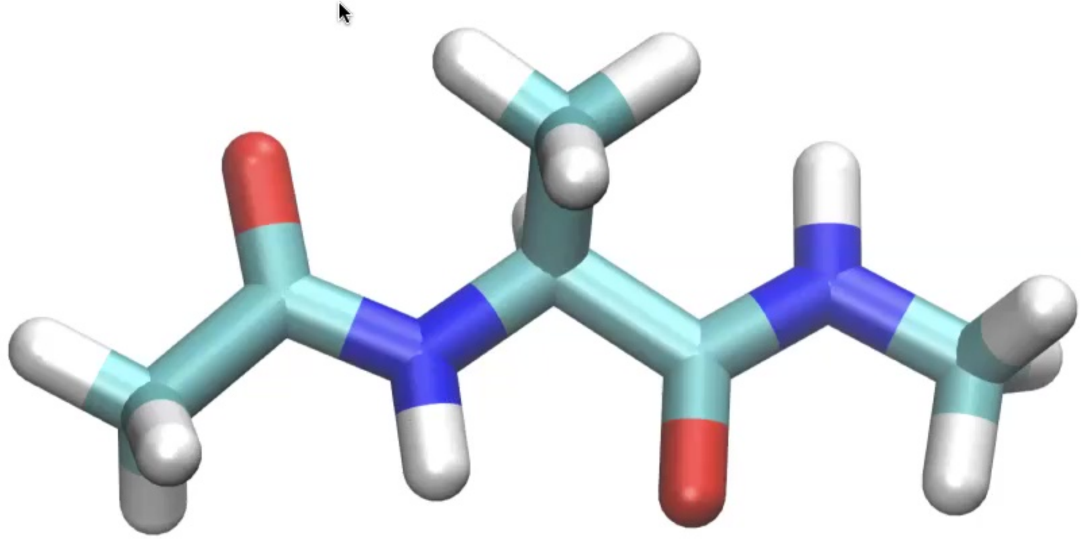
Fermi–Pasta–Ulam–Tsingou problem w/o nonlinearity

Learning the normal modes



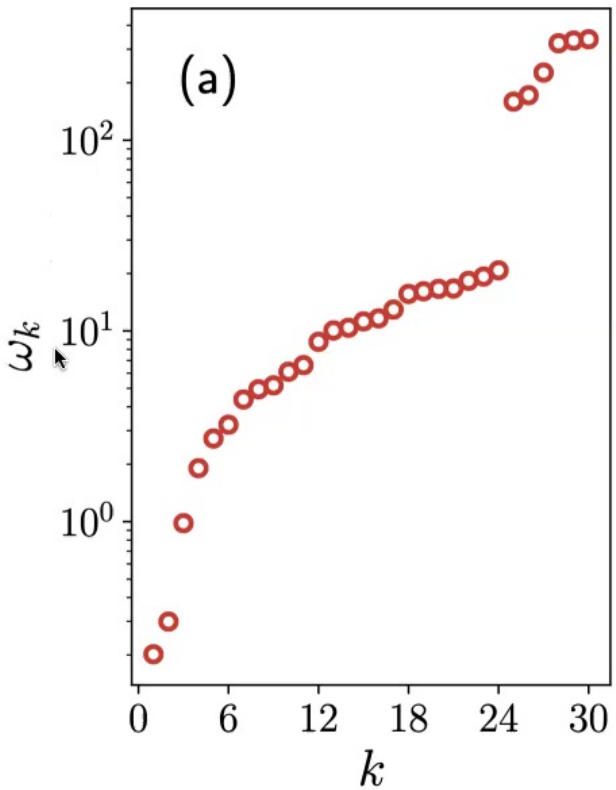
Consistency check: neural nets can learn linear coordinate transformations

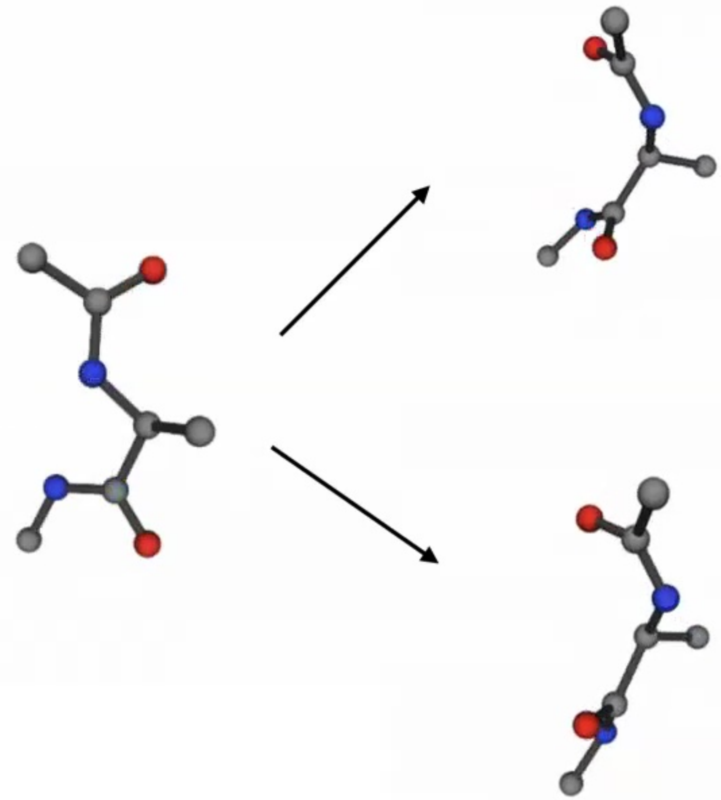
Alanine Dipeptide



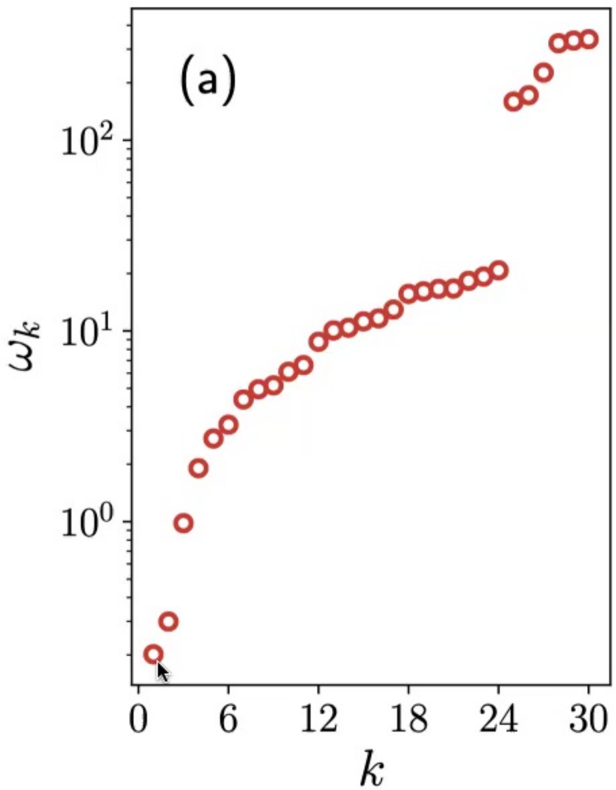
250 ns molecular dynamics simulation data at 300 K

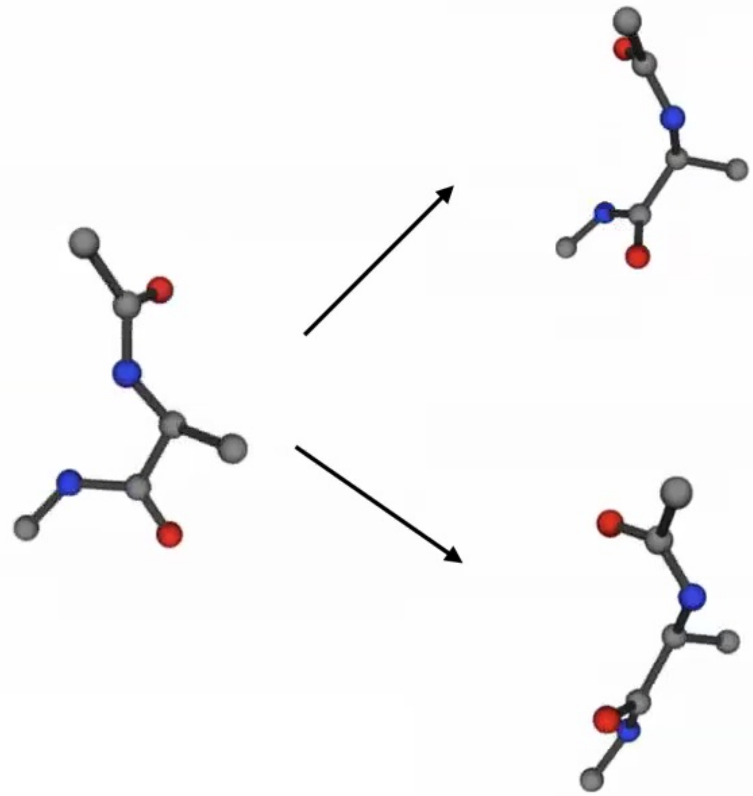
https://markovmodel.github.io/mdshare/ALA2/#alanine-dipeptide



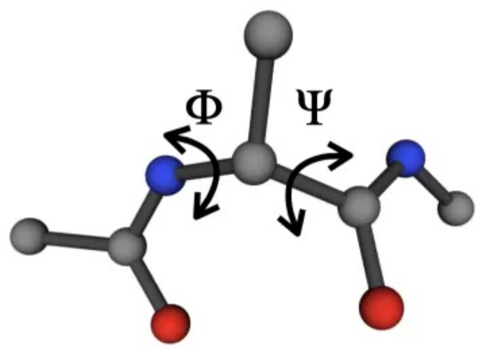


Neural canonical transformation identifies nonlinear slow modes!

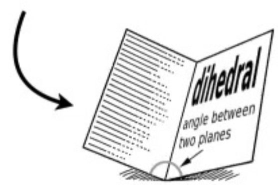


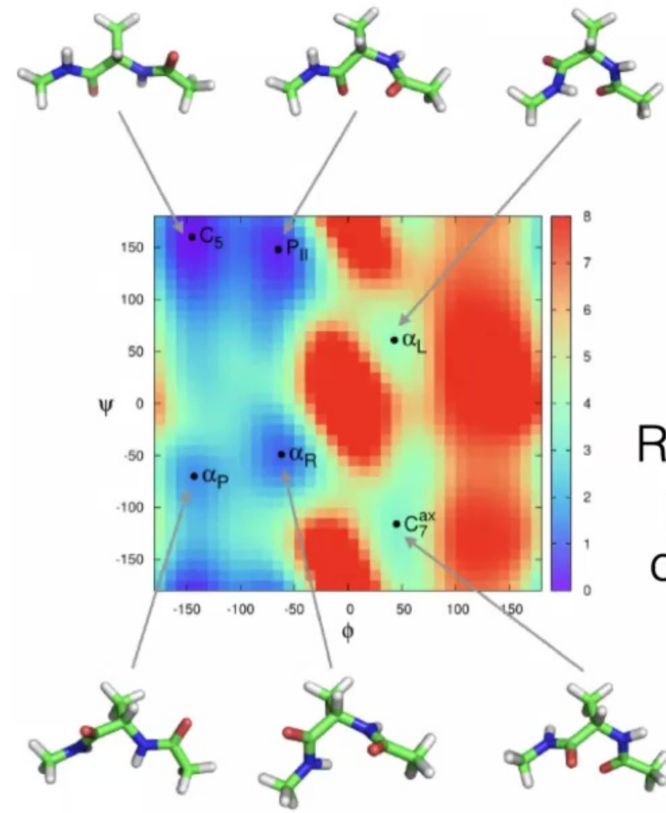


Neural canonical transformation identifies nonlinear slow modes!



slow motion of the two torsion angles

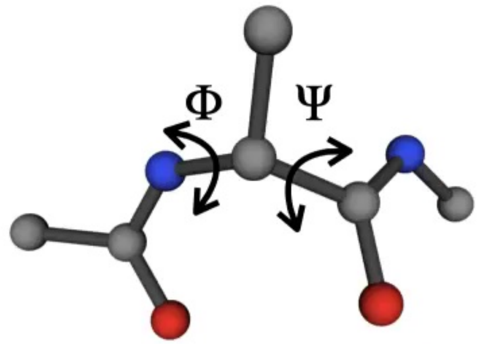




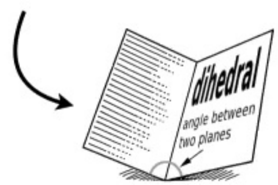
Ramachandran plot of stable conformations

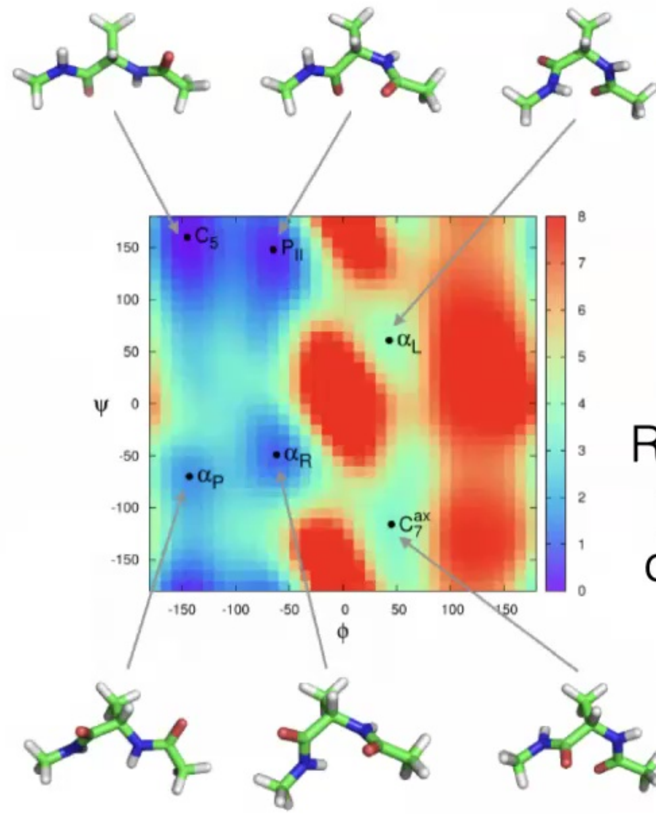
Dimensional reduction to slow collective variables useful for control, prediction, enhanced sampling…

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



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

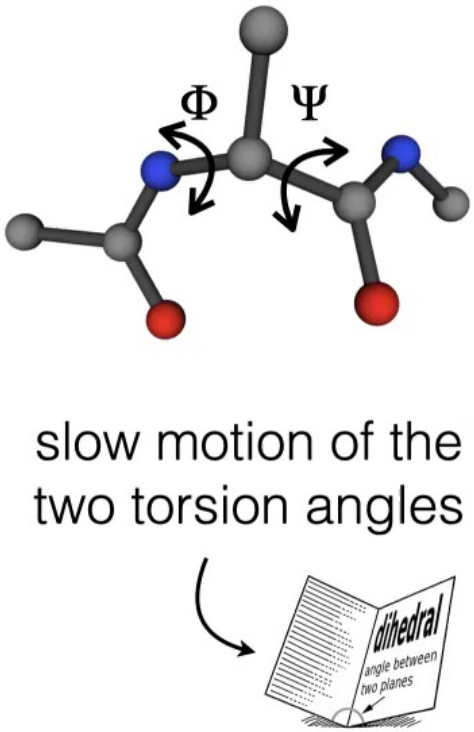
MNIST handwritten digits

Data scientists:

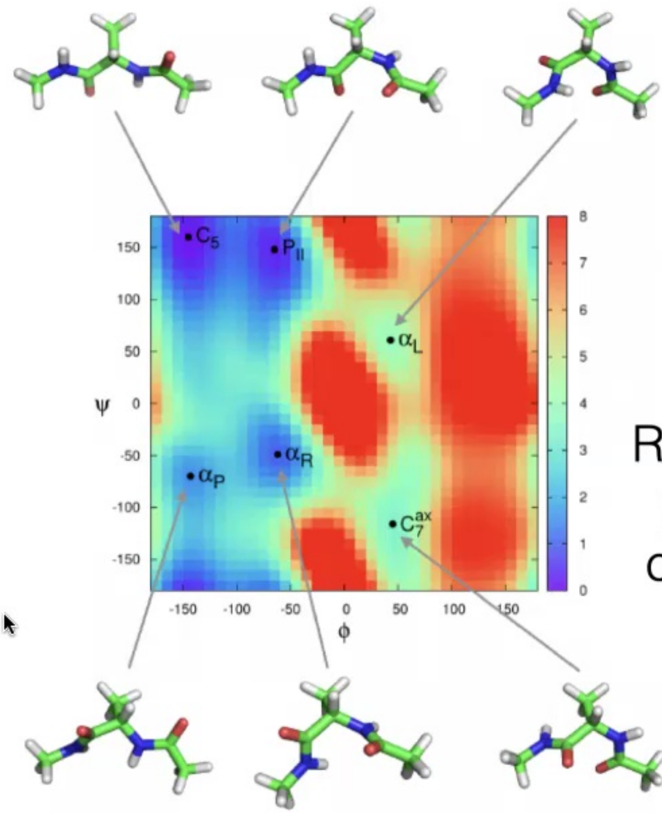
“50,000 grayscale images with 28x28 pixels”

Physical Chemists:

“Stable conformations of a molecule with 784 degrees of freedom”



slow motion of the two torsion angles



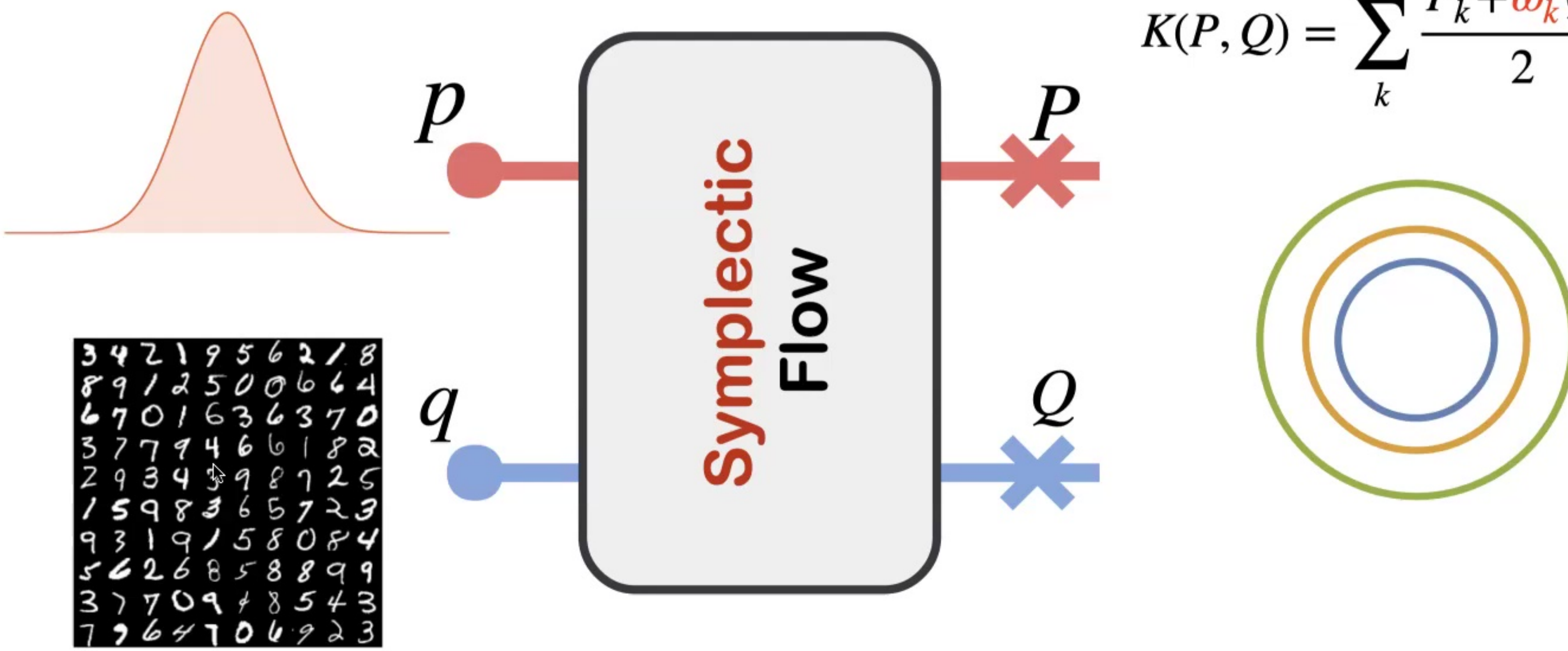
Ramachandran plot of stable conformations

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

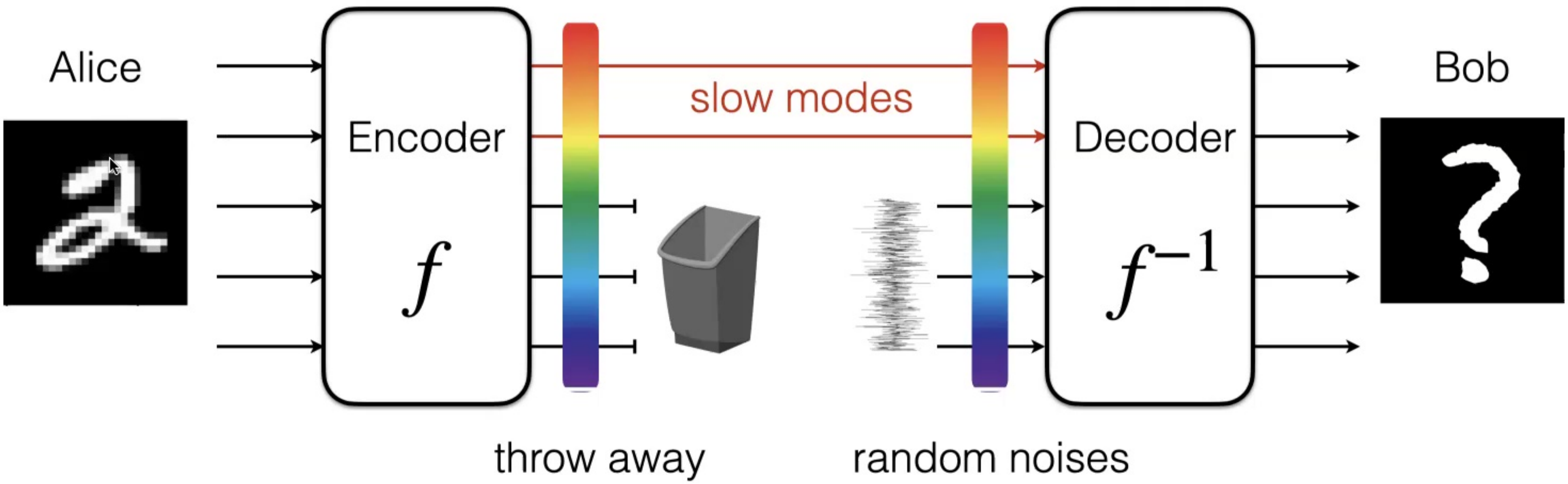
check the paper 1910.00024, PRX '20 for more examples & applications

Learning slow variables of MNIST

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



Neural Alice-Bob game



Conceptual Compression of MNIST

Original

1/784 kept

"A Hamiltonian Extravaganza"

—Danilo J. Rezende@DeepMind

Symplectic ODE-Net, 1909.12077

Hamiltonian Graph Networks with ODE Integrators, 1909.12790

Symplectic RNN, 1909.13334

Equivariant Hamiltonian Flows, 1909.13739

Hamiltonian Generative Network, 1909.13789 http://tiny.cc/hgn

Neural Canonical Transformation with Symplectic Flows, 1910.00024

See also Bondesan & Lamacraft, *Learning Symmetries of Classical Integrable Systems*,1906.04645

“A Hamiltonian Extravaganza”

—Danilo J. Rezende@DeepMind

Symplectic ODE-Net, 1909.12077

Hamiltonian Graph Networks with ODE Integrators, 1909.12790

Symplectic RNN, 1909.13334

Equivariant Hamiltonian Flows, 1909.13739

Hamiltonian Generative Network, 1909.13789 http://tiny.cc/hgn

Neural Canonical Transformation with Symplectic Flows, 1910.00024

See also Bondesan & Lamacraft, *Learning Symmetries of Classical Integrable Systems*,1906.04645

How about quantum systems ?

Neural canonical transformations

"Symplectic" "Unitary"

Classical world

$$(\dot{\boldsymbol{x}}, \dot{\boldsymbol{p}}) = \nabla_{(\boldsymbol{x},\boldsymbol{p})} G_\theta(\boldsymbol{x}, \boldsymbol{p}) \begin{pmatrix} & -\mathbb{I} \\ \mathbb{I} & \end{pmatrix}$$

point transformation

$$G_\theta(\boldsymbol{x}, \boldsymbol{p}) = \boldsymbol{v}_\theta(\boldsymbol{x}) \cdot \boldsymbol{p}$$

Neural canonical transformations, Li, Dong, Zhang, LW, PRX '20

Quantum world

$$(\dot{\boldsymbol{x}}, \dot{\boldsymbol{p}}) = i[(\boldsymbol{x}, \boldsymbol{p}), G_\theta(\boldsymbol{x}, \boldsymbol{p})]$$

"quantized" point transformation

$$G_\theta(\boldsymbol{x}, \boldsymbol{p}) = \frac{1}{2}\{\boldsymbol{v}_\theta(\boldsymbol{x}), \boldsymbol{p}\}$$ DeWitt '52

Quantum generalization to fermions, Xie, Zhang, LW, 2105.08644

see also Cranmer et al, 1904.05903

Even more correspondences

Classical world

Probability density p

Kullback-Leibler divergence

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

Variational free-energy

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

Quantum world

Density matrix ρ

Quantum relative entropy

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

Variational free-energy

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

Neural unitary transformations

Xie, Zhang, LW, 2105.08644, JML '22

$$\rho_{\boldsymbol{\theta},\boldsymbol{\phi}} = e^{-iG_{\boldsymbol{\theta}}t} \left[\sum_n p_{\boldsymbol{\phi}}(n) \,|\, \Psi_n \rangle \langle \Psi_n \,| \right] e^{iG_{\boldsymbol{\theta}}t}$$

the "base" density matrix

$$i\frac{\partial}{\partial t}|\Psi_n\rangle = G_{\boldsymbol{\theta}}|\Psi_n\rangle$$

$$G_{\boldsymbol{\theta}} = \frac{-i}{2}\{\boldsymbol{v}_{\boldsymbol{\theta}}, \nabla\}$$

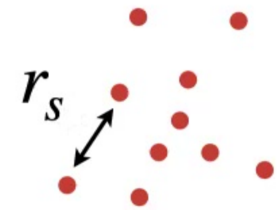
$$\frac{\partial \left|\Psi_n(\boldsymbol{x},t)\right|^2}{\partial t} + \nabla \cdot \left(|\Psi_n(\boldsymbol{x},t)|^2 \boldsymbol{v}_{\boldsymbol{\theta}} \right) = 0$$

Continuous normalizing flow

Variational density matrix with tractable entropy $\mathrm{Tr}(\rho \ln \rho) = \sum_n p \ln p$

Application: Uniform electron gas

$$H = -\sum_{i=1}^{N} \frac{\hbar^2 \nabla_i^2}{2m} + \sum_{i<j} \frac{e^2}{|\boldsymbol{r}_i - \boldsymbol{r}_j|}$$



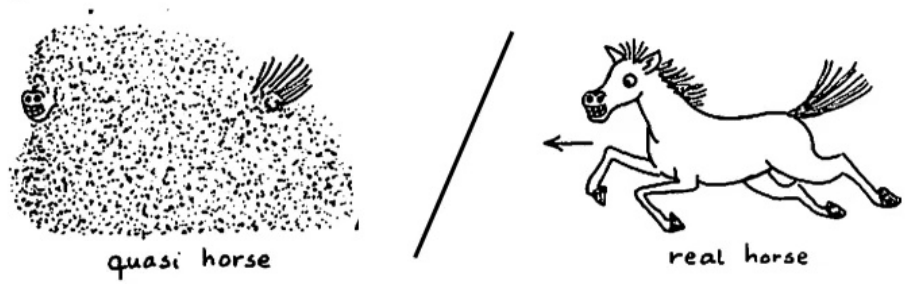
Fundamental model
for metals ($2 < r_s < 6$)

$$E_c^{\mathrm{PBE}}[n] = \int d^3r\, n(\epsilon_c^{\mathrm{ueg}} + \cdots)$$

Input to the density
functional theory calculations

Quasi-particles effective mass

$$\frac{m^*}{m} = $$



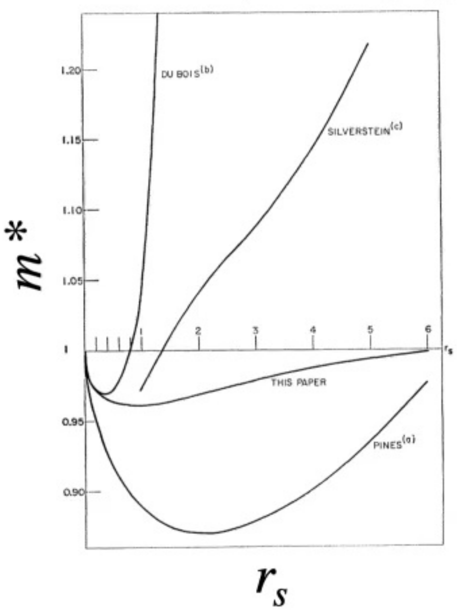
Richard D. Mattuck
A Guide to Feynman Diagrams in the Many-body Problem

A fundamental quantity appears in nearly all physical properties of a Fermi liquid

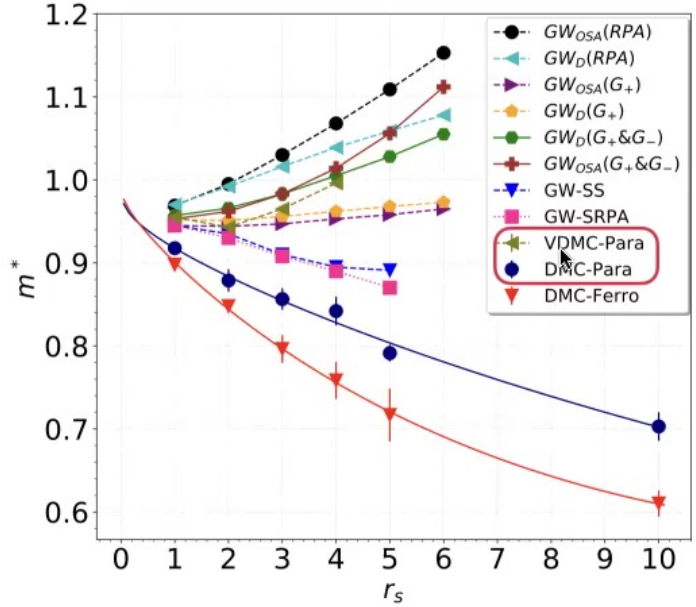
$N(0)$	S	c_V	χ
Density of states	entropy	specific heat	magnetic susceptibility

Quasi-particles effective mass of 3d electron gas

Hedin Phy. Rev. 1965



Azadi, Drummond, Foulkes, PRL 2021



> 50 years of conflicting results !

Two-dimensional electron gas experiments

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

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

A. A. Shashkin,* Maryam Rahimi, S. Anissimova, and S.V. Kravchenko
Physics Department, Northeastern University, Boston, Massachusetts 02115, USA

V.T. Dolgopolov
Institute of Solid State Physics, Chernogolovka, Moscow District 142432, Russia

T. M. Klapwijk
Department of Applied Physics, Delft University of Technology, 2628 CJ Delft, The Netherlands
(Received 13 January 2003; published 24 July 2003)

$m^*/m > 1$

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

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

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

$m^*/m < 1$

Layer thickness, valley, disorder, spin-orbit coupling...

m^* from low temperature entropy

Eich, Holzmann, Vignale, PRB '17

$$s = \frac{\pi^2 k_B}{3} \frac{m^*}{m} \frac{T}{T_F}$$

~~$m^*/m = \left(1 - \frac{\partial \Sigma}{\partial \omega}\right)\left(1 + \frac{m}{k}\frac{\partial \Sigma}{\partial k}\right)^{-1}$~~

~~$m^*/m = k_F/(d\varepsilon/dk)_{k_F}$~~

$$\Rightarrow \frac{m^*}{m} = \frac{s}{s_0}$$

s ← interacting electrons

s_0 ← noninteracting electrons

Not an easy task due to the lack of reliable methods
for interacting electrons at low-temperature with intermediate density

The variational free energy approach

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

Gibbs–Bogolyubov-Feynman-Delbrück–Molière

$$\text{s.t.} \quad \mathrm{Tr}\rho = 1 \qquad \rho \succ 0 \qquad \rho^\dagger = \rho \qquad \langle \boldsymbol{x} | \rho | \boldsymbol{x}' \rangle = (-)^{\mathscr{P}} \langle \mathscr{P}\boldsymbol{x} | \rho | \boldsymbol{x}' \rangle$$

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

Normalized probability distribution

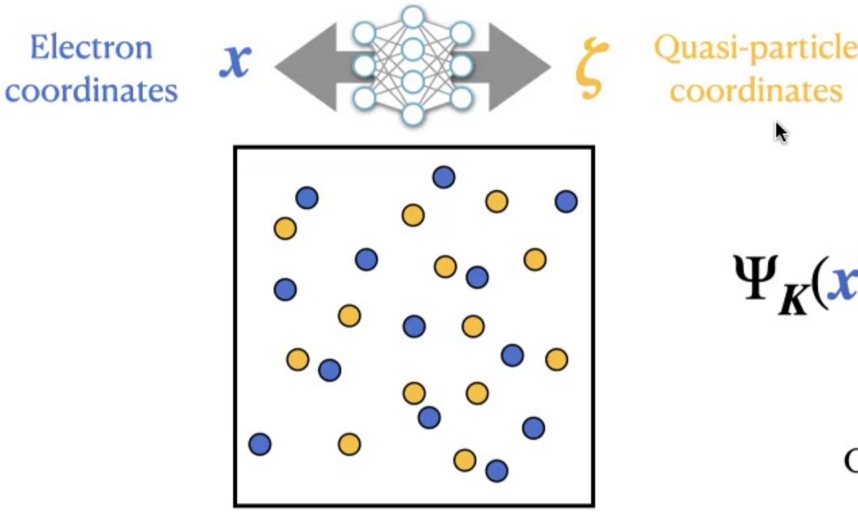
Orthonormal many-electron basis

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

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

There will also be interesting twists for physics considerations

② $\sqrt{\text{Normalizing flow}}$ for $|\Psi_{\boldsymbol{K}}\rangle$



$$\Psi_{\boldsymbol{K}}(x) = \frac{\det(e^{i\boldsymbol{k}_i \cdot \zeta_j})}{\sqrt{N!}} \cdot \left| \det\left(\frac{\partial \zeta}{\partial x}\right) \right|^{\frac{1}{2}}$$

Orthonormal many-body states

Jacobian of the transformation

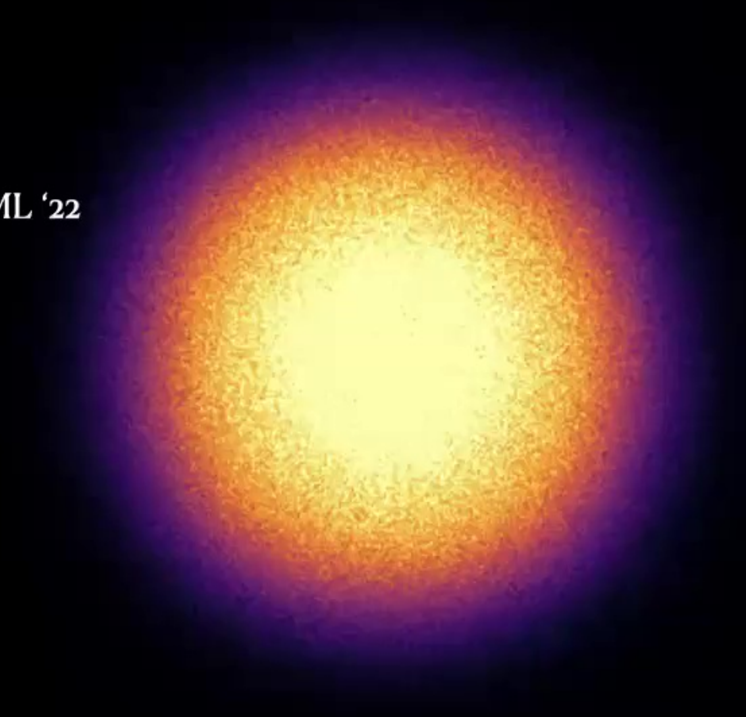
Twist: the flow should be permutation equivariant for fermionic coordinates

we use FermiNet layer Pfau et al, 1909.02487

Fermi Flow

Xie, Zhang, LW, 2105.08644, JML '22

github.com/fermiflow



Continuous flow of electron density in a quantum dot

FAQs

Where to get training data ?

No training data. Data are self-generated from the generative model.

How do we know it is correct ?

Variational principle: lower free-energy is better.

Do I understand the "black box" model ?

a) I don't care (as long as it is sufficiently accurate).

b) $\ln p(\boldsymbol{K})$ contains the Landau energy functional

$$E[\delta n_k] = E_0 + \sum_k \epsilon_k \delta n_k + \frac{1}{2} \sum_{k,k'} f_{k,k'} \delta n_k \delta n_{k'}$$

$\zeta \leftrightarrow x$ illustrates adiabatic continuity.

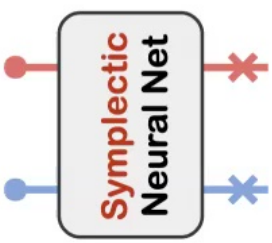
Training approaches

Variational calculation

"learn from Hamiltonian"

$$\mathscr{L} = \int d\boldsymbol{x}\, p(\boldsymbol{x}) \left[\ln p(\boldsymbol{x}) + \beta H(\boldsymbol{x})\right]$$

$\boldsymbol{x}$



$z \sim \mathcal{N}(\mathbf{0}, \Sigma)$

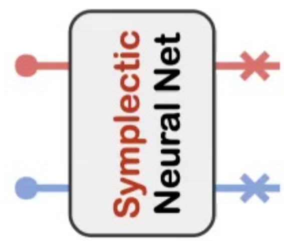
Sample in the latent space

Density estimation

"learn from data"

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

$\boldsymbol{x} \sim \text{dataset}$



z

Sample from dataset in the physical space