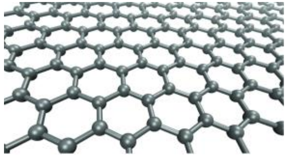
Generative models for molecules in equillibrium

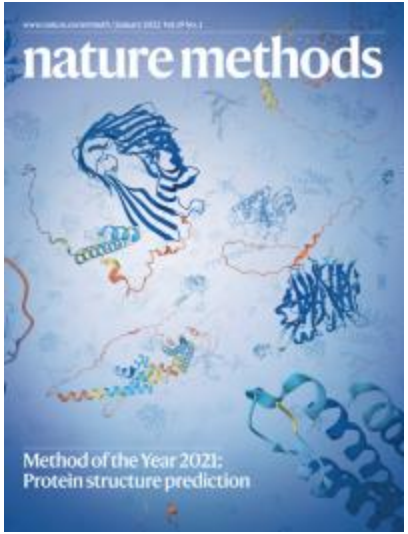
Jonas Köhler

Why generative modeling for molecules?

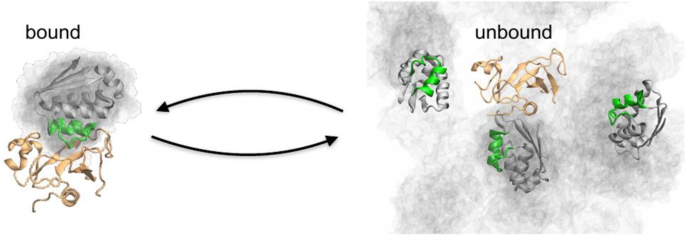
find candidates for drugs and materials (inverse design)



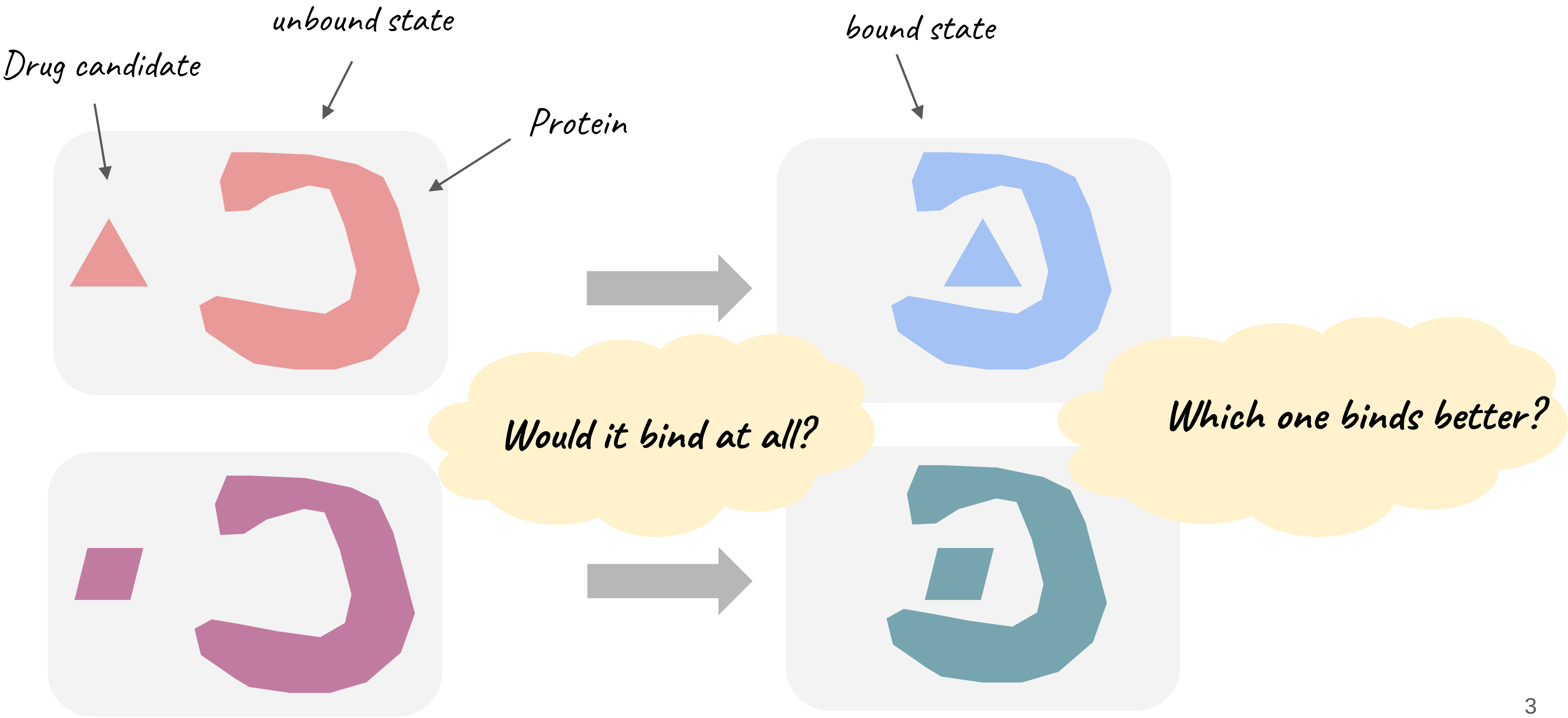
advance science



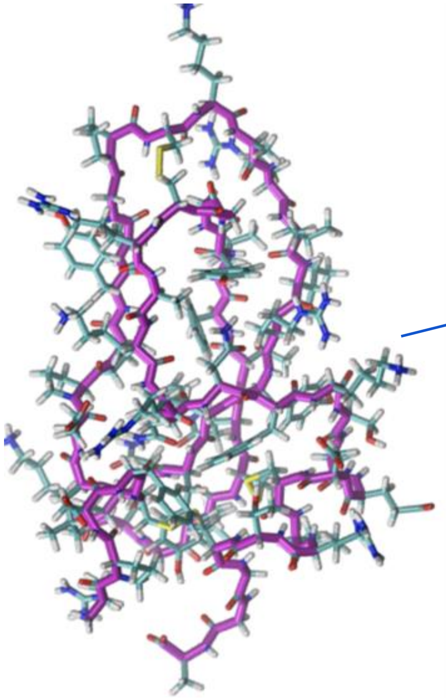
understand molecular origin of diseases



Some Motivation: binding affinity prediction

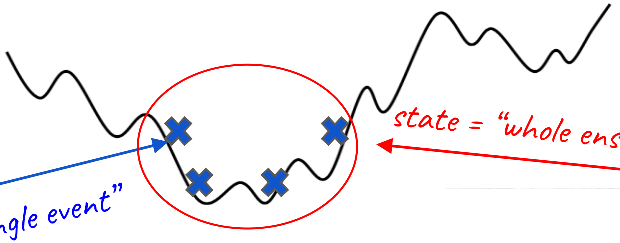


Molecules are not static…



Potential energy

$$u(\mathrm{x})$$

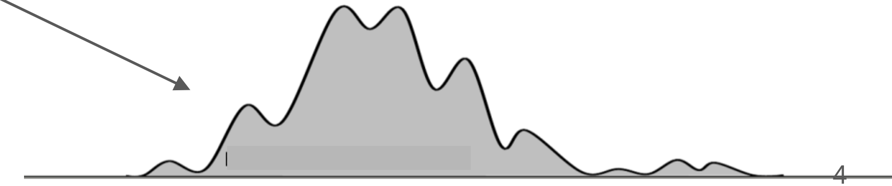


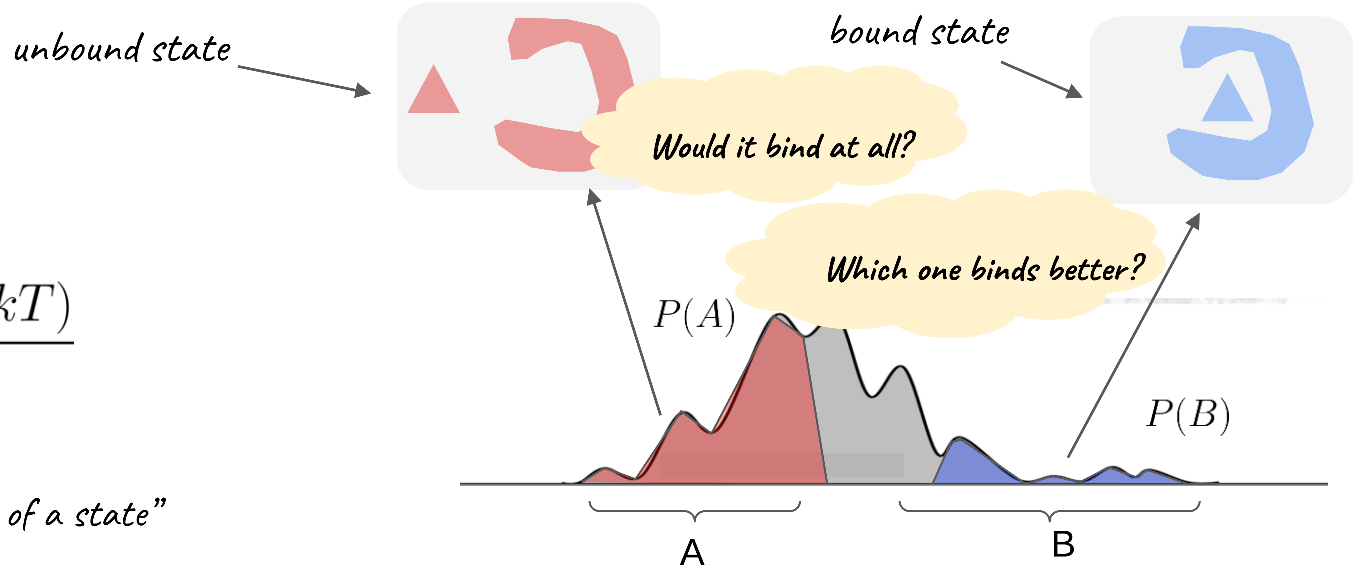
structure = "single event"

state = "whole ensemble"

Boltzmann density

$$\mu(\mathrm{x}) = \frac{\exp(-u(\mathrm{x})/kT)}{Z}$$





Boltzmann density

$$\mu(\mathrm{x}) = \frac{\exp(-u(\mathrm{x})/kT)}{Z}$$

"likelihood of a state"

Partition function

$$Z_A = P(A) = \int_{A \subset \mathbb{R}^d} \mu(\mathbf{x})$$

"log likelihood ratio"

Free energy difference

$$\Delta F_{A \to B} = F_A - F_B$$

"neg. log likelihood. of state"

Free energy

$$F_A = -kT \log Z_A$$

Computing FED requires sampling...

$$\mathbf{x} \sim \exp(-u(\mathbf{x}))/Z$$

:-(

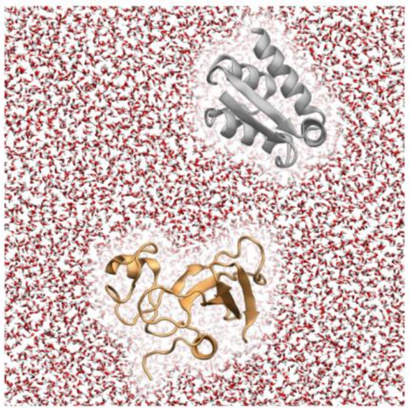
Answers requires sampling...

$$\mathbf{x} \sim \exp(-u(\mathbf{x}))/Z$$

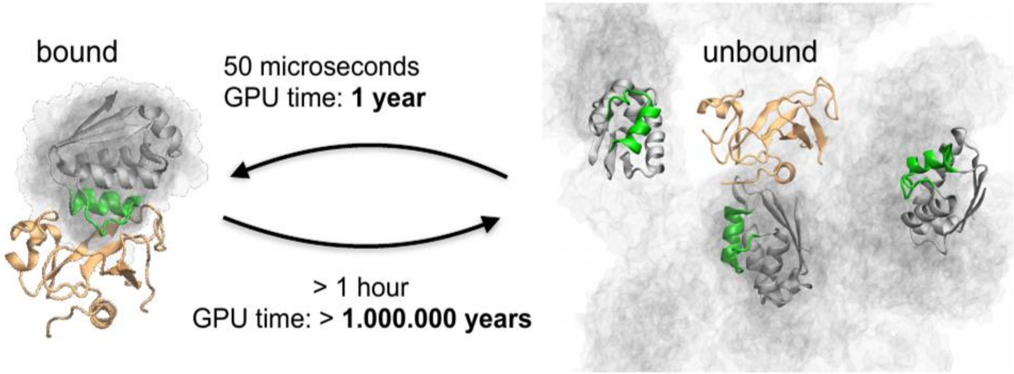
:-(

Classic workhorse:
Molecular / Langevin dynamics simulations

$$\mathbf{x} \leftarrow \mathbf{x} - \nabla_{\mathbf{x}} u(\mathbf{x}) dt + \sqrt{2dt}\, \eta, \quad \eta \sim \mathcal{N}(0, I)$$



easy to make mistakes...

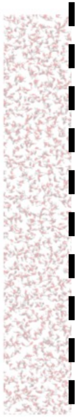


Numerical precision: step size 1-4 fs

Relevant biological scales: 1 ms → hours...

Computing FED requires sampling...

Classic workhorse: Molecular / Langevin dynamics simulations

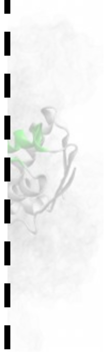


2ms of molecular dynamics

= ~1 Ph.D.

= ~ 500 GJ

3 ✖  = 



Nu

Source: Frank Noé

Relevant biological scales: 1 ms → hours...

Boltzmann Generators. Noé*, Olsson*, JK*, Wu. Science. 2019

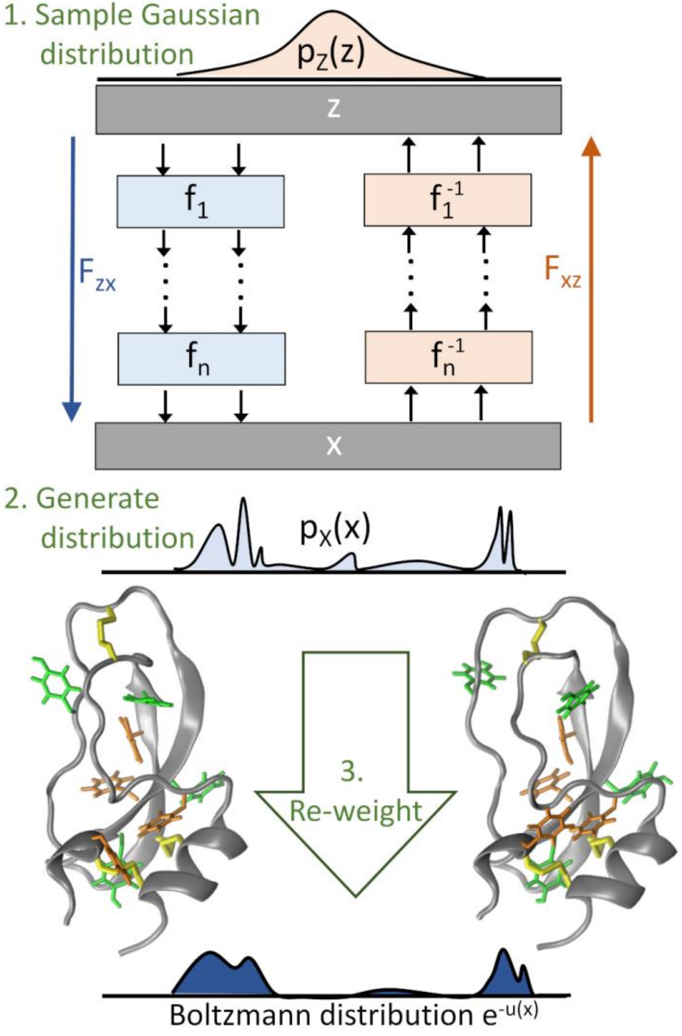
Boltzmann Generators

Frank Noé

Simon Olsson

Hao Wu

1. Sample noise from base distribution
2. Transform via a trainable diffeomorphism (Normalizing Flow)
3. Reweigh against the target



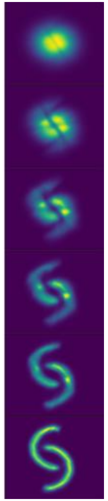
Quick Recap: Normalizing Flows[1]

base density

$$z \sim q(z)$$

diffeomorphism

f



$$x = f(z;\theta) \sim p(x;\theta) \quad \approx \mu(x)$$

target density

$$\mu(\mathrm{x}) = \frac{\exp(-u(\mathrm{x})/kT)}{Z}$$

density of samples

$$p(x;\theta) = q\left(f^{-1}(x;\theta)\right)\left|J_f^{-1}(x;\theta)\right|$$

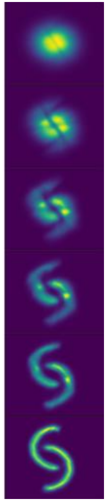
1: Variational inference with normalizing flows. Rezende & Mohammed. ICML. 2015

Figure: Neural ODEs, Chen et al. NeurIPS. 2018

Training mode I: negative log-likelihood

match base

$$f^{-1}(x;\theta) \overset{!}{\sim} q(z)$$



f^{-1}

inverse mode

Minimize

$$\mathcal{L}_{NLL}(\theta) = D_{KL}\left[\mu(x)\|p(x;\theta)\right] + \text{const.}$$

$$x \sim \mu(x)$$

Samples?!

Figure: Neural ODEs, Chen et al. NeurIPS. 2018

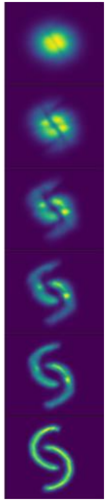
Training mode II: reverse KL

sample base

$$z \sim q(z)$$

forward mode

f



match energy

$$x = f(z;\theta) \overset{!}{\sim} \mu(x)$$

Minimize

$$\mathcal{L}_{KL}(\theta) = D_{KL}\left[p(x;\theta) \| \mu(x)\right] + \text{const.}$$

Mode collapse!

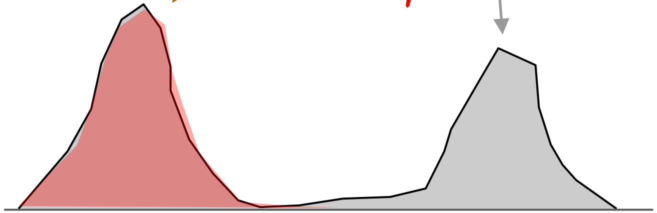


Figure: Neural ODEs, Chen et al. NeurIPS. 2018

Boltzmann Generators. Noé*, Olsson*, JK*, Wu. Science. 2019

Our setup

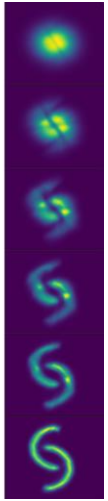
1. NLL on biased samples
(e.g. non-converged MD trajectory)

2. combine with KL training

3. correct with importance sampling

$$\mathbb{E}_{\mu}[O(x)] = \mathbb{E}_{x \sim p}\left[\frac{\mu(x)}{p(x)}\, O(x)\right]$$

$z \sim q(z)$ $\qquad \mathcal{L}_{NLL}(\theta)$ ← keep modes!



f $\qquad f^{-1}$

biased trajectory

$\mathcal{L}_{KL}(\theta)$ $\qquad x \sim \tilde{\mu}(x)$

better fit

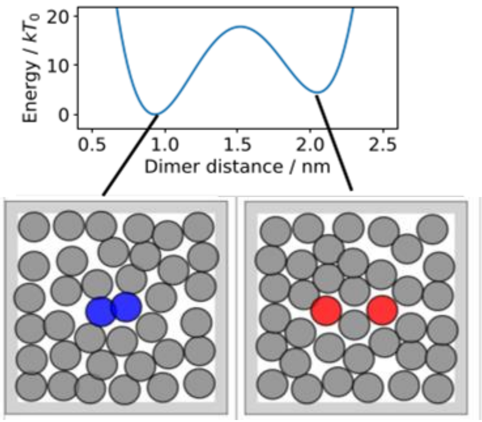
Joint loss:

$$\mathcal{L}(\theta) = \alpha \cdot \mathcal{L}_{KL}(\theta) + \beta \cdot \mathcal{L}_{NLL}(\theta)$$

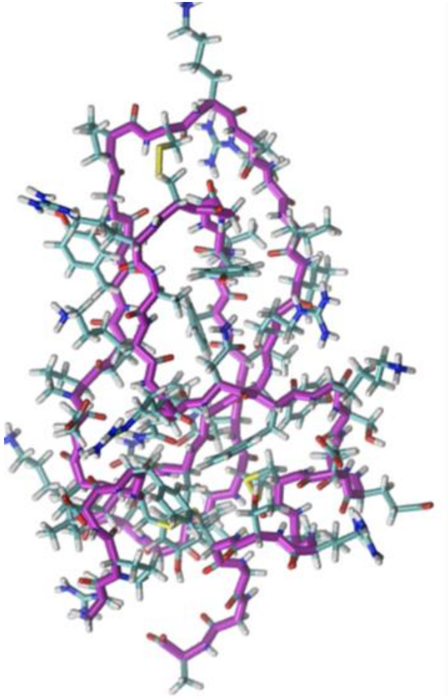
convex combination

Figure: Neural ODEs, Chen et al. NeurIPS. 2018

Test systems



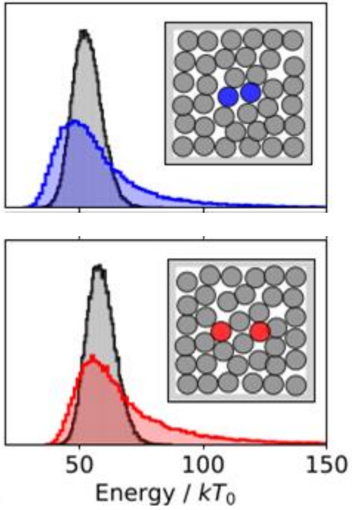
dimer in particle box

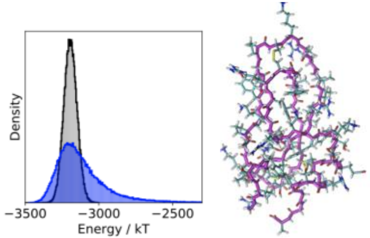


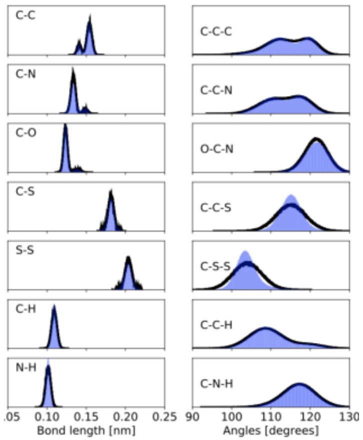
protein (BPTI) in implicit solvent

Results

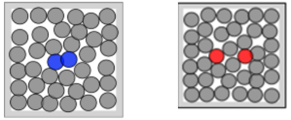
energies + marginals match

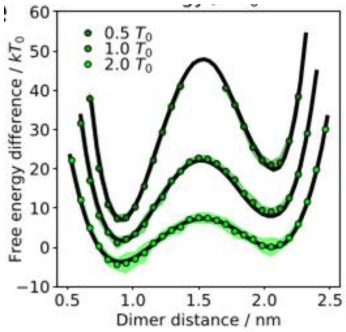


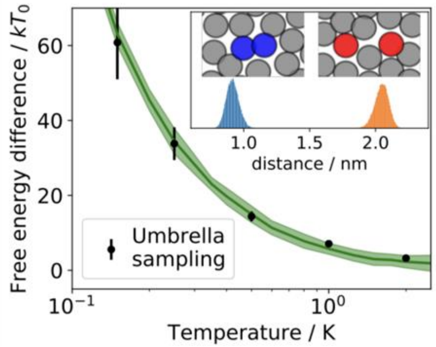




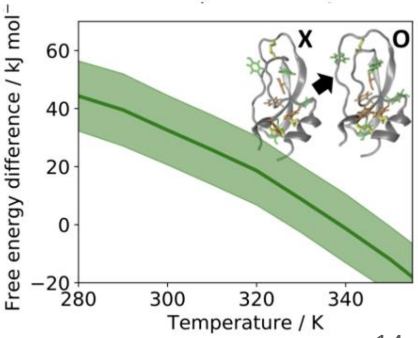
reweighing works

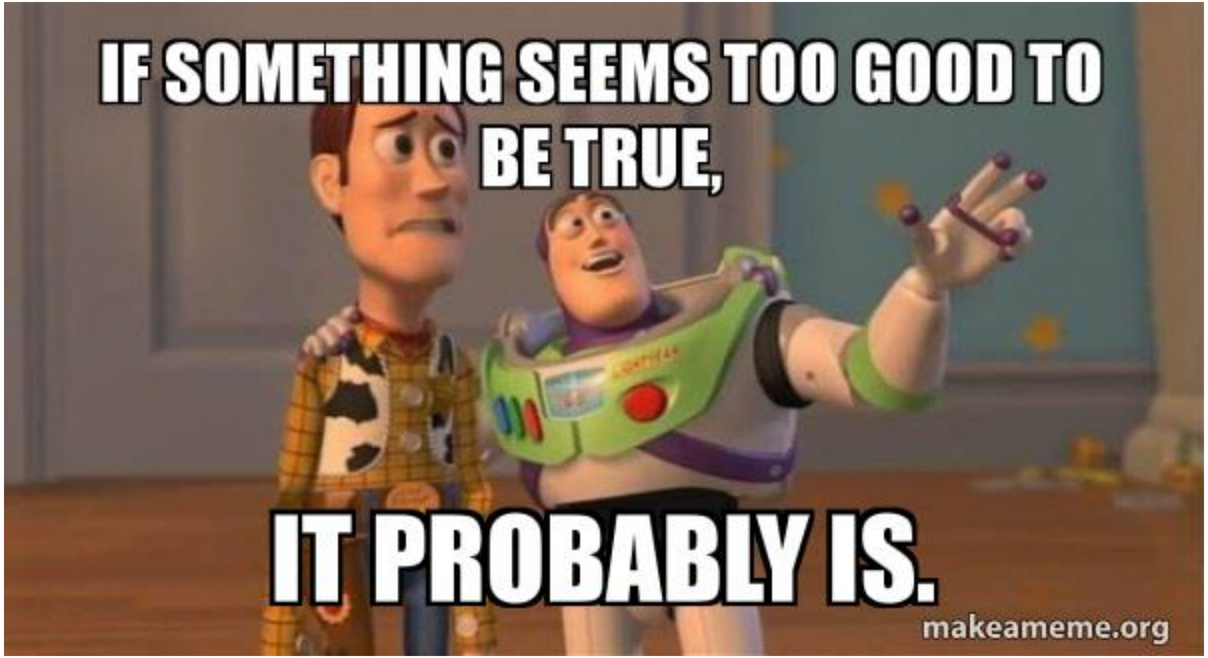






estimate free energy diff.

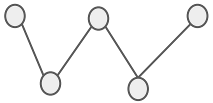



IF SOMETHING SEEMS TOO GOOD TO BE TRUE,
IT PROBABLY IS.
makeameme.org

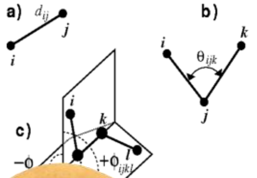
Actual picture of the method
at this state...

Topology / representation?

XYZ - space

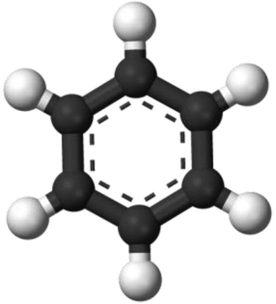


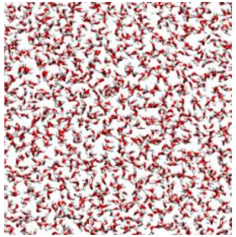
Internal Coordinates + Whitening



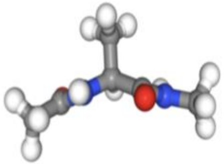
What are the problems?

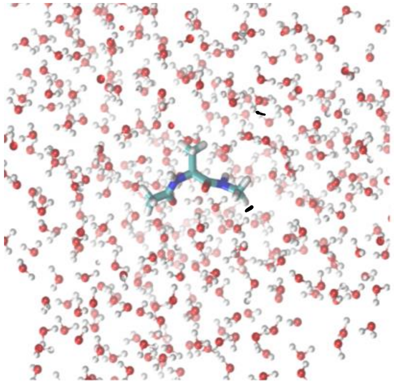
Symmetries?





Bulk systems?

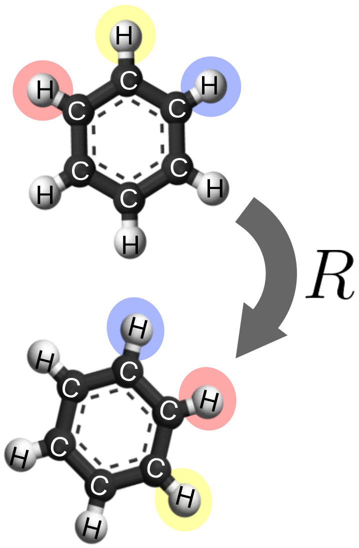




Equivariant Flows

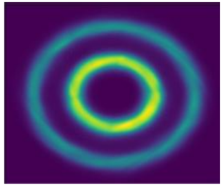
TL/DR: normalizing flows with group symmetries

Symmetries



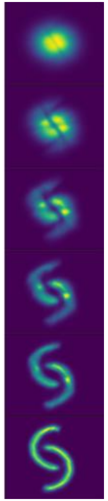
Invariant energy / density

$$\forall R \in \rho(G):\, u(Rx) = u(x)$$



$$z \sim q(z)$$

f



$$x = f(z;\theta) \sim p(x;\theta)$$

Arbitrary flow maps

$$p(Rx;\theta) \neq p(x;\theta)$$

Bad for reweighing!

Handles data inefficiently!

Figure: Neural ODEs, Chen et al. NeurIPS. 2018

Equivariant Flows. Köhler*, Klein*, Noé. ICML. 2020

Equivariant Flows

Constraint on group representations

$$\mu(\rho(g)x) = \mu(x)$$
$$\Downarrow$$
$$|\det \rho(g)| = 1.$$

Important for molecules:

$$G \leq O(n)$$

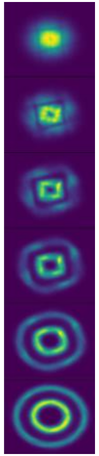
permutations *rotations*

symmetric

$$z \sim q(z)$$

equivariant

$$f$$



$$v(x) = \frac{\partial x}{\partial t}$$

$$v(x) = \nabla \phi(x)$$

invariant

$$x = f(z;\theta) \sim p(x;\theta)$$

symmetric

1: Equivariant Hamiltonian Flows. Rezende et al. arxiv:1909.13739

Figure: Neural ODEs, Chen et al. NeurIPS. 2018

Smooth Flows

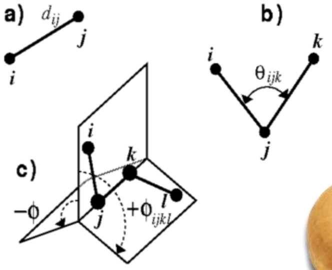
TL/DR: fix broken topology with smooth transforms on hypertorus!

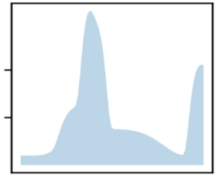
Constrained manifold of internal coordinates

Bond Lengths $d_{ij} \in (0,\infty)$

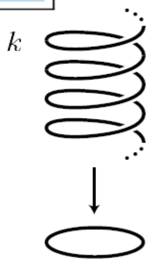
Angles $\theta_{ijk} \in [0,\pi]$

Torsions $\phi_{ijkl} \in S^1$





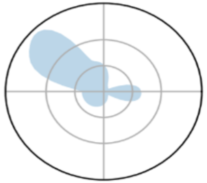
$p(x)$

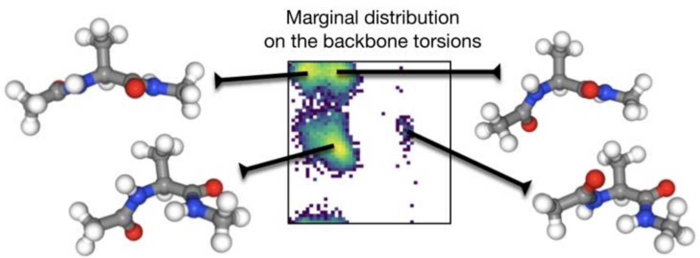


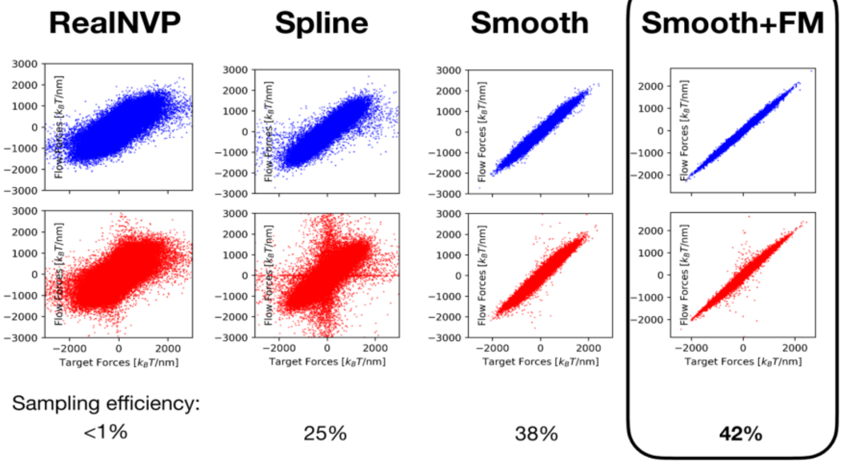
compact support

= finite sum!

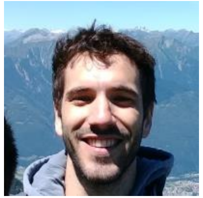
$$\tilde{p}(x) = \sum_{k\in\mathbb{Z}} p(x+k)$$







Rigid body flows for molecular crystals



Michele Invernizzi

Pim de Haan

Frank Noé

TL/DR: smooth and equivariant flows on SE(3)

Motivation: solvent systems and crystals

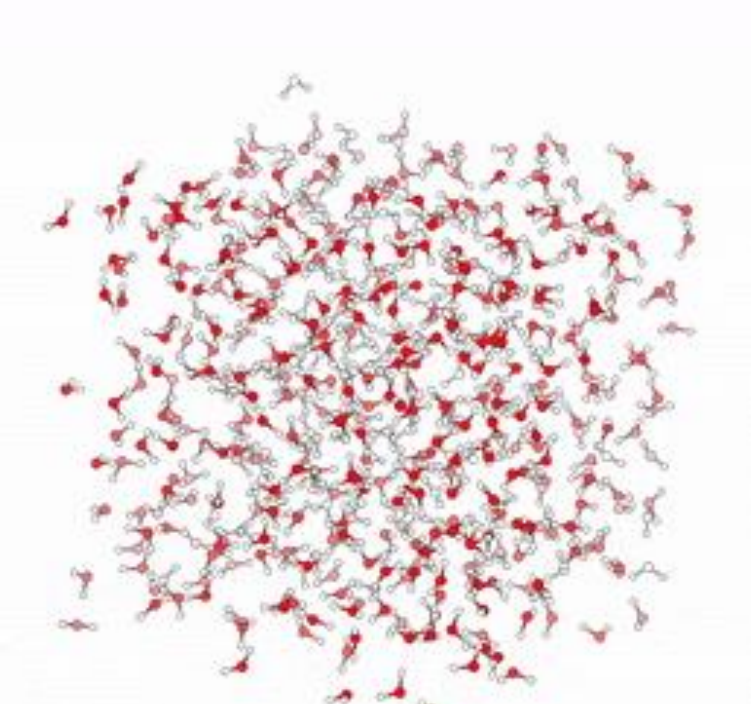
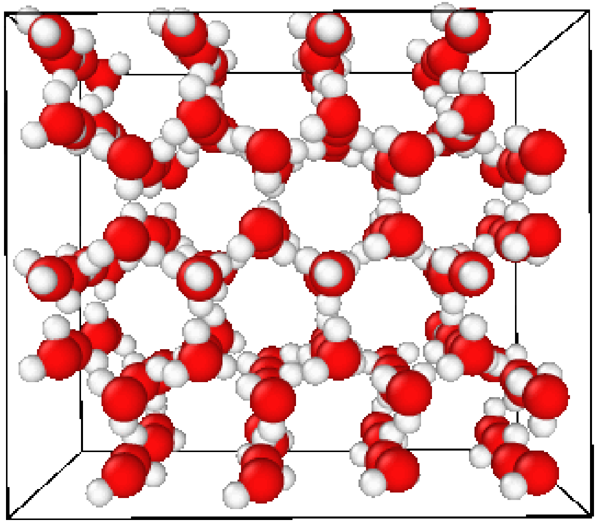
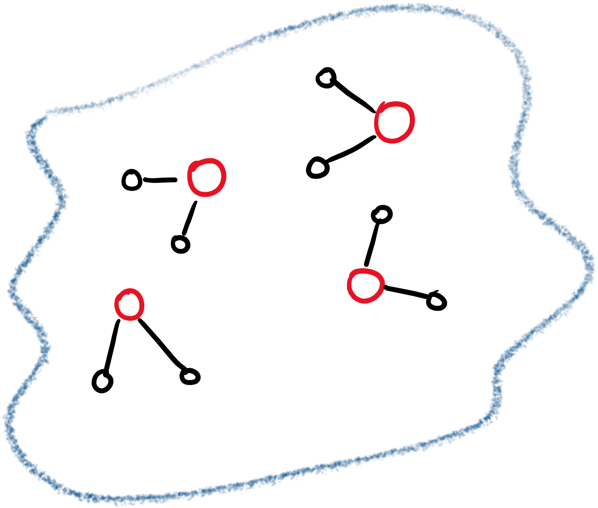


Figure: Wikipedia





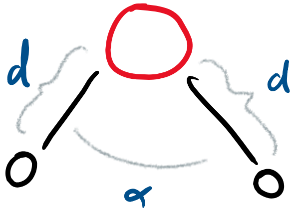
$x \in \mathbb{R}^{n \cdot a \cdot 3}$

for water $a = 3$

Usually fixed:

$1\times$ angle

$2\times$ bond length



Degrees of freedom:

position $r \in \mathbb{R}^{3n}$

rotation $R \in \underline{\underline{(SO(3))}}^n$

$6 < 9$ dof!

Support manifold:

$$\mathcal{M} = \left\{ x \in \mathbb{R}^{n \cdot a \cdot 3} \;\middle|\; \mu(x) \neq 0 \right\} = \left[\mathbb{R}^3 \times SO(3) \right]^n \neq \mathbb{R}^{n \cdot a \cdot 3}$$

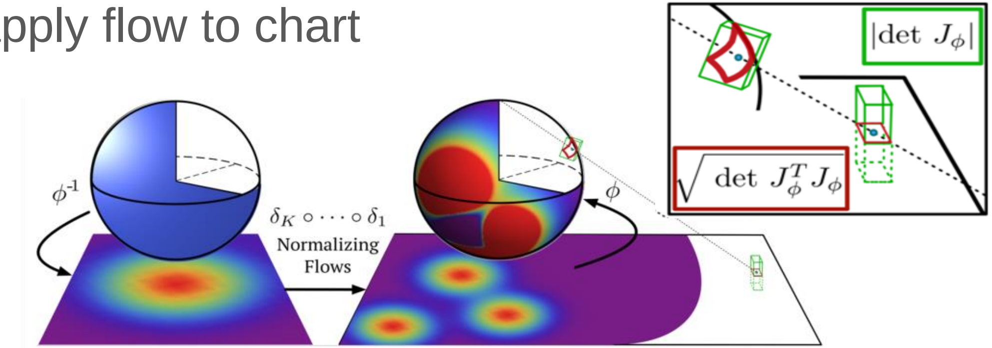
Charting

Cut manifold open into charts and apply flow to chart

- Easy to implement
- Fast

- Non-smooth solutions!

Figure: Gemici (2015)



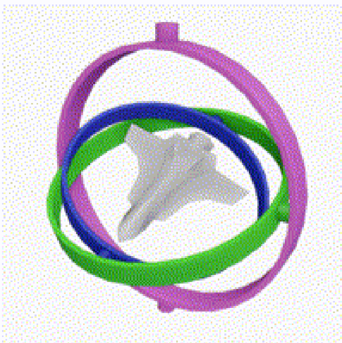


Figure: Wikipedia

Normalizing flows on riemannian manifolds, Gemici M., et. al. arXiv:1611.02304, 2015

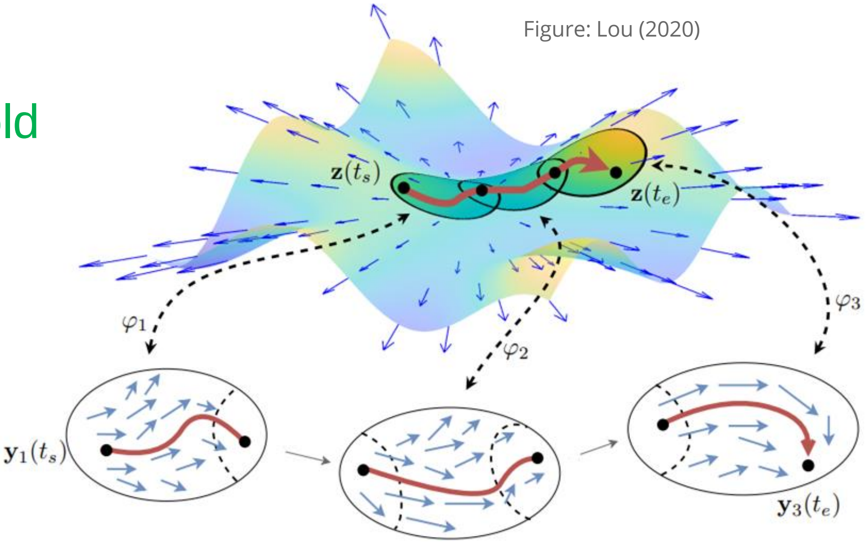
Continuous flows on manifolds

Integrate NN dynamics on manifold

- Works on every Riemannian manifold
- Smooth

- Difficult to train
  - Likelihood easy with flow-matching…
  - Rev. KL: adjoint method
- Slow integration
- Not scalable to high dimensions

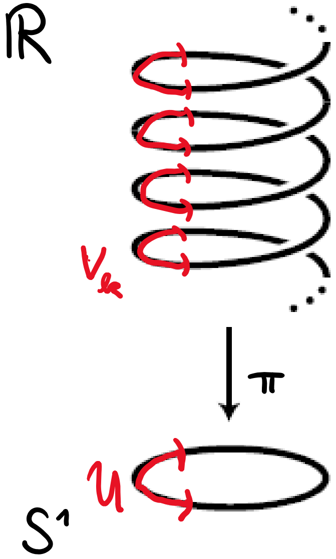
Figure: Lou (2020)



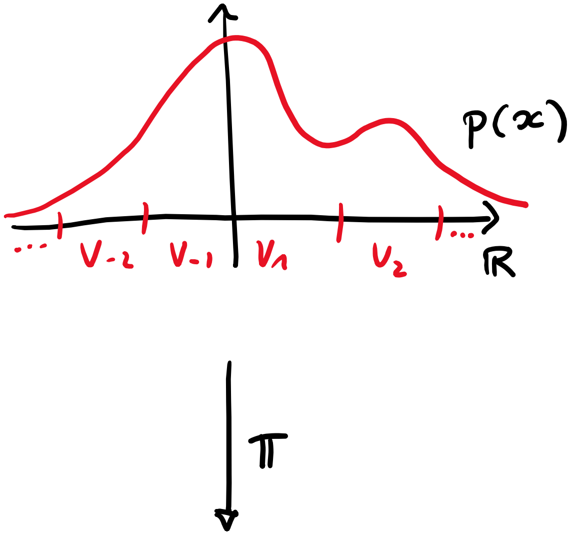
Lou, Aaron, et al. "Neural manifold ordinary differential equations." NeurIPS 2020

Covering flows

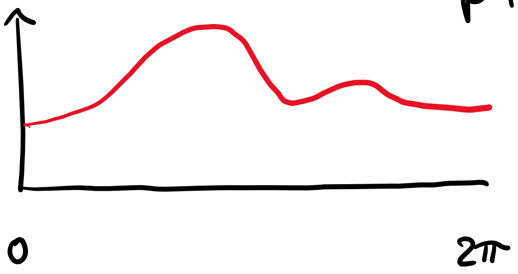
$\pi : \mathbb{R} \to S^1, \; x \mapsto \exp(i \cdot x)$



$\pi^{-1}(U) \cong U \times \mathbb{Z}$



$\tilde{p}(r) = \sum_{k \in \mathbb{Z}} p(x + k)$



Quaternions: $q = \underbrace{a \cdot 1}_{\text{scalar}} + \underbrace{b \cdot i + c \cdot j + d \cdot k}_{\text{vector}} \in \mathbb{R}^4$

$\boxed{\text{Rule: } i \cdot j \cdot k = -1}$

vector $= \mathrm{vec}(q)$

conjugate: $q^* = a - b \cdot i - c \cdot j - d \cdot k$

action on $\mathbb{R}^3$: $p = (x, y, z) \in \mathbb{R}^3 \longrightarrow \hat{p}\,(0, x, y, z) \in \mathbb{R}^4$

$$q \curvearrowright p = \mathrm{vec}\left(q^* \cdot \hat{p} \cdot q\right)$$

For $q \in S^3 = \{ q \in \mathbb{R}^4 \mid \|q\| = 1 \}$: ① $q \curvearrowright p$ is 3D rotation !

② $-q$ and q give same rotation

$\boxed{\Rightarrow \text{surjective map } \pi: S^3 \longrightarrow SO(3)}$ $\Rightarrow$ „$S^3 \to SO(3)$ double cover"

Main Result:

Section:

$\pi \circ s = 1$

$$
\begin{tikzcd}[column sep=huge, row sep=large]
S^3 \arrow[r, "\overbrace{\text{sign-equi } \Phi}^{\Phi(-q) = -\Phi[q]}"] \arrow[d, bend left, "\pi"] & S^3 \arrow[d, "\pi"] \\
SO(3) \arrow[u, dashed, bend left, "s"] \arrow[r, "\underbrace{\psi}_{\text{induced transformation}}"'] & SO(3)
\end{tikzcd}
$$

double cover

Result:

For <u>any</u> [even stoch.] s: ① ψ is a flow
② vol. change ψ = vol. change Φ !

$\Rightarrow$ no sum over preimages!

Return of the gradient flows

Strictly convex $\phi : \mathbb{R}^4 \to \mathbb{R}$

$$\Phi_{CG}(\boldsymbol{x}) = \frac{\nabla_{\boldsymbol{x}}\phi(\boldsymbol{x})}{\|\nabla_{\boldsymbol{x}}\phi(\boldsymbol{x})\|}$$

Result: ϕ sign-invariant $\Rightarrow$ Φ_{CG} satisfies constraints!

Boltzmann Generators:

simple prior $\mathcal{N}(0, I)$

flow $\Phi \downarrow$

target $\exp(-u(x))$

$\Longrightarrow$

Learned Free Energy Perturbation

„easy" system $\exp(-u_0(x))$ high temp.

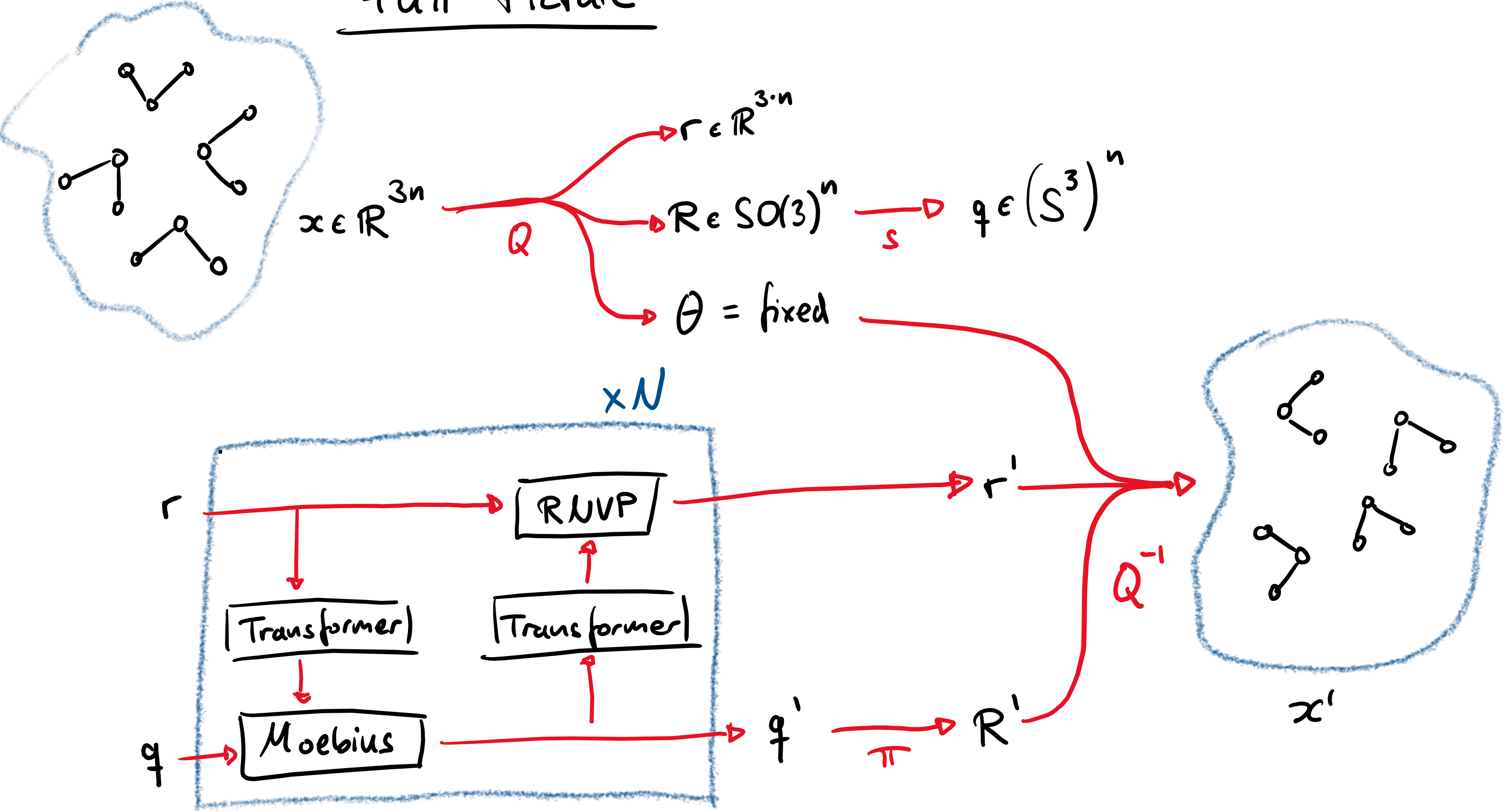
flow $\Phi \downarrow$

target $\exp(-u_1(x))$ low temp.

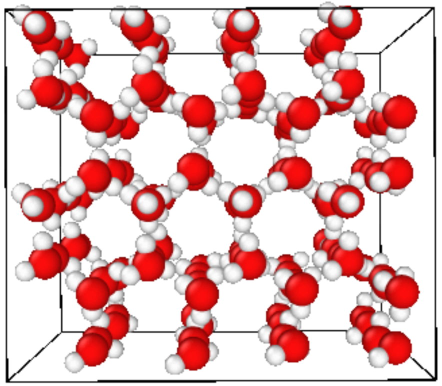
$$\boxed{\Delta F[u_0; u_1] \leq KL[p_{flow} \,\|\, p_{target}]}$$

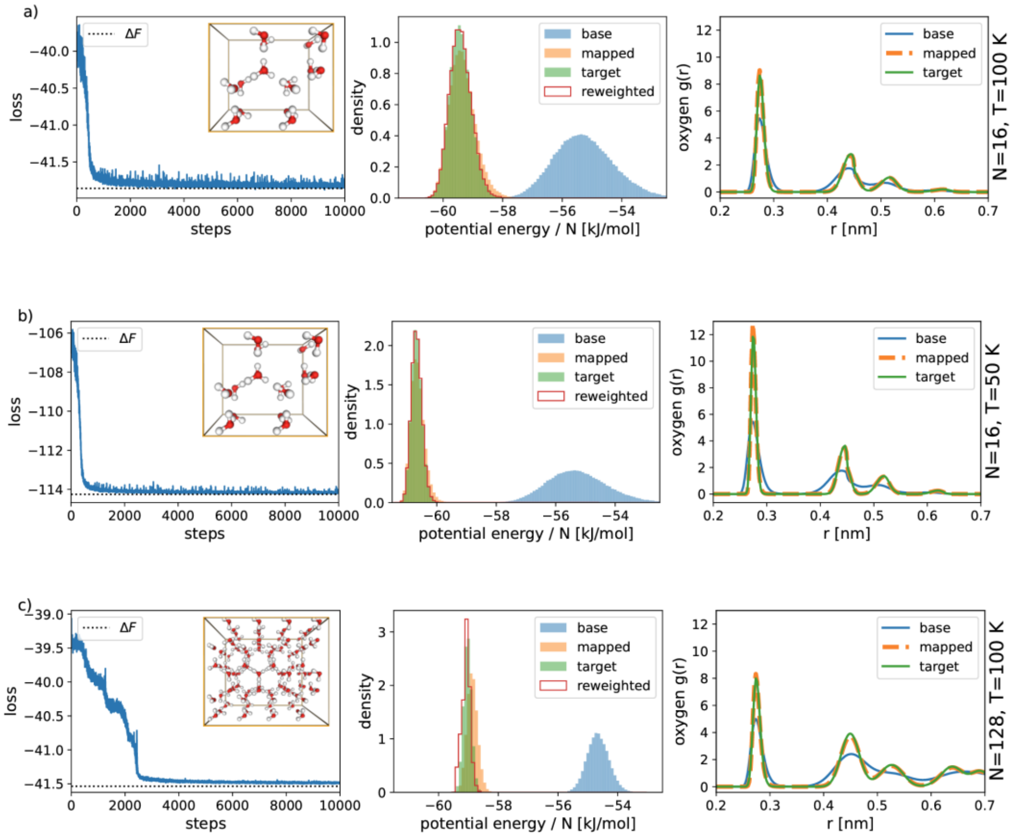
Targeted free energy estimation via learned mappings, Wirnsberger et. al., JCP 2020

Full Picture



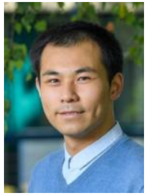
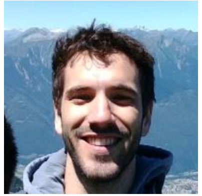
Results: Ice in different thermodynamic states





Target	MBAR	LFEP
N=16, T=100 K	-41.857 ± 0.007	-41.859 ± 0.002
N=16, T=50 K	-114.251 ± 0.007	-114.252 ± 0.005
N=128, T=100 K	-41.535 ± 0.002	-41.534 ± 0.003

Thanks!

Freie Universität Berlin

UNIVERSITY OF AMSTERDAM

DAEDALUS
RTG 2433

Microsoft