LPENS
LABORATOIRE DE PHYSIQUE
DE L'ÉCOLE NORMALE SUPÉRIEURE

Assisting sampling of physical systems, at thermal equilibrium, with generative models

Algorithms in Structural Bioinformatics
Cargèse, Nov 22, 2024

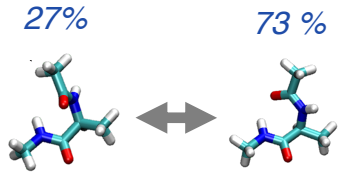
Marylou Gabrié
Laboratoire de Physique de l'ENS

Based on works with:
Grant Rotskoff (Stanford), Eric Vanden-Eijnden (Courant Institute, NYU)
Christoph Schönle (École Polytechnique)
Gabriel Stoltz & Tony Lelièvre (École Nationale des Ponts & Chaussés)
Pilar Cossio (Flatiron, CCM), Olga Lopez Acevedo & Ana Molina Taborda (Universidad de Antioquia)
Samuel Tamagnone & Alessandro Laio (SISSA)

INTRODUCTION

Setting the stage of the sampling problem

- Information about the system is the Boltzmann distribution: $\rho_*(x) = \frac{1}{\mathcal{Z}_\beta} e^{-\beta U(x)}$ (known up to normalization)



- Although we will do ML, there is **no extensive dataset** to start with

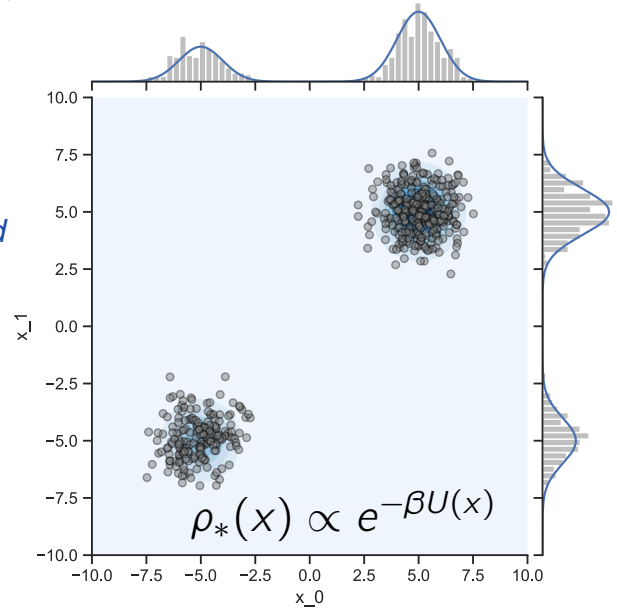
pick your preferred notation

- Averages with respect to $\rho_*(x)$ describe the thermodynamics $\langle f(x) \rangle_{\rho_*} = \mathbb{E}_{\rho_*}[f(x)]$
  → too costly to compute by numerical integration ...

- Monte Carlo approximations consist in sampling

  with $x_i \sim \rho_*(x)$, $\frac{1}{N} \sum_{i=1}^{N} f(x_i) \underset{N \to \infty}{\longrightarrow} \mathbb{E}_{\rho_*}[f(x)]$

ex: one particle in 2d with two wells



- Overall goal: Sample from a target distribution to characterize the thermodynamics!

- Disclaimer: No information about the dynamics!

How to sample?

▷ Monte Carlo simulations were first proposed in the late 1940s in problems related to high-energy physics.

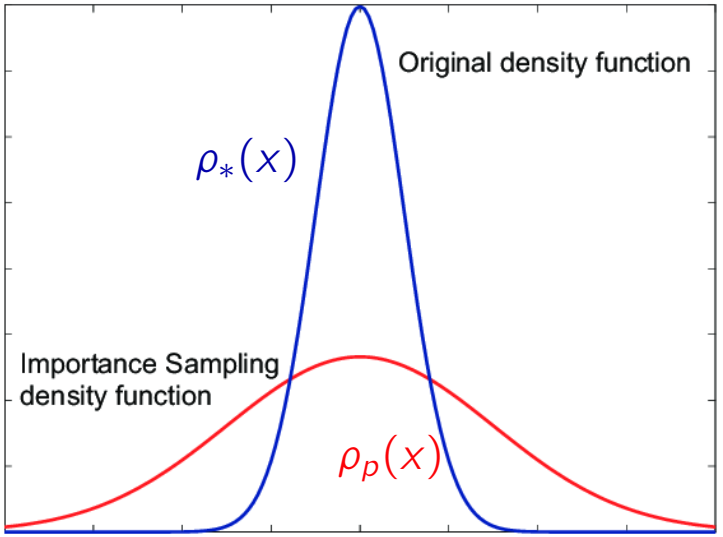
▷ From these early beginning, the premises of two very important algorithms were set

- Importance Sampling
- Markov Chain Monte Carlo
  - All purpose framework: Metropolis-Hastings sampling

    NB: Metropolis can also be used to look for a free energy minima (cf Alisa's talk)
  - Some versions inspired by physical dynamics you might have heard about:
    Langevin dynamics, Metropolis-Adjusted Langevin (MALA), Hamiltonian Monte Carlo (HMC)

Importance Sampling

- Context: $\rho_*(x) = \frac{1}{\mathcal{Z}} e^{-U(x)}$ with unkown $\mathcal{Z}$
- Task: Compute expectations $\mathbb{E}_\rho[f(x)] = \int_\Omega f(x)\rho_*(x)\mathrm{d}x$
- Method:
  - Samples from proposal distribution $x_i \sim \rho_p(x_i)$ e.g. Gaussian, factorized, …
  - Self-normalized weights $w_i = \frac{e^{-U(x_i)}/\rho_p(x_i)}{\sum_{i=1}^N e^{-U(x_i)}/\rho_p(x_i)}$
  - Compute $\mathbb{E}_{\rho^*}[f(x)] \approx \frac{1}{N}\sum_{i=1}^N w_i f(x_i)$
  - Asymptotically “unbiased” $\mathbb{E}_\rho[f(x)] = \lim_{N\to\infty} \frac{1}{N}\sum_{i=1}^N w_i f(x_i)$



Markov Chain Monte Carlo

▷ Idea: design transition kernel $\pi(x_{t+1}|x_t)$ such that chain $x_0, x_1, \ldots, x_t$ produces samples from ρ_* for t large

▷ Important example:

Metropolis-Hastings sampler

Initialize: x_0

Iterate:

- Propose $x_{t+1} \sim \rho_p(x_{t+1}|x_t)$
- Accept/Reject with prob.

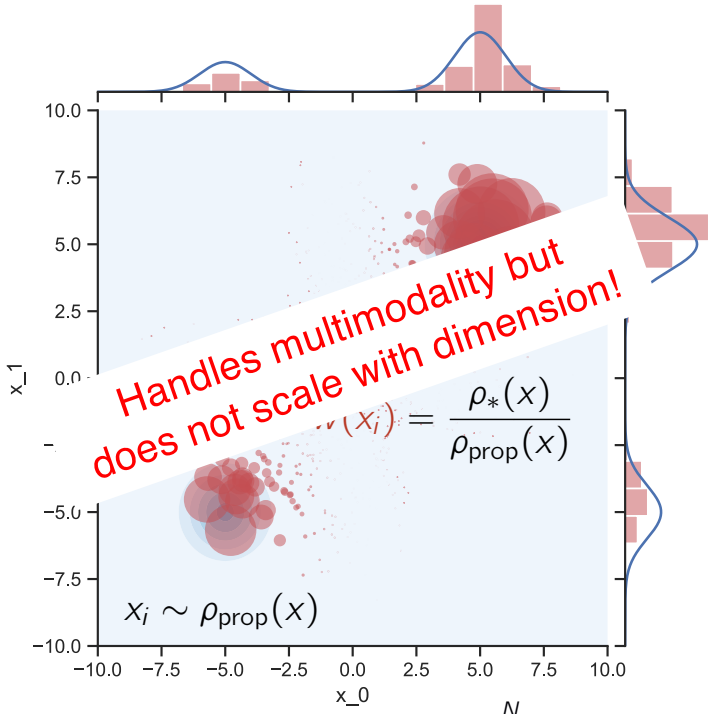
$$\text{acc}(x_{t+1}|x_t) = \min\left[1, \frac{\rho_*(x_{t+1})\rho_p(x_t|x_{t+1})}{\rho_*(x_t)\rho_p(x_{t+1}|x_t)}\right]$$

- If reject stay $x_{t+1} = x_t$

[e.g. Liu. *Monte Carlo Strategies in Scientific Computing*, 2004
Brooks et al. *Handbook of MCMC*, 2011
C. P. Robert & W. Changye, *Markov Chain Monte Carlo Methods, A survey with some frequent misunderstandings*, 2020]

Why is sampling so hard?

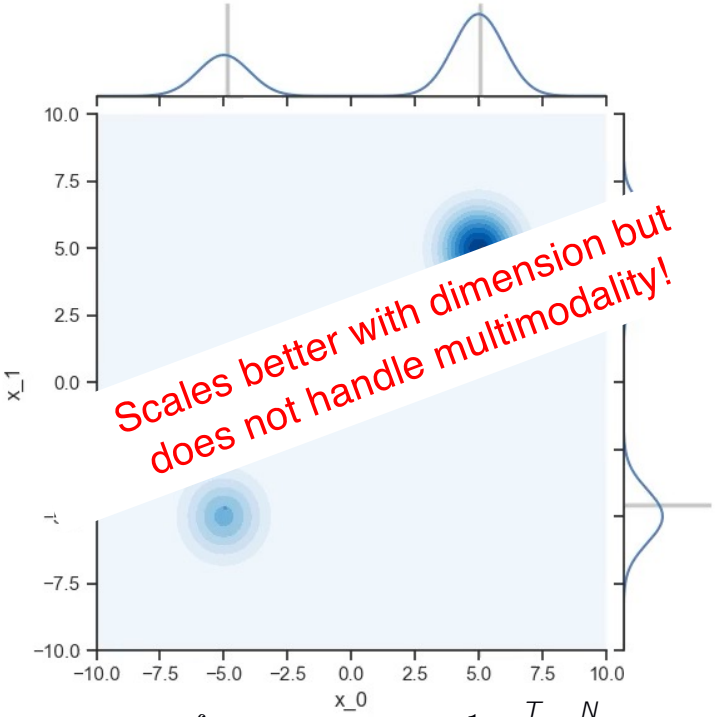
▷ Shoot and reject/reweight algorithms:
(e.g. Importance Sampling IS)



$$\mathbb{E}_{\rho_*}[f(x)] = \int_\Omega f(x)\rho_*(x)\mathrm{d}x \approx \frac{1}{N}\sum_{i=1}^{N} w(x_i)f(x_i)$$

High variance!

▷ Local exploration Markov chain Monte Carlo (MCMC) *(e.g. Metropolis Adjusted Langevin)*



$$\mathbb{E}_{\rho_*}[f(x)] = \int_\Omega f(x)\rho_*(x)\mathrm{d}x \approx \frac{1}{NT}\sum_{t=1}^{T}\sum_{i=1}^{N} f(x_i^t)$$

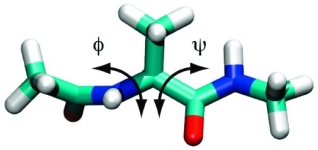
High bias!

Enhanced sampling algorithms

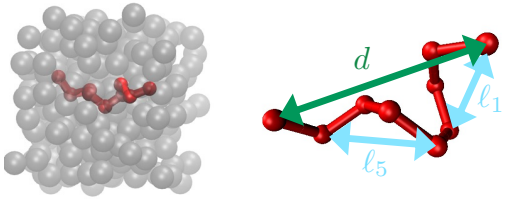
- Temperature based:
  - Couple the system to higher temperature copies of the same system driving the exploration and transitions between metastable states.
  - Example: Parallel Tempering (a.k.a. Replica Exchange) [Marinari & Parisi (1992), Geyer & Thomson (1995), Neal (1998) etc.]
  - Challenges: Computationally costly, hard to tune (how high should the high temperature be? how close in temperature should be the replicas?)

- Collective variable based: [Fu et al. "Enhanced Sampling Based on Collective Variables." 2023]
  - In cases where the metastability of interest can be described by an identifiable small collection of variables.
  - Example: Umbrella Sampling, Meta-dynamics, Adaptive Biasing Force
  - Challenges: Identifying this collective random variables (typically few)

Dihedral angles



Main chain of solvated polymer



Ongoing revolution of generative modellig

▷ Ground-breaking progress has been made into generating high-quality data with neural nets
Architectures: GANs, VAE, Normalizing flows, Neural ODE, Score-based Diffusion models etc..
Famous models: Midjourney, ChatGPT etc..

▷ Deep latent generative models produce highly structured data points at negligible cost

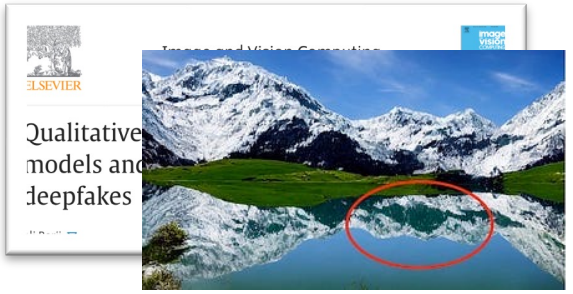
$$T_\theta : \Omega \mapsto \Omega$$

$$z \sim \rho_B(z) \quad \theta_1 \quad \theta_2 \quad \theta_3 \quad x = T_\theta(z) \quad \sim \rho_\theta(x)$$

push-forward distribution

[Song et al. ICLR 2021]

New dog picture for each new base variable!



[Ali Borji, Image and Vision Computing 2023]

▷ How can they be used for sampling?

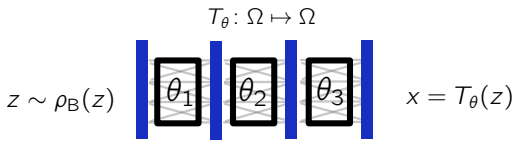
Q1: Guarantees on the quality of the output?

Q2: Training without extensive data?

Task	generative modelling	sampling
Input	$\mathcal{D} = \{x_i\}_{i=1}^N$	$\rho_*(x) \propto e^{-\beta U(x)}$

Outline

1.1 Normalizing Flows (NF)



▷ Parametrized invertible map $T_\theta : \Omega \mapsto \Omega \quad \Omega \subset \mathbb{R}^d$

Transport!

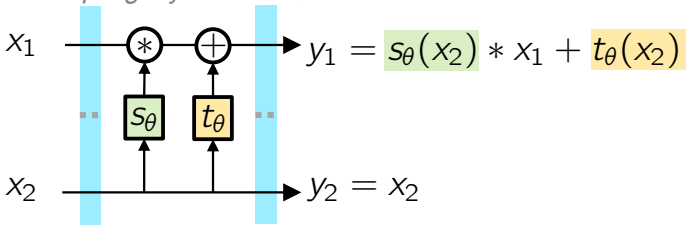
- Base distribution $z \sim \rho_B(z)$
- Push-forward distribution $x = T_\theta(z) \sim \rho_\theta(x) = \rho_B(T_\theta^{-1}(x)) \det \left| \nabla_x T_\theta^{-1} \right|$

[Tabak & Vanden Eijnden Commun. Math. Sci. 2010, Papamakarios, et al JMLR 2021 (review)]

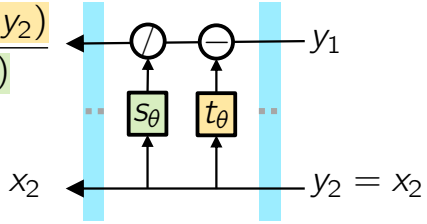
▷ How to construct an invertible parametrizable map? → The coupling layer trick

Dinh et al "Density Estimation Using Real Nvp." ICLR 2027

Affine coupling layer $T_\theta(x)$



Inverse layer $T_\theta^{-1}(y)$



Block triangular Jacobian:

$$\nabla_x T_\theta(x) = \begin{bmatrix} s_\theta(x_2) I_{d/2} & A \\ 0 & I_{d/2} \end{bmatrix}$$

- Can be generalized beyond affine maps, as long as $y_1 = f_{x_2}(x_1)$ and $x_1 = f_{y_2}^{-1}(y_1)$ are available + easy Jacobians e.g. Rational Quadratic Splines (RQS)

  Durkan et al, "Neural Spline Flows." NIPS 2019.

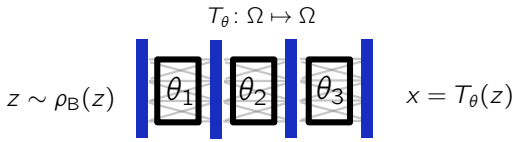
▷ Compose to get expressive maps $T_\theta = T_{\theta_4} \circ T_{\theta_3} \circ T_{\theta_2} \circ T_{\theta_1}$

▷ Crucial property: Easy to sample from and tractable likelihood!

Discrete data? Autoregressive models!

Benigno et al. JMLR 2016. "Neural Autoregressive Distribution Estimation."

1.1 Normalizing Flows (NF) training



▷ Parametrized invertible map $T_\theta : \Omega \mapsto \Omega \quad \Omega \subset \mathbb{R}^d$

Transport!

- Base distribution $z \sim \rho_B(z)$
- Push-forward distribution $x = T_\theta(z) \sim \rho_\theta(x) = \rho_B(T_\theta^{-1}(x)) \det \left| \nabla_x T_\theta^{-1} \right|$

[Tabak & Vanden Eijnden Commun. Math. Sci. 2010, Papamakarios, et al JMLR 2021 (review)]

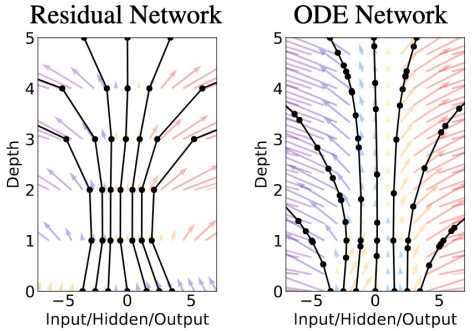
▷ Training is straightforward thanks to the tractable likelihood

▷ Maximum likelihood

- Given data samples $x_i \quad i = 1 \cdots N$
- Likelihood of parameter $\theta \in \mathbb{R}^P$: $\ell[\theta] = \prod_{i=1}^{N} \rho_\theta(x_i)$ (independent samples)
- Negative log-likelihood loss: $L[\theta] = -\sum_{i=1}^{N} \log \rho_\theta(x_i)$ + (Stochastic) Gradient Descent!

1.2 Continuous normalizing flow, Neural ODE

Ordinary Differential Equation



▷ Parametrized invertible map $T_\theta : \Omega \mapsto \Omega \quad \Omega \subset \mathbb{R}^d$

- Base distribution $x_0 \sim \rho_B(x_0)$
- Output defined through a parametrized dynamical process

$$\frac{\mathrm{d}x_t}{\mathrm{d}t} = v_\theta(x_t, t) \qquad x_1 = x_0 + \int_0^1 v_\theta(x_t, t)\mathrm{d}t$$

- Learnable velocity field $v_\theta(x, t) : \Omega \times [0, 1] \to \Omega$
- Time plays an analogous role to depth in model defined by composition
- Typically uses adaptive integrators (non-constant time step)

▷ Computing the likelihood also requires integrating an ODE

- Instantaneous change of variable formula $\frac{\mathrm{d} \ln \rho_{\theta,t}(x_t)}{\mathrm{d}t} = -\nabla_x \cdot v_\theta(x_t, t)$

$$\implies \quad \ln \rho_{\theta,t}(x_t) = \ln \rho_B(x_0) - \int \nabla_x \cdot v_\theta(x_t, t)\mathrm{d}t$$

▷ Training? Still by maximum likelihood. What is hard about it?

- Computing derivative through integrals → adjoint method → github.com/rtqichen/torchdiffeq

Chen et al. “Neural Ordinary Differential Equations.” NeurIPS 2019

1.2 Continuous normalizing flows with invariances

- Continuous NF are heavier than their discrete counterparts, so why use them?
  - Possibly more expressive
  - More importantly allows to incorporate symmetries
- Idea is to constrain our learned distribution $\rho_\theta(x)$ to be invariant under a given group symmetry $\mathcal{G}$

  For $g \in \mathcal{G}$, $\rho_\theta(g[x]) = \rho_\theta(x)$

  e.g. rotations, translations, permutations
- We still want the nice properties of transport-based models (easy to sample + tractable likelihood)
  - Consider an **invariant** based distribution $\rho_B(g[x]) = \rho_B(x)$
  - **Equivariant** velocity filed $v_\theta(g[x], t) = g[v_\theta(x, t)]$
  - The resulting $\rho_{\theta,t}(x_t) = \rho_B(x_0)e^{-\int \nabla_x \cdot v_\theta(x_t,t)\mathrm{d}t}$ is invariant with respect to $\mathcal{G}$
- How to construct an equivariant velocity field?
  - Option 1: Consider invariant function engineered by hand $\phi_\theta : \Omega \to \mathbb{R}$ $\phi_\theta(g[x]) = \phi_\theta(x)$, take the gradient $v_\theta(x) = \nabla_x \phi_\theta(x)$
  - Option 2: Rely on dedicated libraries relying on more fundamental principles, e.g. github.com/e3nn/e3nn github.com/QUVA-Lab/e2cnn (cf Ilia's talk!)

Köhler et al, "Equivariant Flows: Exact Likelihood Generative Learning for Symmetric Densities" ICML 2020
Garcia Satorras et al, "E (n) equivariant normalizing flows", Neurips 2021

1.3 Diffusion models

▷ Noising and denoising stochastic differential equations

Data noising process

$$\mathrm{d}x_\tau = -x_\tau \mathrm{d}\tau + \sqrt{2}\mathrm{d}B_\tau \implies x_{\tau=T} \sim \rho_B(x) = \mathcal{N}(0, I_d)$$

$$x_{\tau=0} \sim \rho_{\text{data}}(x_0)$$

Denoising process

$$t = T - \tau$$

Instantaneous distribution

$$\implies \mathrm{d}x_t = x_t \mathrm{d}t + 2\nabla_x \ln \rho_t(x_t)\mathrm{d}t + \sqrt{2}\mathrm{d}B_t$$

$$x_{t=0} \sim \rho_B(x_0)$$

▷ The backward SDE samples (approximately) from the data distribution at time T Generative process !

But the so-called score function $s(x, t) = \nabla_x \ln \rho_t(x)\mathrm{d}t$ is unknown

▷ Score can be learned by denoising score matching (no need to compute the likelihood anymore)

- Loss $L(\theta) = \int_0^T \frac{\mathrm{d}t}{N} \sum_{i=1}^{N} \|s_\theta(x_t^i) - \nabla \ln \rho_t(x_t^i|x_0)\|^2$

Conditional distribution: Gaussian, easy to evaluate and to sample

$$x_t^i \sim \rho_t(x_t^i|x_0)$$

▷ Although, one **<u>can compute the likelihood</u>** using the ODE equivalent to the reverse SDE

$$\mathrm{d}x_t = x_t \mathrm{d}t + 2\nabla_x \ln \rho_t(x_t)\mathrm{d}t + \sqrt{2}\mathrm{d}B_t \iff \frac{\mathrm{d}x_t}{\mathrm{d}t} = x_t + \nabla_x \ln \rho_t(x_t)$$

$x_t \sim \rho_t(x_t)$

Approximated by $\frac{\mathrm{d}x_t}{\mathrm{d}t} = \overbrace{x_t + s_\theta(x_t, t)}^{v_\theta(x_t, t)}$

$$\implies \rho_{\theta,t}(x_t) = \rho_B(x_0) e^{-\int \nabla_x \cdot v_\theta(x_t, t)\mathrm{d}t}$$

Sohl-Dickstein et al. “Deep Unsupervised Learning Using Nonequilibrium Thermodynamics.” ICML 2015,
Ho et al. “Denoising Diffusion Probabilistic Models.” NeurIPS 2020,
Song et al. “Score-Based Generative Modeling through Stochastic Differential Equations.” ICLR 2021

1.3 Diffusion models

- Advantages
  - Can incorporate invariances

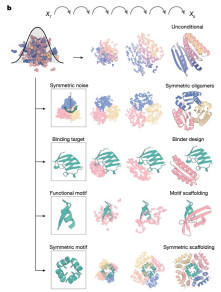
    e.g. for rotation invariance: add to the invariant base distribution and an equivariant score of the CNF the assumption of an isotropic noise.

    Hoogeboom et al, "Equivariant Diffusion for Molecule Generation in 3D" ICML 2022

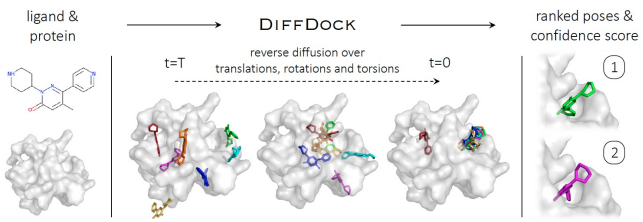
  - Computationally lighter than Continuous Normalizing flows → in particular the loss is much cheaper to evaluate

- Already quite impressive for generating molecular structure

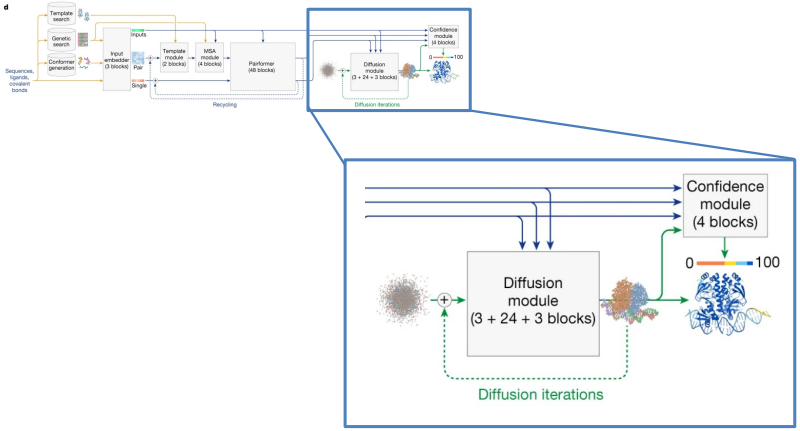
RF diffusion



DiffDock



AlphaFold 3



Corso et al, “DiffDock: Diffusion Steps, Twists, and Turns for Molecular Docking”, ICLR 2023
Watson et al. “De Novo Design of Protein Structure and Function with RFdiffusion.” Nature 2023
Abramson et al. “Accurate structure prediction of biomolecular interactions with AlphaFold 3”. Nature 2024.

1.4 Flow matching / stochastic interpolant models

▷ Build a bridge between two arbitrary distributions

$\rho_1(x_1)$ data distribution

$\rho_0(x_0)$ arbitrary base distribution

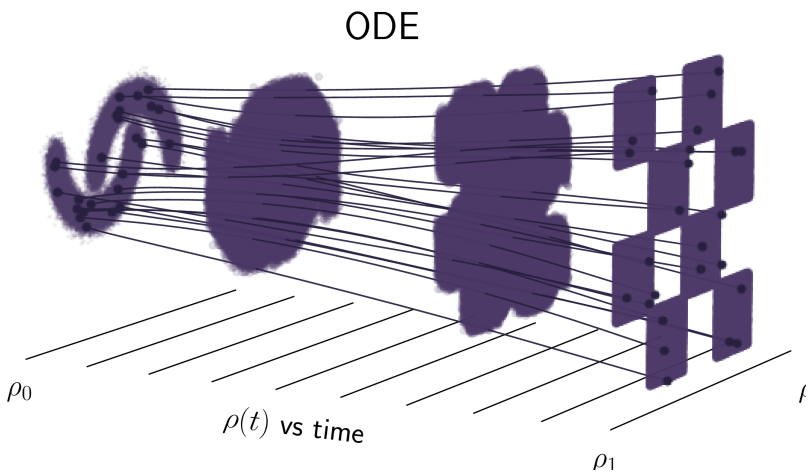
Interpolation function $I_t(x_0, x_1) = x_t = tx_1 + (1-t)x_0$

Instantaneous distribution $\rho_t(x) = \int \mathrm{d}x_0 \int \mathrm{d}x_1 \delta\big(x - I_t(x_0, x_1)\big)\rho_0(x_0)\rho_1(x_1)$

▷ The distribution $\rho_t(x)$ is also generated by an ODE $\dfrac{\mathrm{d}x_t}{\mathrm{d}t} = v(x_t, t)$

with $v = \arg\min \int_0^1 \mathrm{d}t\, \mathbb{E}_{\rho_0,\rho_1}\left[\|v(x_t) - \partial_t I_t(x_0, x_1)\|^2\right]$

Loss function for learning the velocity



https://github.com/malbergo/stochastic-interpolants

▷ Close cousins to diffusion models, differences/advantages

- Base distribution can be arbitrary
- The process bridges the two distribution exactly between times 0 and 1
- But requires to choose an interpolant (how?)

Albergo et al, “Building Normalizing Flows with Stochastic Interpolants”, ICLR 2022
Lipman et al, “Flow Matching for Generative Modeling,” ICLR 2022
Albergo et al, “Stochastic Interpolants: A Unifying Framework for Flows and Diffusions”, JMLR 2024

Summary of part 1

- ▷ Transport based generative models allows to **evaluate the likelihood** of the generator

- ▷ **Normalizing flows** (NFs) rely on compositions of simple invertible layers

- ▷ **Continuous normalizing flows** (CNFs) parametrize the transport with a velocity field and an ODE
  - They are computationally heavier than NFs but are typically more expressive.
  - They can be made invariant to desired symmetries by considering:
    - invariant base distributions
    - equivariant velocity field.

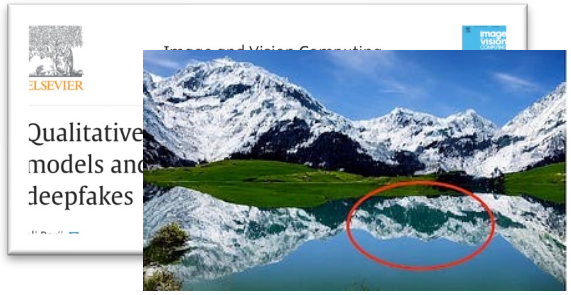
- ▷ **Diffusion models** parametrize the transport with a Stochastic Differential Equation.
  - They learn the score of a noising process to generate data through the backward denoising process.
  - They are computationally lighter than continuous normalizing flows but can be made invariant equally.

- ▷ **Flow matchings** (a.k.a. stochastic interpolant) are close cousins with more flexibility for the base distributions

Outline

Recall: main challenges in sampling with generative models

Q1: Guarantees on the quality of the output?



[Ali Borji, Image and Vision Computing 2023]

Q2: Training without extensive data?

Task	generative modelling	sampling
Input	$\mathcal{D} = \{x_i\}_{i=1}^N$	$\rho_*(x) \propto e^{-\beta U(x)}$

Generative models for sampling: first ideas

Variational Inference with Normalizing Flows ICML 2015

Danilo Jimenez Rezende DANILOR@GOOGLE.COM
Shakir Mohamed SHAKIR@GOOGLE.COM
Google DeepMind, London

Flow-based generative models for Markov chain Monte Carlo in lattice field theory PRD 2019

M. S. Albergo,[1,2,3] G. Kanwar,[4] and P. E. Shanahan[4,1]

Boltzmann generators: Sampling equilibrium states of many-body systems with deep learning

Frank Noé*†, Simon Olsson*, Jonas Köhler*, Hao Wu

Science 2019

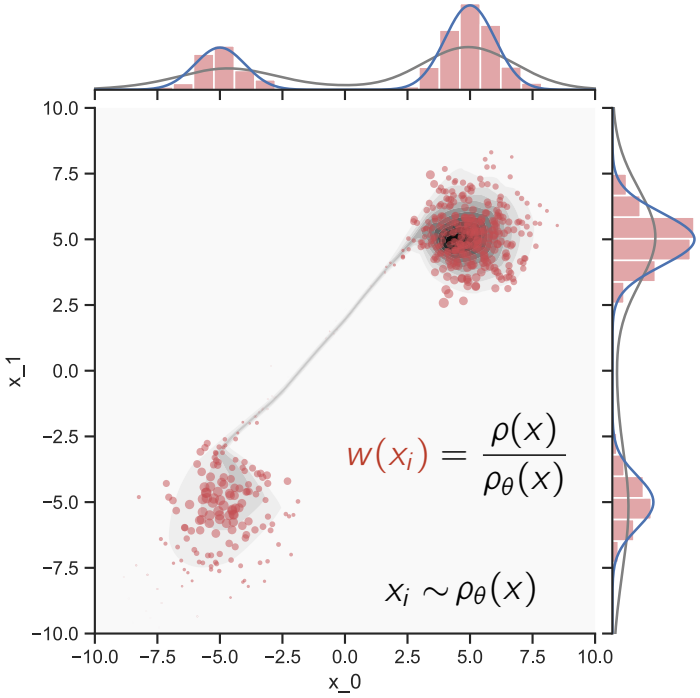
Solving Statistical Mechanics Using Variational Autoregressive Networks PRL 2019

Dian Wu,[1] Lei Wang,[2,3,4,*] and Pan Zhang[5,†]

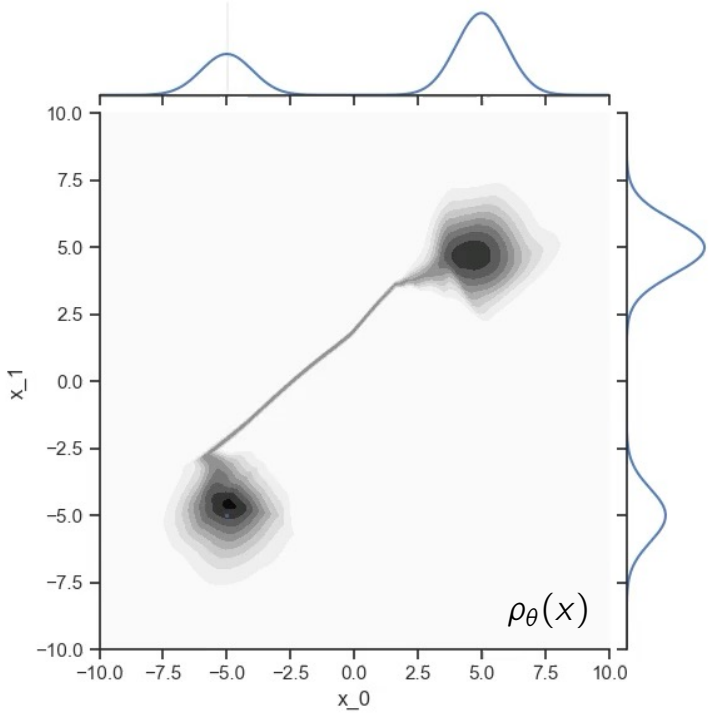
2.1 NF assisted sampling

Jumping ahead: suppose you can train a NF model $\rho_\theta(x) \approx \rho_*(x)$, what do you gain?

- ▷ Importance sampling:
  *rely on **adapated** tractable proposal!*



- ▷ Markov Chain Monte Carlo: e.g. Metropolis Hastings
  *rely on **global** proposal!* $\rho_{\text{prop}}(x_{t+1}|x_t) = \rho_\theta(x_{t+1})$



Computing the weights and the accept-reject criteria requires the computation of the model likelihood!

Q1: Guarantees on the quality of the output?

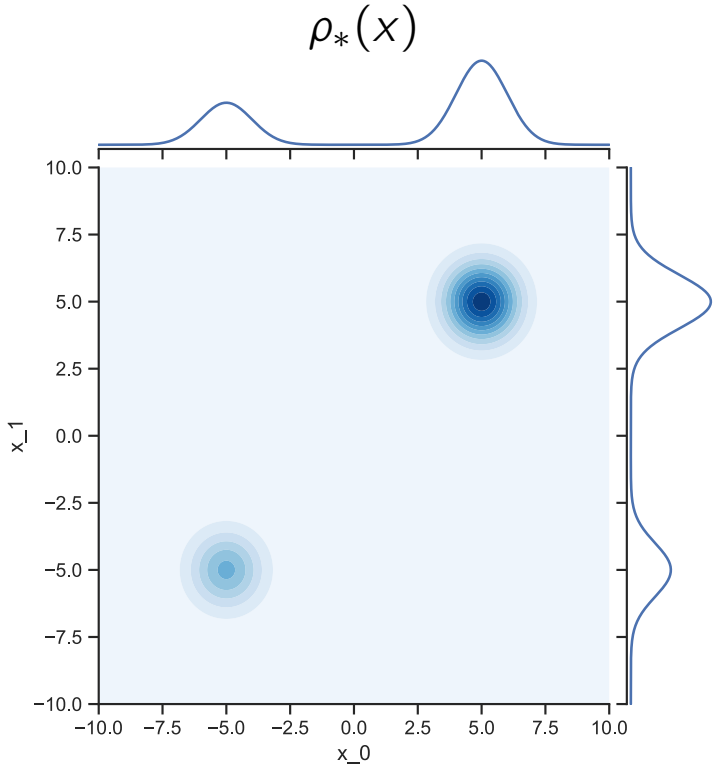
Q2: Training without extensive data?

2.2 Variational training of NFs for sampling

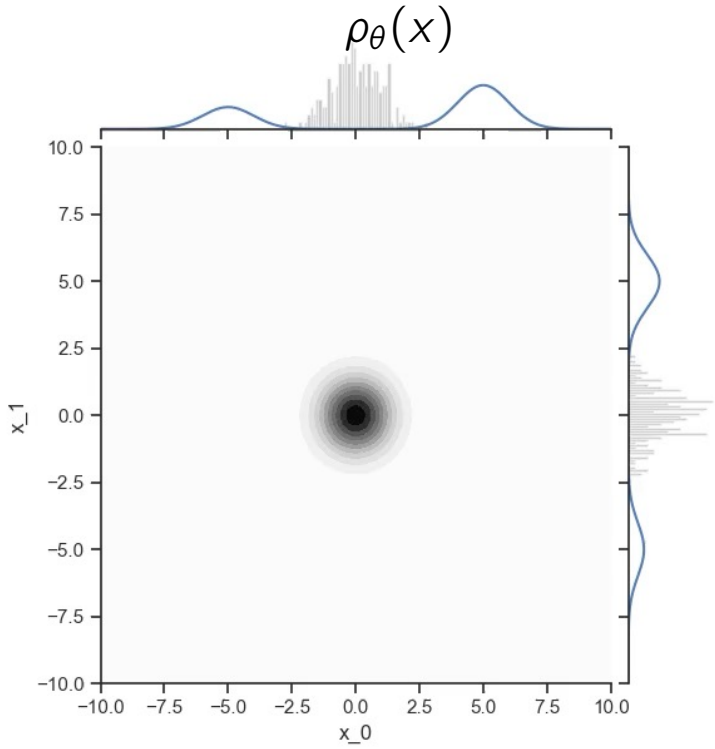
▷ A data-free learning objective: the (reverse) Kullback-Leibler divergence $D_{\mathrm{KL}}(\rho_\theta \| \rho_*)$

$$D_{\mathrm{KL}}(\rho_\theta \| \rho_*) = \int \log \frac{\rho_\theta(x)}{\rho_*(x)} \rho_\theta(x) \mathrm{d}x \approx \frac{1}{N} \sum_{i=1}^{N} \log \frac{\rho_B(z_i) \det |\nabla_{z_i} T_\theta|}{\rho_*(T_\theta(z_i))} \qquad z_i \sim \rho_B(z)$$

$\rho_*(x)$



$\rho_\theta(x)$

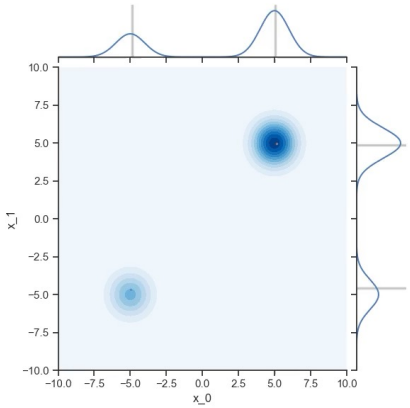


Mode collapse!
Adhoc fixes in these first papers (annealing and adding data!)

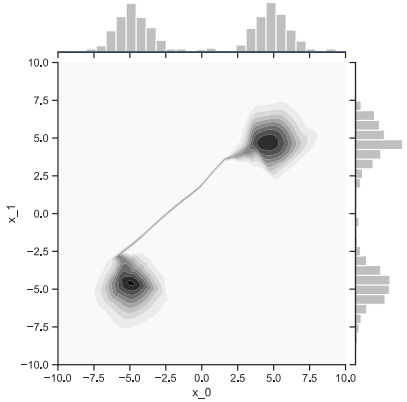
Weiss, P. (1907). L'hypothèse du champ moléculaire et la propriété ferromagnétique.
Rezende et al. ICML 205, Wu et al. PRL 2019, Albergo et al PRD 2019

2.3 Adaptive training and sampling: Learning a non-local proposal while sampling

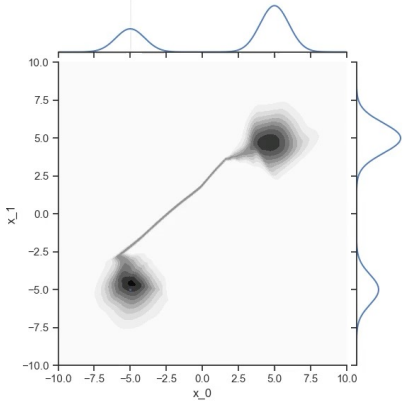
▷ Simultaneously training the flow and sampling by looping over 3 steps:



$$x^i_{t+1} \sim \pi_{\text{local}}(x^i_{t+1}|x^i_t)$$



$$\theta^* = \arg\max_\theta \sum_{i,t} \log \rho_\theta(x^i_t)$$

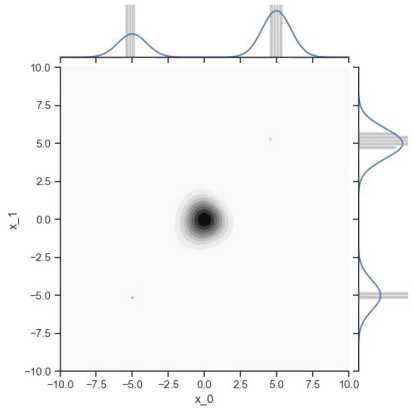


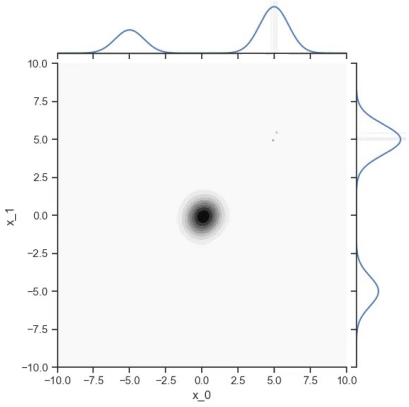
$$\rho_{\text{prop}}(x_{t+1}|x_t) = \rho_\theta(x_{t+1})$$

▷ FlowMC full algorithm adressing the key questions

Q1: Guarantees on the quality of the output? ✓

Q2: Training without extensive data? ✓





[MG, Rotskoff, Vanden-Eijnden PNAS 2022 , Parno & Marzouk 2018 SIAM Journal of Uncertainty; Naesseth et al. Neurips 2020]

No free lunch!

2.4 Computing free energy differences

- Once trained to approximate a target distribution $\rho_*(x) = \frac{e^{-\beta U_*(x)}}{\mathcal{Z}_*}$, transport-based generative models $\rho_\theta(x)$ can assist free energy estimation

- Idea behind targeted free energy perturbation (TFEP): Jarzynski, C. "Targeted Free Energy Perturbation."PRE 2002

$$F_* = -\frac{1}{\beta} \ln \mathcal{Z}_* = -\frac{1}{\beta} \ln \int e^{-\beta U_*(x)} \mathrm{d}x$$

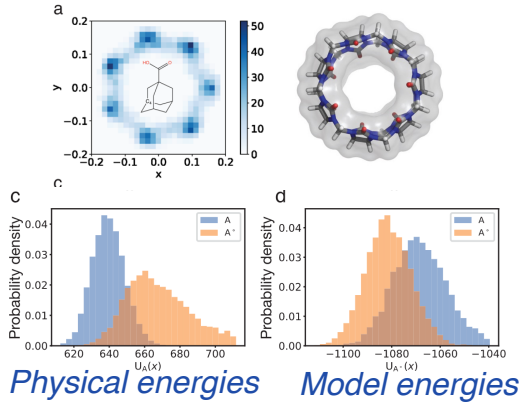
Very much à la importance sampling

NB: If the model is not normalized, we obtain a free energy difference between the model and the target.

- Benett Acceptance Ratio (BAR) calculations are related and have typically smaller variance.

- Proof of concepts for computing the free energy differences with NFs
  - between the target model and a simplified theory (used as a base)
  - between a crystal (used as a base) and the thermodynamical state at a given temperature
  - of binding events (using a NF for bound structure and one for unbound)

Wirnsberger et al."Targeted Free Energy Estimation via Learned Mappings." J. Chem. Phys. 2021.
Ding et al. "DeepBAR: A Fast and Exact Method for Binding Free Energy Computation", J. Chem. Phys. 2021
Jia et al, "Normalizing Constant Estimation with Gaussianized Bridge Sampling", AABI 2019 Proceedings



Ding et al. "DeepBAR", J. Chem. Phys. 2021

Summary for part 2

- The tractable likelihood of transport based models make it possible to insert them in Monte Carlo approximation strategies guaranteeing the quality of the outputs.
- Training is performed either using a data-free variational objective (but be careful for mode collapse) or a self-consistent loop of sampling and training.
- The trained model can also be employed to compute free energy differences.

Outline

1. Transport based generative models

1.1 Normalizing flows
1.2 Continuous normalizing flows
1.3 Diffusion models
1.4 Stochastic interpolants/flow matching

2. Generative models for sampling

2.1 NF Assisted sampling
2.2 Variational training
2.3 Adaptive training and sampling
2.4 Free energy computations

Case study: Sampling Ag_6 nanoclusters

3. Remaining challenges and ongoing research

Case study: Using CVs to sample

Outline

1. Transport based generative models

1.1 Normalizing flows
1.2 Continuous normalizing flows
1.3 Diffusion models
1.4 Stochastic interpolants/flow matching

2. Generative models for sampling

2.1 NF Assisted sampling
2.2 Variational training
2.3 Adaptive training and sampling
2.4 Free energy computations

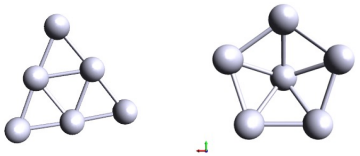
Case study: Sampling Ag_6 nanoclusters

3. Remaining challenges and ongoing research

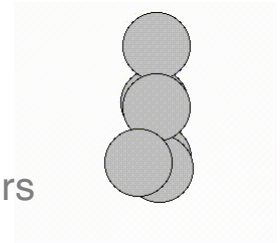
Case study: Using CVs to sample

An example of the pipeline put into practice: Sampling Ag_6 nanoclusters at quantum accuracy

- Multiple isomers of silver nanoclusters are credibly stable at room temperature



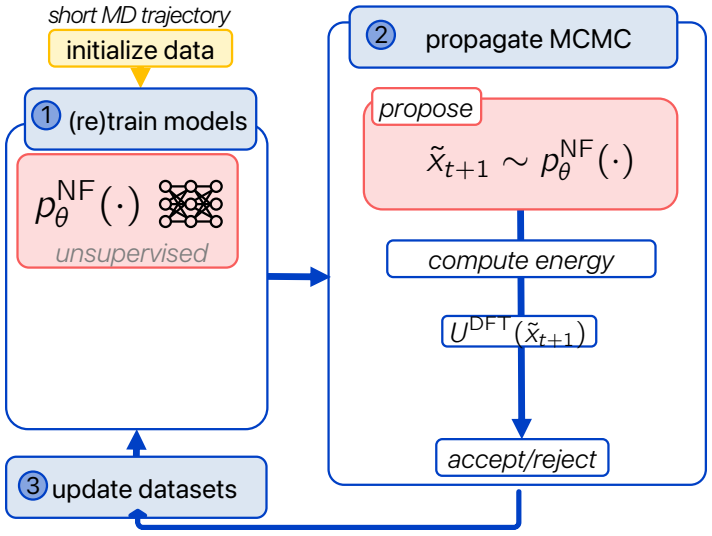
- Transitions between metastable states cannot be observed through molecular dynamics as they happen on long time scales



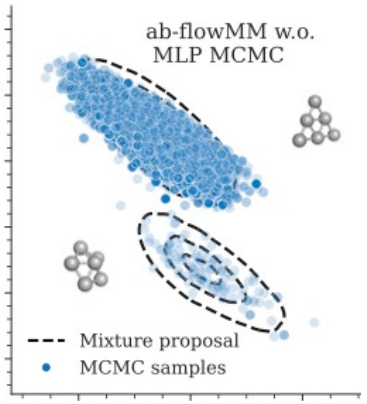
- We can use the flowMC pipeline to equilibrate an MCMC across isomers

ASE GPAW!

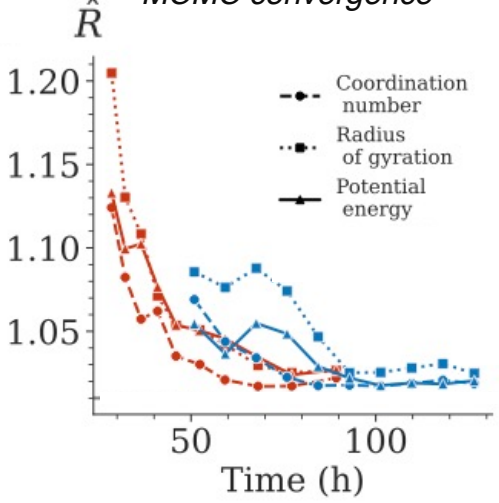
flowMC workflow



Projected samples



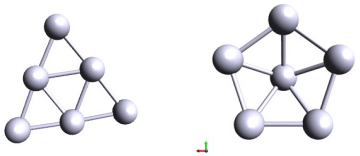
MCMC convergence



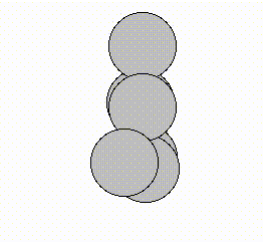
Molina-Taborda, Cossio, Lopez-Acevedo, MG. "Active Learning of Boltzmann Samplers and Potential Energies with Quantum Mechanical Accuracy." JCTC 2024

An example of the pipeline put into practice: Sampling Ag_6 nanoclusters at quantum accuracy

- Multiple isomers of silver nanoclusters are credibly stable at room temperature

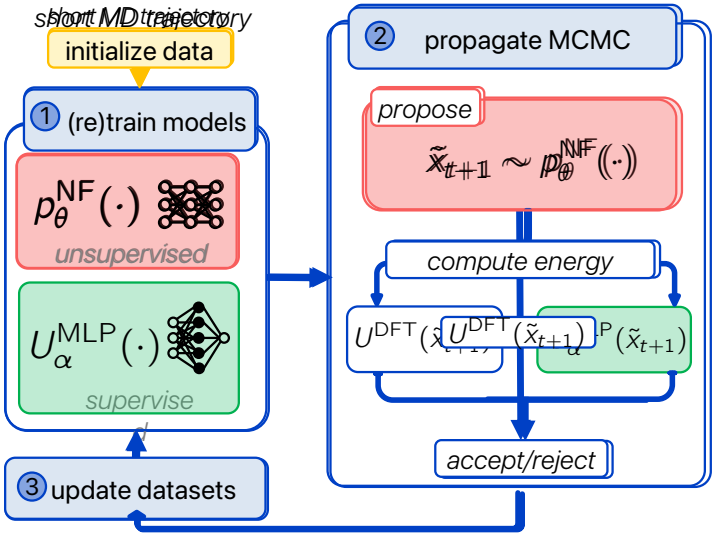


- Transitions between metastable states cannot be observed through molecular dynamics as they happen on long time scales

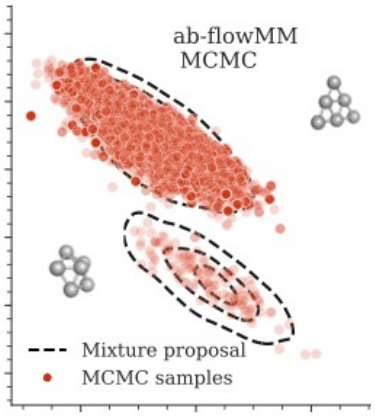


- We can also ease computational cost by adding a Machine Learning Potential regressing the DFT predictions

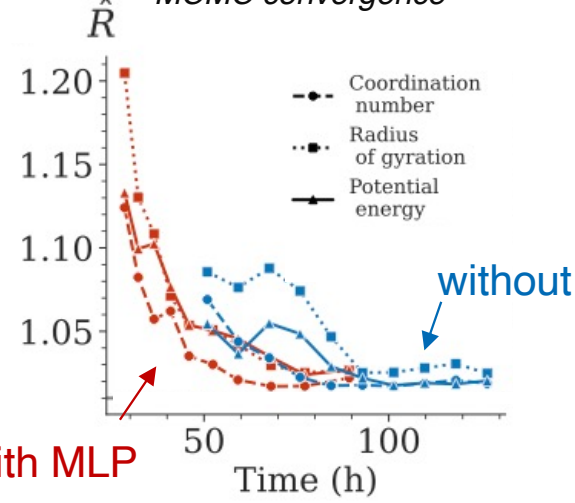
flowMC workflow



Projected samples



MCMC convergence



Molina-Taborda, Cossio, Lopez-Acevedo, MG. "Active Learning of Boltzmann Samplers and Potential Energies with Quantum Mechanical Accuracy." JCTC 2024

Outline

1. Transport based generative models

1.1 Normalizing flows
1.2 Continuous normalizing flows
1.3 Diffusion models
1.4 Stochastic interpolants/flow matching

2. Generative models for sampling

2.1 NF Assisted sampling
2.2 Variational training
2.3 Adaptive training and sampling
2.4 Free energy computations

Case study: Sampling Ag_6 nanoclusters

3. Remaining challenges and ongoing research

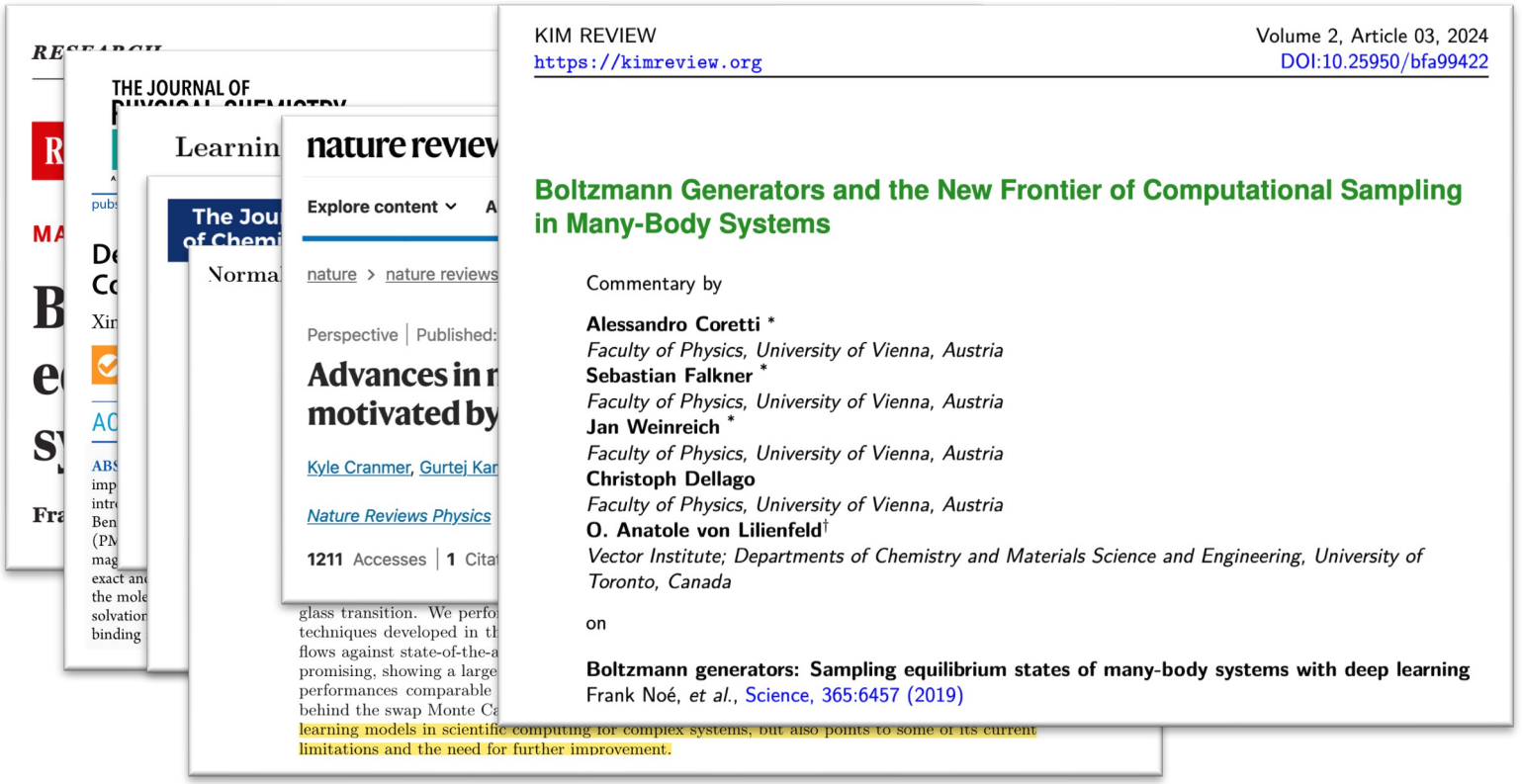
Case study: Using CVs to sample

State of affairs

▷ Proofs of concepts of using generative models for sampling physical systems

An incomplete selection

A recent review/opinion paper



▷ Limits arise when it becomes difficult to obtain $\rho_\theta(x) \approx \rho_*(x)$ through training, either because of the **complexity of the landscape** or because of the **dimension**

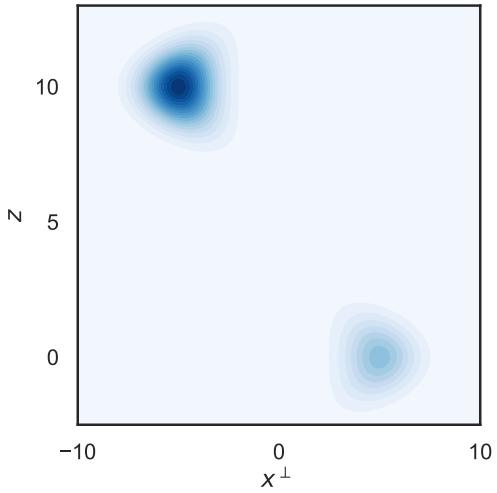
[Del Debbio et al PRD 2021; Ciarella et al MLST 2023; Grenioux, Durmus, Moulines & MG ICML 2023]

To overcome the curse of dimensionality,
combine learning with physics' knowledge!

Leveraging a low dimensional representation of metastability

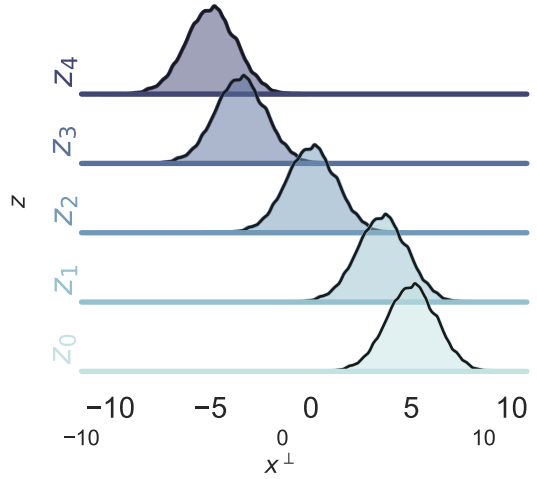
▷ Sometimes the measure can decompose into hard/easy d.o.f.

$$\rho_*(x) = \rho_*(z, x^\perp) = \rho_*(z)\rho_*(x^\perp|z)$$

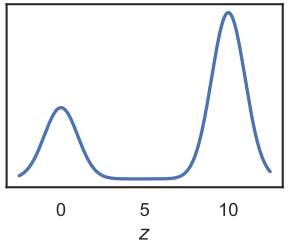


$\rho_*(x^\perp|z)$

High-dimensional but "easy" distributions



Multi-modal but lower dimensional



▷ Remains the marginal on the variable(s) describing the metastability (a.k.a. Collective Variables)

$$\rho_*(z) = \int \mathrm{d}x^\perp \rho(z, x^\perp) \propto e^{F(z)}$$

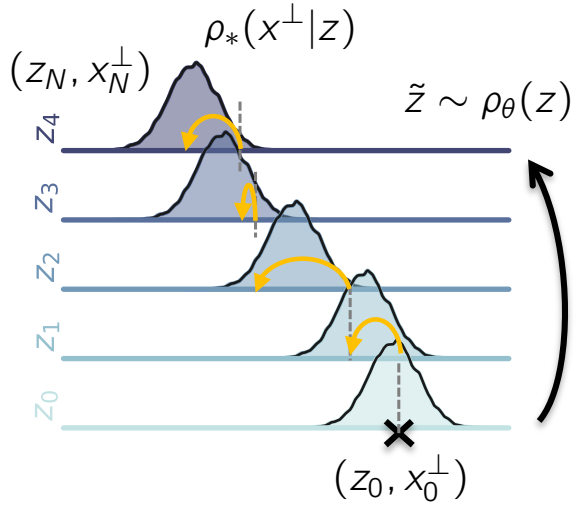
▷ Intuition: focus efforts on low-dimensional problem (use ML!)

▷ Issues: the marginal is intractable so we cannot directly derive a sampling algorithm for it

Non-equilibrium path proposals & accept/reject

Super detailed-balance [Frenkel, PNAS 2001], Residence weight algorithm [Athènes PRE 2001], Non-equilibirum candidate Monte Carlo [Nilmeier et al PNAS 2011], [Chen & Roux, JChemPhys 2015]

- ▷ Create a jump towards a newly proposed $\tilde{z} \sim \rho_\theta(z)$
- ▷ Discrete N steps schedule for CV (e.g. linear interpolation)
  $z_n = z_0 + n\dfrac{(\tilde{z} - z_0)}{N}$
  $(z_N = \tilde{z})$
- ▷ Propose relaxation-steps for the rest recursively $x_n^\perp \sim \mathcal{K}(x_n^\perp | z_n, x_{n-1}^\perp)$
  Relaxation kernel
- ▷ Apply the accept/reject criteria
  - Path-proposal probability density $\rho_\theta(\tilde{z}) \prod_{n=0}^{N} \mathcal{K}(x_n^\perp | z_n, x_{n-1}^\perp)$
  - Considering the same path backward, it as probability $\rho_\theta(z_0) \prod_{n=N}^{0} \mathcal{K}(x_{n-1}^\perp | z_{n-1}, x_n^\perp)$
  - Acceptance criteria sufficient for getting to the invariant distribution $\rho_*(x) = \dfrac{1}{\mathcal{Z}_\beta} e^{-\beta U(x)}$

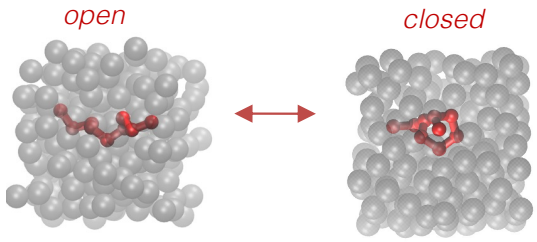


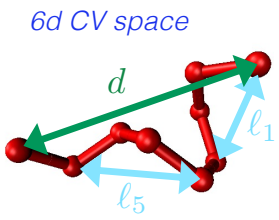
$$\mathrm{acc}(Z, X^\perp | x_0) = \min\left[1\,,\ \frac{\rho^*(x_N)}{\rho^*(x_0)}\frac{\rho_\theta(z_0)}{\rho_\theta(z_N)}\frac{\prod_n \mathcal{K}(x_{n-1}^\perp | z_{n-1}, x_n^\perp)}{\prod_n \mathcal{K}(x_n^\perp | z_n, x_{n-1}^\perp)}\right]$$

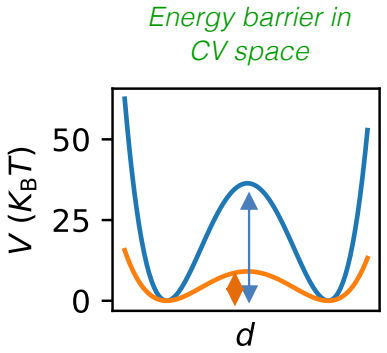
[Tamagnone, Laio, MG, "Coarse-Grained Molecular Dynamics with Normalizing Flows", JCTC 2024,
Schönle, MG, Lelièvre, Stoltz "Sampling metastable systems using collective variables and Jarzynski-Crooks paths", 2405.18160]

How to learn the coarsed-grained proposal: adaptively!

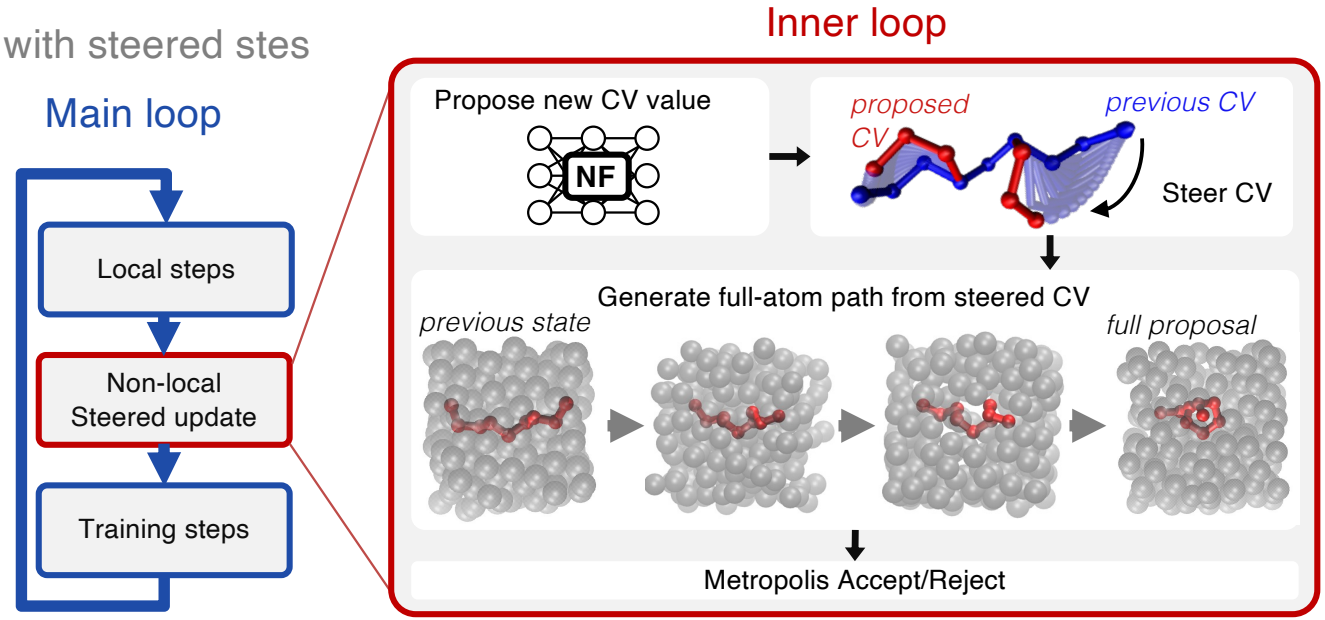
- A test system of a 9-bead polymer in solvent with two stabilized states:





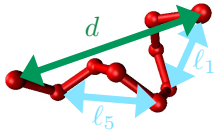


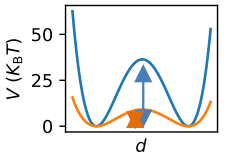
- The adaptive loop with steered stes



[Tamagnone, Laio, MG, "Coarse-Grained Molecular Dynamics with Normalizing Flows", JCTC 2024]

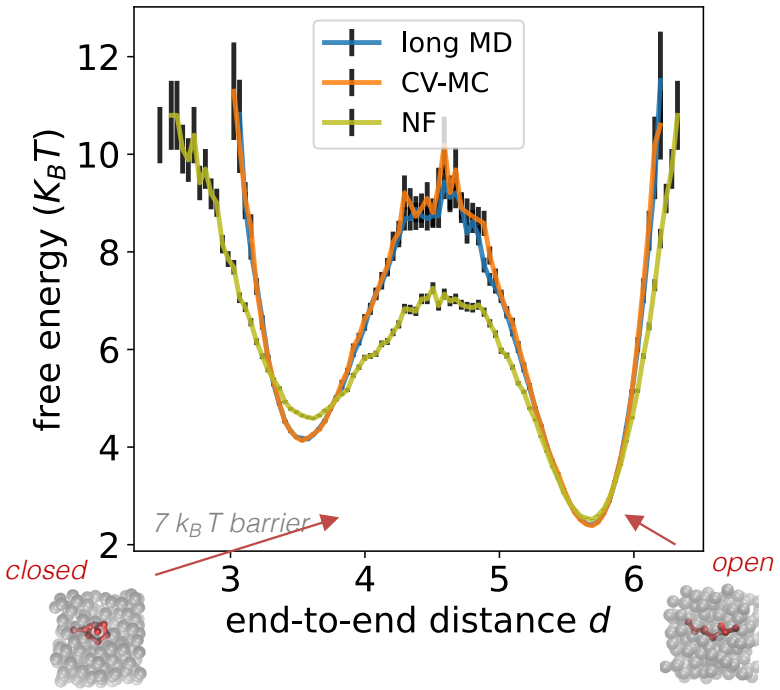
Promising results for this hybrid ML-CV approach



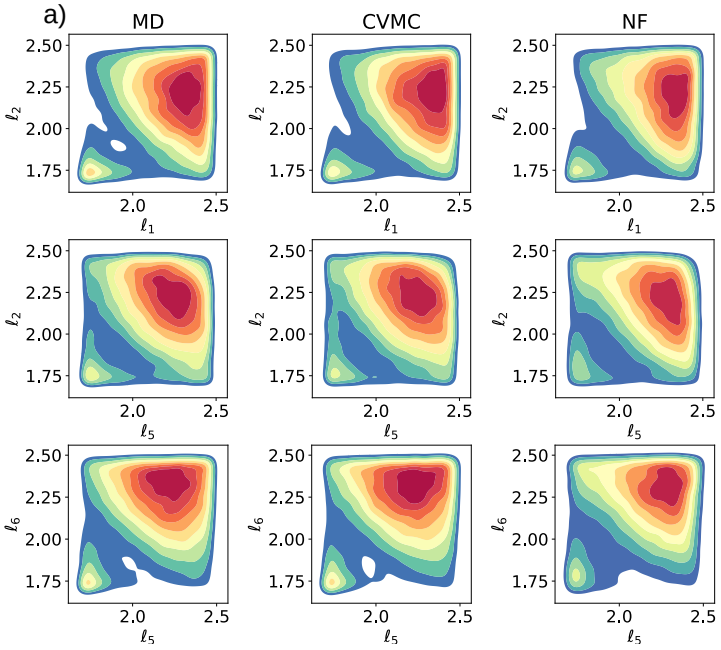


▷ Produces converged free energies in CV space with 25% acceptance of non-local moves (35,000 steps in steered path dependending weakly on the energy barrier in CV)

Free energy projection 1d



Free energy projections 2d



[Tamagnone, Laio, MG, "Coarse-Grained Molecular Dynamics with Normalizing Flows", JCTC 2024]

Perspective/Conclusion

- Progess in generative modelling suggests a road to powerfull **machine learning enhanced samplers** which can speed up the simulation of equilibrium thermodynamics (or Bayesian posteriors)!

- These methods appear to be efficient “all-purpose” samplers for problems in moderate dimension.

- Reaching the level of training accuracy required for complex systems such as proteins is not trivial.

- **Combining machine learning & physics** will allow us to fully explore the potential of these approaches.

Thank you!

Main collaborators

Grant Rotskoff (Stanford)

Eric Vanden-Eijnden (Courant Institute, NYU)

Christoph Schönle (École Polytechnique)

Gabriel Stoltz & Tony Lelièvre (École Nationale des Ponts & Chaussés)

Ana Molina Taborda, Olga Lopez Acevedo (Universidad de Antioquia)

Pilar Cossio (Flatiron, CCM)

Samuel Tamagnone & Alessandro Laio (SISSA)