DeepMind

Generative models, manifolds and symmetries: tools and applications

Danilo J. Rezende

Arnold Sommerfeld School summer school, Munich, 2022

https://danilorezende.com/slides/

Abstract

The study of symmetries in physics has revolutionized our understanding of the world. Inspired by this, the development of methods to incorporate internal (Gauge) and external (space-time) symmetries into machine learning models is a very active field of research. We will discuss general methods for incorporating symmetries in ML, and our work on invariant generative models. We will then present its applications to quantum field theory on the lattice (LQFT) and molecular dynamics (MD) simulations. In the MD front, we'll talk about how we constructed permutation and translation-invariant normalizing flows on a torus for free-energy estimation. In the LQFT front, we'll present our work that introduced the first U(N) and SU(N) Gauge-equivariant normalizing flows for pure Gauge simulations and its extension to incorporate "pseudo-fermions", leading to the first proof of principle of a full QCD simulation with normalizing flows in 2D.

Outline

Monday

Tools, building blocks

- Flows
- Model corrections
- Models on Manifolds
- Incorporating Symmetries

Tuesday

Applications

- Free-energy estimation in solids
- Lattice-QFT
  - Yukawa
  - Schwinger
  - QCD

DeepMind

1 Tools

The problem

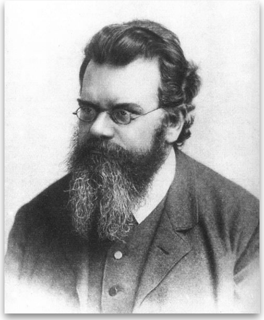


Image credit:
[Wikipedia]

inverse temperature

energy

coordinates

$$p(x) = \frac{e^{-\beta U(x)}}{Z}$$

normalizing constant

$$Z = \int \mathrm{d}x \; e^{-\beta U(x)}$$

The problem

$$p(x) = \frac{e^{-\beta U(x)}}{Z}$$

coordinates → x; inverse temperature → β; energy → U; normalizing constant → Z

Goals

1) Draw samples
2) Expectations under $p(x)$ $\Longrightarrow$ **Material properties**
3) Estimate Z $\Longrightarrow$ **Free energy**

Problem summary

We are **given an energy function** with known invariances…

$$U(x)$$

… that defines a Boltzmann distribution …

$$p(x) = \frac{e^{-\beta U(x)}}{Z}$$

… under which **we want to compute expectations and free energies.**

$$\langle O \rangle = \mathbb{E}_{p(x)}\left[O(x)\right]$$

DeepMind

Normalizing flows

Why normalizing flows?

- Scientific applications require *high-accuracy predictions with reliable error estimation*
- Model de-biasing methods (e.g. IS, MCMC) require fast *exact model likelihoods and sampling*
- This excludes certain families of generative models such as GANs, energy-based and diffusion models
- Auto-regressive, latent variable and flow models are compatible with the desiderata of MCMC corrections

Change of variable formula

Our goal is to define a density $q(\mathbf{x})$ over a *D*-dimensional vector $\mathbf{x}$.

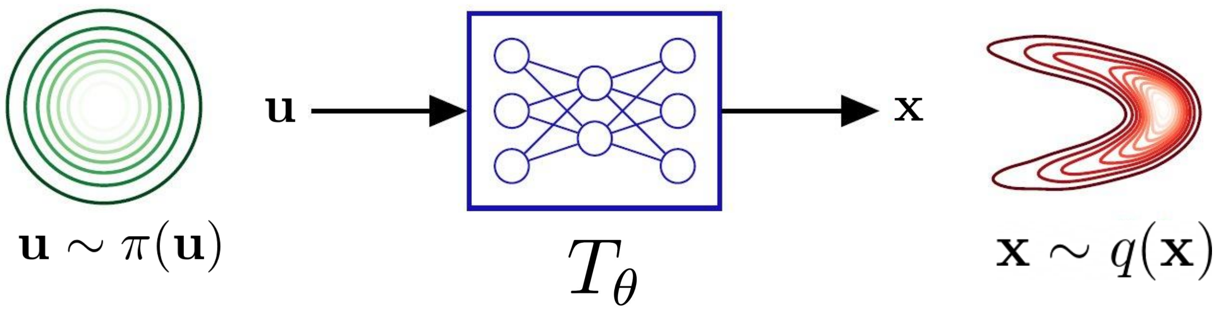
We can achieve this by transforming samples from a **base distribution** $\mathbf{u} \sim \pi(\mathbf{u})$,

$$q(\mathbf{x}) = \pi(\mathbf{u})\left|\det J_T(\mathbf{u})\right|^{-1}$$

where T is an invertible transformation and $\mathbf{x} = T(\mathbf{u})$.

Basic concept of NFs

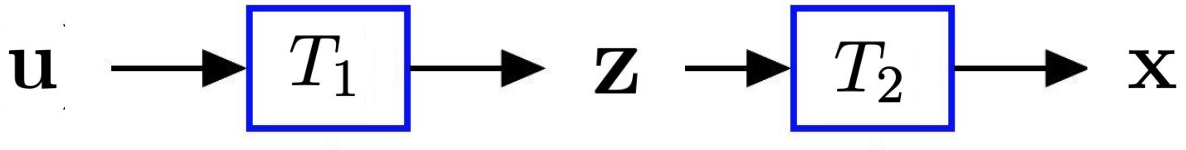
Goal: Use ML to transform a simple base density into a complex density.



We assume the transformation to be a diffeomorphism with tractable Jacobian determinant.

Rezende and Mohamed, *Variational inference with normalizing flows*, ICML (2015)
Papamakarios *et al.*, *Normalizing flows for probabilistic modeling and inference*, JMLR *(2021)*
Kobyzev, Prince and Brubaker, *Normalizing Flows: An Introduction and Review of Current Methods*, IEEE PAM (2021)

NFs are composable



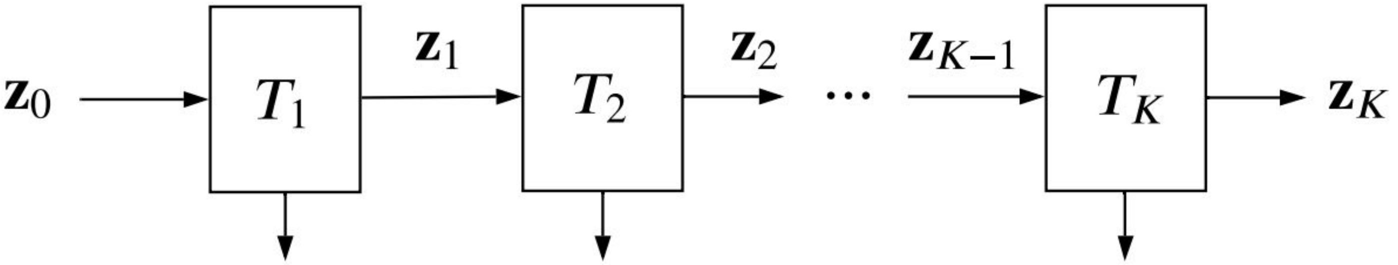
$$T_2 \circ T_1$$ ← is also a flow

$$(T_2 \circ T_1)^{-1} = T_1^{-1} \circ T_2^{-1}$$

$$\det J_{T_2 \circ T_1}(\mathbf{u}) = \det J_{T_2}(T_1(\mathbf{u})) \cdot \det J_{T_1}(\mathbf{u})$$

Composing multiple layers

$$T = T_K \circ \cdots \circ T_1$$



$$\log|\det J_{T_1}(\mathbf{z}_0)| + \log|\det J_{T_2}(\mathbf{z}_1)| + \cdots + \log|\det J_{T_K}(\mathbf{z}_{K-1})| = \log|\det J_T(\mathbf{z}_0)|$$

Papamakarios *et al., Normalizing flows for probabilistic modeling and inference, JMLR (2021)*

Composing multiple layers

$$T = T_K \circ \cdots \circ T_1$$

$$\log |\det J_T(\mathbf{z}_0)| = \log \left| \prod_{k=1}^{K} \det J_{T_k}(\mathbf{z}_{k-1}) \right| = \sum_{k=1}^{K} \log |\det J_{T_k}(\mathbf{z}_{k-1})|$$

Papamakarios *et al., Normalizing flows for probabilistic modeling and inference,* *JMLR* *(2021)*

Density evaluation and sampling

Sampling $$\mathbf{x} = T(\mathbf{u}) \qquad \mathbf{u} \sim \pi(\mathbf{u})$$

Density of samples $$q(\mathbf{x}) = \pi(\mathbf{u})|\det J_T(\mathbf{u})|^{-1}$$

Forward & Jacobian determinant

Arbitrary density $$q(\mathbf{x}) = \pi\left(T^{-1}(\mathbf{x})\right)|\det J_{T^{-1}}(\mathbf{x})|$$

Inverse & Jacobian determinant

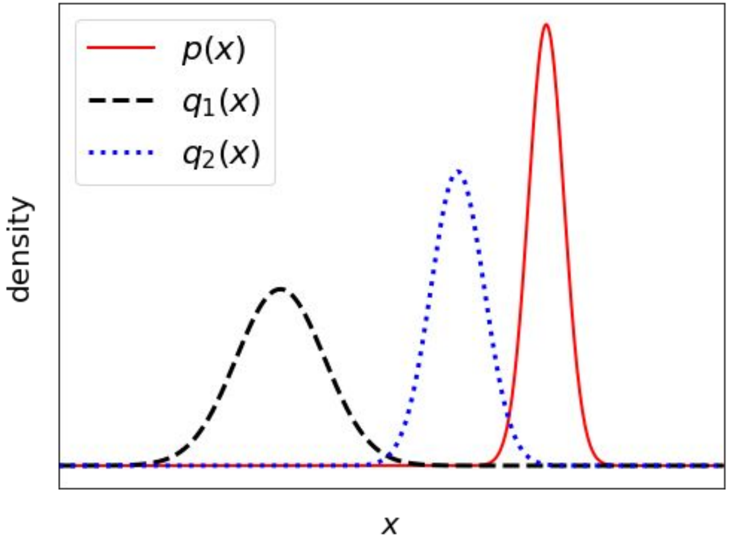
Slide credit: George Papamakarios

DeepMind

What should models optimize for?

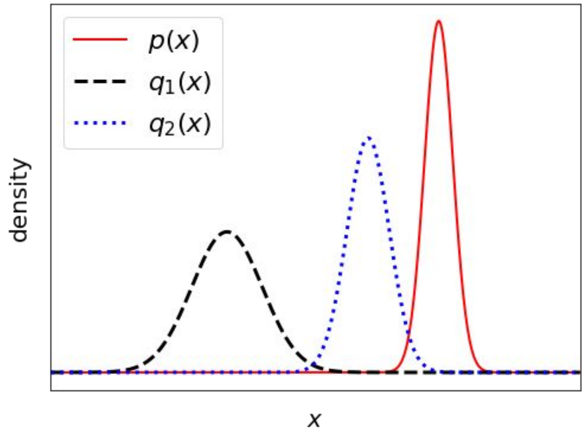
Training objective

- Our goal is to fit a model q_θ to an (unnormalised) target distribution p^*.
- How to quantify the difference between the distributions below?



Divergence

- We need a loss function $\mathcal{L}$ that quantifies the *divergence* or discrepancy between the two distributions q_θ and p^*.
- We can then update the parameters θ based on the gradient $\frac{\partial \mathcal{L}}{\partial \theta}$ (e.g. with Adam).



Kingma and Ba, Adam: A method for stochastic optimization, ICLR (2015)

Kullback–Leibler divergences

The **Kullback–Leibler (KL) divergences** are popular choices for training flows:

- *forward-KL:* $$D_{\mathrm{KL}}(p \,\|\, q) = \mathbb{E}_{p(x)}\left[\log p(x) - \log q(x)\right]$$
- *reverse-KL:* $$D_{\mathrm{KL}}(q \,\|\, p) = \mathbb{E}_{q(x)}\left[\log q(x) - \log p(x)\right]$$

$D_{\mathrm{KL}}(p \,\|\, q) \geq 0$ with equality if and only if the two distributions are equal.

Papamakarios *et al., Normalizing flows for probabilistic modeling and inference, JMLR (2021)*

Forward KL

The forward KL can be rewritten as

$$
\begin{aligned}
D_{\mathrm{KL}}(p \,\|\, q_\theta) &= \mathbb{E}_{p(x)}\left[\log p(x) - \log q_\theta(x)\right] \\
&= -\mathbb{E}_{p(x)}\left[\log q_\theta(x)\right] + \text{const} \\
&= -\mathbb{E}_{p(x)}\left[\log \pi\left(T_\theta^{-1}(x)\right) + \log\left|\det J_{T_\theta^{-1}}(x)\right|\right] + \text{const}
\end{aligned}
$$

= Maximum likelihood estimation

Papamakarios *et al., Normalizing flows for probabilistic modeling and inference, JMLR* (2021)

Forward KL

Given samples $\{x_n\}_{n=1}^N$ from p^* we can use the MC approximation

$$D_{\mathrm{KL}}(p \,\|\, q_\theta) \approx -\frac{1}{N}\sum_{n=1}^{N}\left[\log \pi\left(T_\theta^{-1}(x_n)\right) + \log\left|\det J_{T_\theta^{-1}}(x_n)\right|\right] + \text{const}$$

We can evaluate this for a batch of samples to approximate $\dfrac{\partial \mathcal{L}}{\partial \theta}$

Independent of the parameters.

Papamakarios *et al., Normalizing flows for probabilistic modeling and inference,* *JMLR* (2021)

Reverse KL

The reverse KL can be rewritten as

$$p^*(x) = Zp(x) = e^{-\beta U(x)}$$

$$D_{\mathrm{KL}}(q_\theta \parallel p^*) = \mathbb{E}_{q_\theta(x)}\left[\log q_\theta(x) - \log p^*(x)\right]$$

$$= \mathbb{E}_{q_\theta(x)}\left[\log q_\theta(x) + \beta U(x)\right]$$

$$= \mathbb{E}_{\pi(u)}\left[\log \pi(u) - \log\left|\det J_{T_\theta}(u)\right| + \beta U(T_\theta(u))\right]$$

energy

reparametrization

Papamakarios *et al., Normalizing flows for probabilistic modeling and inference,* JMLR *(2021)*

Reverse KL

Given samples $\{u_n\}_{n=1}^{N}$ from the base $\pi(u)$ we can approximate this KL as

$$D_{\mathrm{KL}}(q_\theta \parallel p^*) \approx -\frac{1}{N}\sum_{n=1}^{N}\left[\log \pi(u_n) - \log\left|\det J_{T_\theta}(u_n)\right| + \beta U(T_\theta(u_n))\right]$$

We can evaluate this for a batch of samples to approximate $\dfrac{\partial \mathcal{L}}{\partial \theta}$

Papamakarios *et al., Normalizing flows for probabilistic modeling and inference*, JMLR *(2021)*

KL Comparison

Forward KL

- "Training by sample"
- requires samples from the target
- "mode covering"

Reverse KL

- "Training by energy"
- requires an energy function
- "mode seeking"

Training objectives can be combined

Papamakarios *et al., Normalizing flows for probabilistic modeling and inference, JMLR (2021)*
Noé et al., *Boltzmann generators: Sampling equilibrium states of many-body systems with deep learning,* Science (2019)
Wirnsberger et al., *Targeted free energy estimation via normalizing flows,* JCP (2020)

DeepMind

Correcting models: reliable predictions with (some) guarantees

Goal:

Guarantees of correctness (controlled systematic errors)

Model inaccurate ⟶ Results correct but slow

Model accurate ⟶ Results correct and fast

Debiasing expectations

The goal is to compute expectations w.r.t. the target:

$$\langle O \rangle = \mathbb{E}_{p(x)}\left[O(x)\right]$$

Averages under the model are, in general, biased:

$$\mathbb{E}_{q_\theta(x)}\left[O(x)\right] \neq \mathbb{E}_{p(x)}\left[O(x)\right]$$

Common strategies: ***Metropolis–Hastings MCMC*** *and* ***Importance Sampling***

Metropolis–Hastings MCMC

Generate a new trial move using the flow: (independent proposal distribution)

$$y \sim q_\theta(y)$$

Accept $x \to y$ with probability:

$$\mathrm{acc}(x, y) = \min\left[1, \frac{p^*(y) q_\theta(x)}{p^*(x) q_\theta(y)}\right]$$

Using the collected samples $\{x_n\}_{n=1}^{N}$ estimate expectations as:

$$\langle O \rangle \approx \hat{O} = N^{-1} \sum_{i=1}^{N} O(x_i)$$

Albergo, Kanwar and Shanahan, *Flow-based generative models for Markov chain Monte Carlo in lattice field theory*, PRD (2019).
Nicoli et al., *Asymptotically Unbiased Estimation of Physical Observables with Neural Samplers*, PRE (2020).
Gabrié, Rotskoff and Vanden-Eijnden, *Adaptive Monte Carlo augmented with normalizing flows*, PNAS (2022).

Importance Sampling (IS)

The model is the proposal distribution: q_θ

Draw *N* samples from the model: $\{x_n\}_{n=1}^N$

Compute importance weights: $w_n = \frac{p^*(x_n)}{q_\theta(x_n)}$

Estimate expectations as: $\langle O \rangle_p \approx \frac{\sum_{n=1}^N w_n O(x_n)}{\sum_{n=1}^N w_n}$

Effective sample size for IS

IS estimate:

Assume we want to estimate:

$$I = \int \mathrm{d}x \; h(x) p(x)$$ th

MC estimate:

$$\hat{I} = \frac{1}{N} \sum_{n=1}^{N} h(x_n) \qquad x_n \sim p(x)$$

$$\tilde{I} = \sum_{n=1}^{N} \bar{w}_n h(x_n) \qquad x_n \sim q(x)$$

$$\bar{w}_n = \frac{w_n}{\sum_{n=1}^{N} w_n}$$

Martino et al. *Effective Sample Size for Importance Sampling based on discrepancy measures*, Signal Processing (2017).

Effective sample size for IS

The effective sample size (ESS) is defined as:

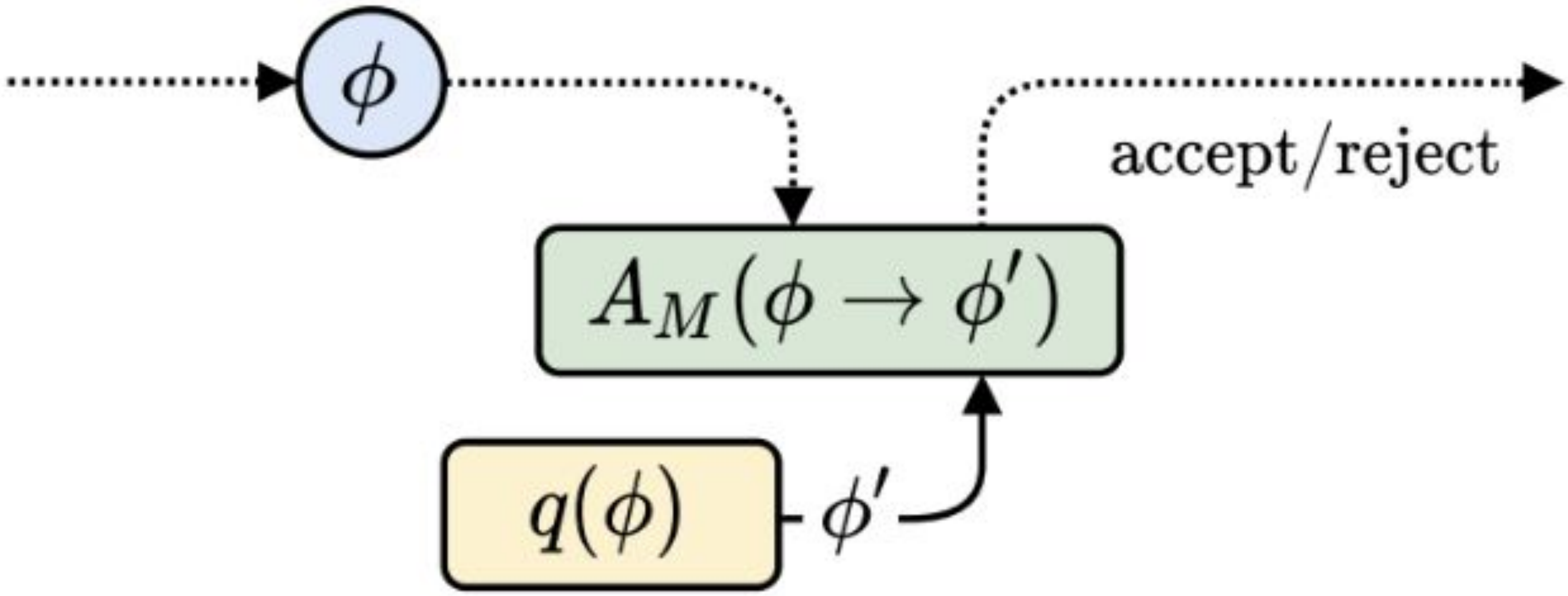
$$ESS = N \frac{\mathrm{Var}_p[\hat{I}]}{\mathrm{Var}_q[\tilde{I}]}$$

It can be estimated as:

$$\widehat{ESS} = \frac{1}{\sum_{n=1}^{N} \bar{w}_n^2} = \frac{\left(\sum_{n=1}^{N} w_n\right)^2}{\sum_{n=1}^{N} w_n^2}$$

Martino et al. *Effective Sample Size for Importance Sampling based on discrepancy measures,* [Signal Processing] (2017).

Sampling with flows: simple accept/reject bias correction



Albergo, M.S., Kanwar, G. and Shanahan, P.E., 2019. Flow-based generative models for Markov chain Monte Carlo in lattice field theory. Physical Review D, 100(3), p.034515.

DeepMind

Continual Repeated Annealed Flow Transport Monte Carlo (CRAFT)

Continual Repeated Annealed Flow Transport Monte Carlo (CRAFT)

Alex Matthews

DeepMind

Michael Arbel

INRIA

Univ. Grenoble Alpes

Danilo J. Rezende

DeepMind

Arnaud Doucet

DeepMind

github.com/deepmind/annealed_flow_transport

Slide credit: Alex Matthews

Continual Repeated Annealed Flow Transport Monte Carlo (CRAFT)

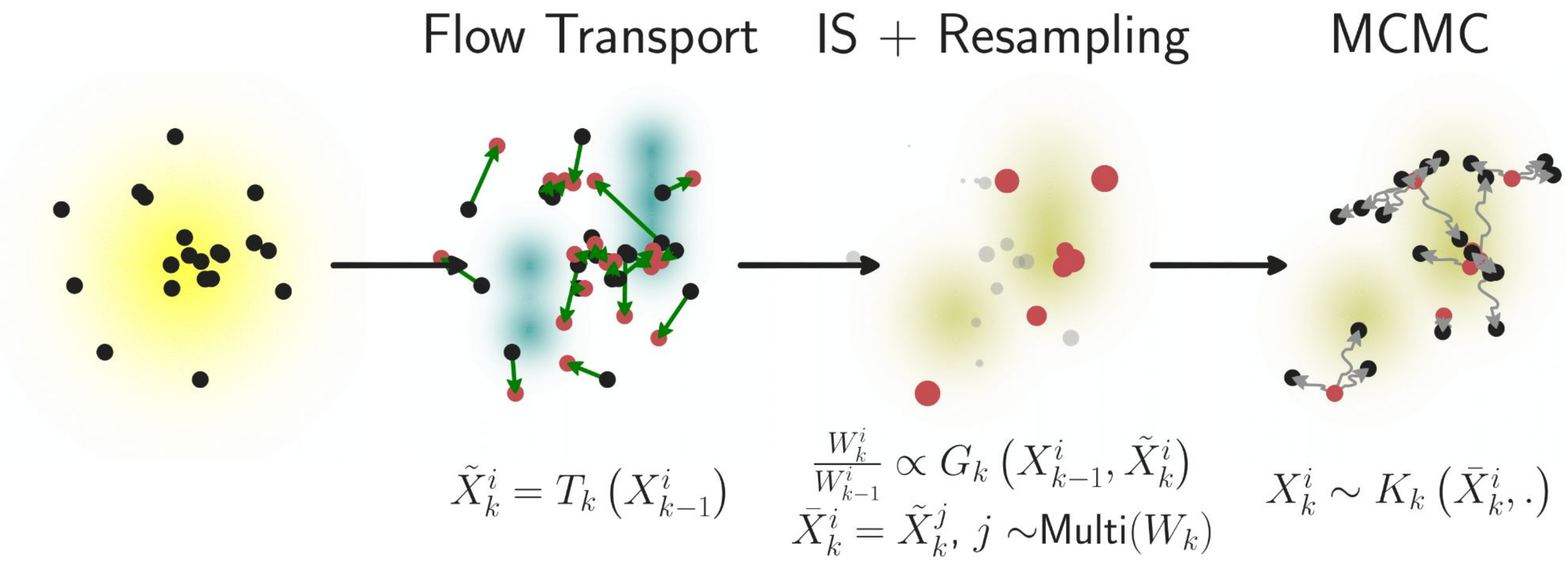


Image credit: Alex Matthews

Matthews, A.G., Arbel, M., J. Rezende, D. and Doucet, A., 2022. Continual Repeated Annealed Flow Transport Monte Carlo

CRAFT flow training objective

Previous distribution passed through a flow *T*

Current distribution

Zero if flow transport perfect.

$$H = \sum_{l=1}^{K} \mathcal{KL}[T_{\#}^{(l)} \pi_{l-1} || \pi_l]$$

Sum over transitions between temperatures.

Estimate objective and gradients using current importance sampling estimate.

Gradients local to transitions so no need to backprop through discrete steps.

Further analysis in paper.

Slide credit: Alex Matthews

CRAFT for Phi4 Lattice-QFT

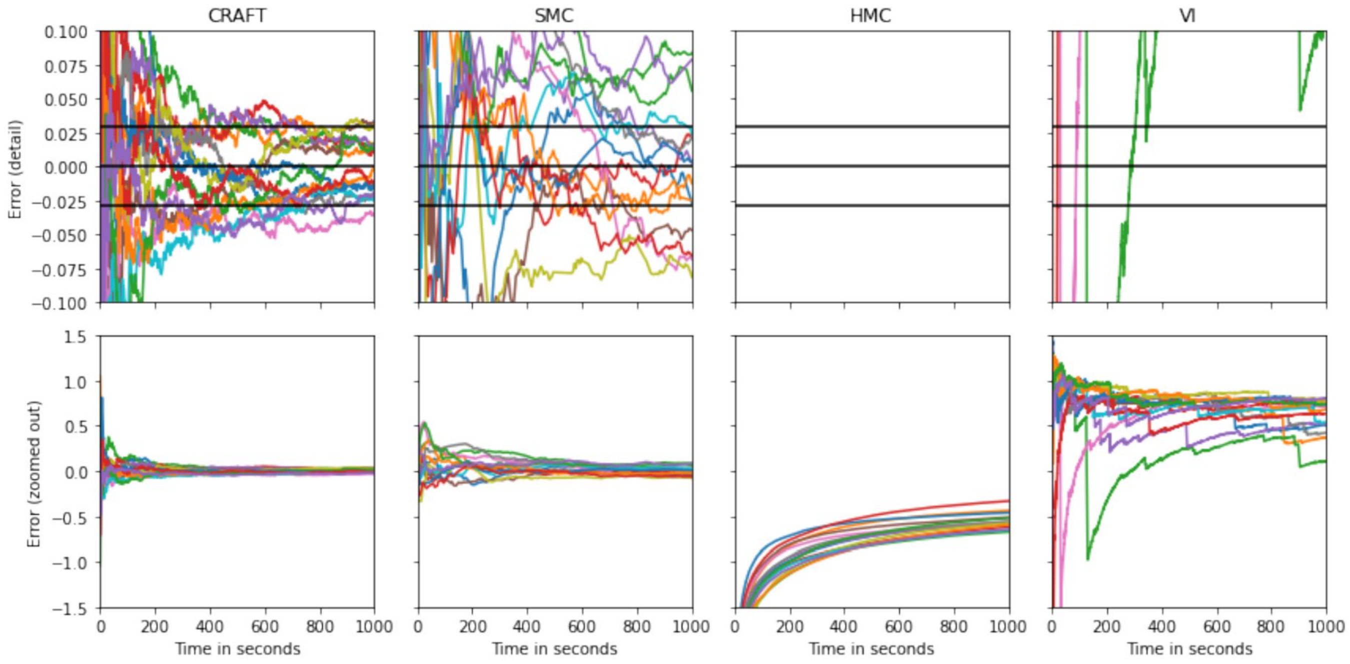


Figure 3. Timed comparison of MCMC methods for the ϕ^4 example, based on fifteen repeats. CRAFT, SMC and VI serve as proposal mechanisms for Particle MCMC. HMC is applied directly to the target. Error is in estimating two point susceptibility, an example, physically relevant expectation. Note HMC never reaches the detailed level of error in the top row.

Matthews, A.G., Arbel, M., J. Rezende, D. and Doucet, A., 2022. Continual Repeated Annealed Flow Transport Monte Carlo

DeepMind

2 Manifold Constraints

General principles

- Most existing ML techniques and tools assume data leaves in R^n and cannot be adapted in a straightforward way to manifolds
- There are very few tools that can be broadly applied to manifold data
- Solutions need to be custom-made for each problem in general

Continuous-time flows on Riemannian manifolds

$$\frac{d\mathbf{z}_t}{dt} = g_\phi(t, \mathbf{z}_t)$$

$$\ln p_t(x_t) = \ln p_0(x_0) - \int_0^t ds \, |G(x_s)|^{-\frac{1}{2}} \operatorname{tr} \frac{\partial}{\partial x_s} \sqrt{|G(x_s)|} g(s, x_s)$$

Metric matrix on local chart

Grathwohl, W., Chen, R.T., Bettencourt, J., Sutskever, I. and Duvenaud, D., 2018. Ffjord: Free-form continuous dynamics for scalable reversible generative models. arXiv preprint arXiv:1810.01367.

Mathieu, E. and Nickel, M., 2020. Riemannian continuous normalizing flows. arXiv preprint arXiv:2006.10605.

Cost-Concave Potential Flows on Riemannian manifolds

$$\phi : \mathcal{M} \to \mathbb{R}$$

Intrinsic distance

\phi is cost-concave

$$\phi(x) := \inf_{y \in \mathcal{M}} \left[\frac{1}{2} d(x,y)^2 + \psi(y) \right]$$

Diffeomorphism

$$f : \mathcal{M} \to \mathcal{M}$$

$$f(x) := \exp_x(-\nabla \phi(x))$$

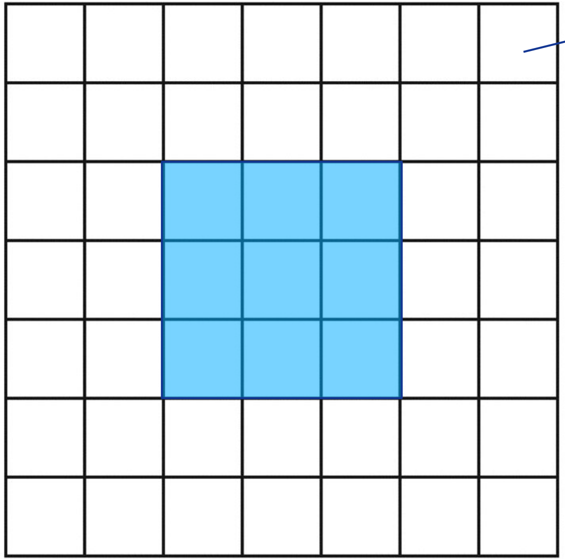
Villani, C., 2009. Optimal transport: old and new (Vol. 338, p. 23). Berlin: Springer.

DeepMind

Convnets on manifolds and fiber bundles: A general solution, formulation

Convnets on manifolds and fiber bundles

Convnets in R^n

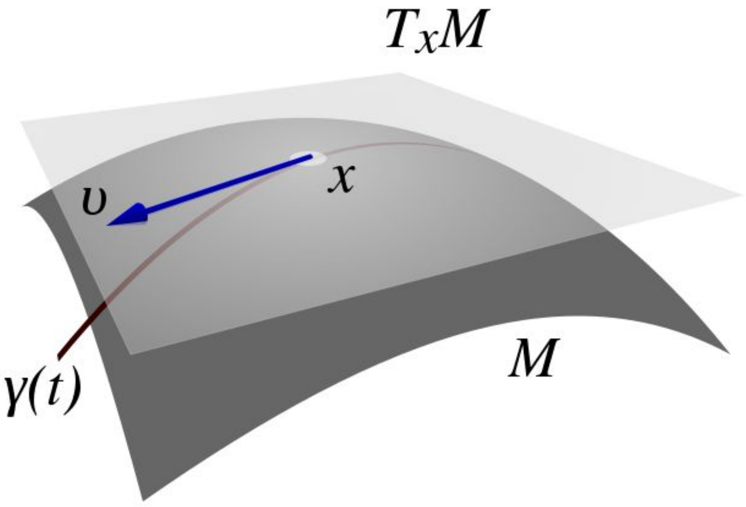


$$X_\nu(y)$$

$$Y_\mu(x) = \sum_{y \in N(x), \nu} w_{\mu,\nu}(x,y) X_\nu(y)$$

Convnets on manifolds and fiber bundles

We can't trivially extend convnets to fiber bundles: Linear combinations of elements belonging to different fibers are neither invariant nor equivariant



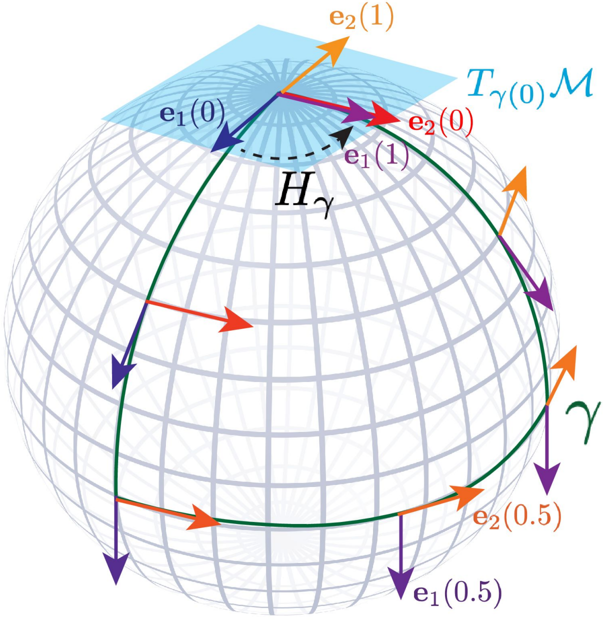


Image credit: Disentangling by Subspace Diffusion, Pfau et al, arxiv 2020

Convnets on manifolds and fiber bundles

General solution: Elements of different fibers need to be "parallel transported" to a "common fiber" before taking linear combinations

$$Y_\mu(x) = \sum_{y \in N(x)} w(x,y)[\Gamma(x,y) \circ X(y)]_\mu$$

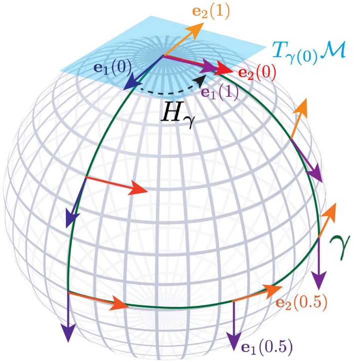


Image credit: Disentangling by Subspace Diffusion, Pfau et al, arxiv 2020

DeepMind

3 Symmetry Constraints

General principles

Given a group...

... and a map $g \in G$

... with group action $f : \mathcal{A} \to \mathcal{B}$

T_g

Invariance

$$f \circ T_g = f$$

Equivariance

$$f \circ T_g = T_g \circ f$$

Pitts, *Nominal sets: Names and symmetry in computer science*, Cambridge Univ. Press (2013).

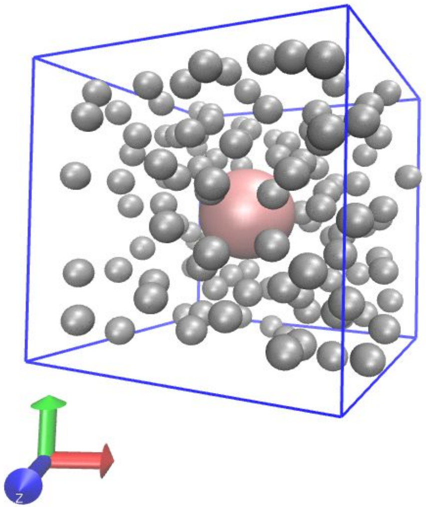
Why respect symmetries?

Many real-world problems have **known symmetries.**

- **Physics**
  - time reversal,
  - identical particles,
  - wave functions, or
  - gauge invariance, …
- **Point-cloud modelling** (e.g. 3D objects)
- **Image detection** (e.g. rotations)

⇨ **Can have dramatic effects on training!**

$$p(x) = \frac{e^{-\beta U(x)}}{Z}$$



Examples of common symmetries

- **Permutations**
  - symmetric
  - antisymmetric

$$p(\ldots, x_i, \ldots, x_j, \ldots) = p(\ldots, x_j, \ldots, x_i, \ldots)$$

$$\psi(\ldots, x_i, \ldots, x_j, \ldots) = -\psi(\ldots, x_j, \ldots, x_i, \ldots)$$

- **Translations and/or rotations**
  - SE(3)
  - Octahedral symmetries

$$x \to Rx + \mu$$

- **Gauge invariance**
  - U(n)
  - SU(n)

$$(\Omega \cdot U)_\mu(x) = \Omega(x)U_\mu(x)\Omega^\dagger(x+\hat{\mu})$$

DeepMind

General mechanisms to incorporate symmetry and equivariance in ML

Building Invariance: Group convolutions

Invariant map $$\bar{\phi}(x) = \int d\mu(g)\phi(T_g \circ x)$$

Equivariant map $$\hat{\phi}(x) = \int d\mu(g)(T_{g^{-1}} \circ \phi)(T_g \circ x)$$

Example:
group convolution nnets

Building Invariance: Group convolutions

Invariant map $$\bar{\phi}(x) = \int d\mu(g)\phi(T_g \circ x)$$

Equivariant map $$\hat{\phi}(x) = \int d\mu(g)(T_{g^{-1}} \circ \phi)(T_g \circ x)$$

Example:
group convolution nnets

Not scalable with dimension of G!

Building Invariance: Direct use of known group invariants

Example: Pairwise interactions for translational invariance (e.g. graphnets, transformers)

$$\phi(x) = \sum_{ij} f(||x_i - x_j||)$$

Building Invariance: Direct use of group invariants

Example: Trace-networks for matrix conjugation invariance

$$X \to UXU^{\star}$$

$$U \in SU(n)$$

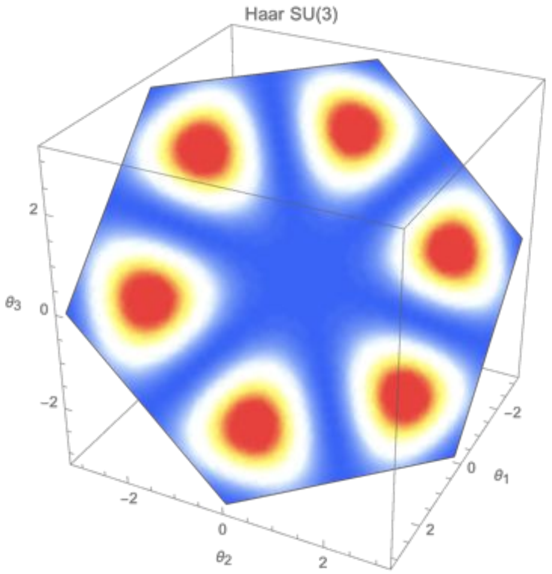
$$\phi(X) = f(\mathrm{Tr}(X), \mathrm{Tr}(XX), \ldots)$$

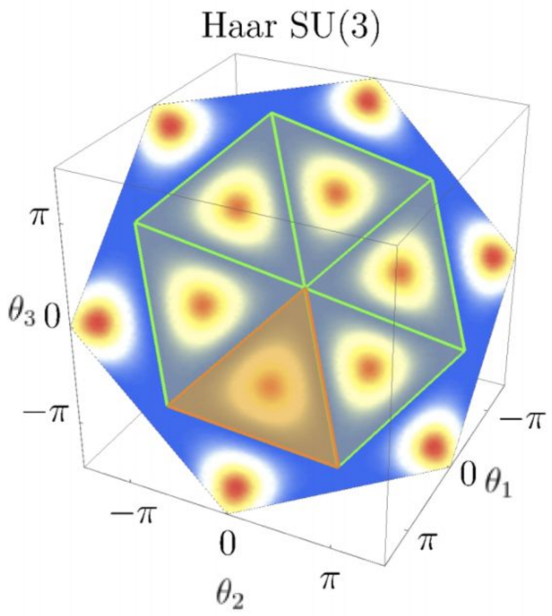
Example from Physics: Wilson loops

$$W_C = \mathrm{Tr}(Pe^{i\oint U_\mu dx^\mu})$$

Building Invariance: Canonicalization maps

Canonicalize -> Flow on cell -> Uncanonicalize





Building Equivariance: Equivariance from Invariance

Lemma 2 (Equivariance from invariance) *Let $f : \mathbb{R}^D \to \mathbb{R}$ be invariant with respect to G, and assume that $\mathbf{R}_g$ is orthogonal for all $g \in G$. Then $\nabla_{\mathbf{u}} f(\mathbf{u})$ is equivariant with respect to G.*

Example: Permutation equivariant gradient maps

$$f(x) = h(\sum_i \phi(x_i))$$

$$\nabla_{x_i} f(x) = h'(\sum_j \phi(x_j))\nabla_{x_i}\phi(x_i)$$

Papamakarios, G., Nalisnick, E., Rezende, D.J., Mohamed, S. and Lakshminarayanan, B., 2019. Normalizing flows for probabilistic modeling and inference. arXiv preprint arXiv:1912.02762.

Convnets on manifolds and fiber bundles: Gauge symmetries

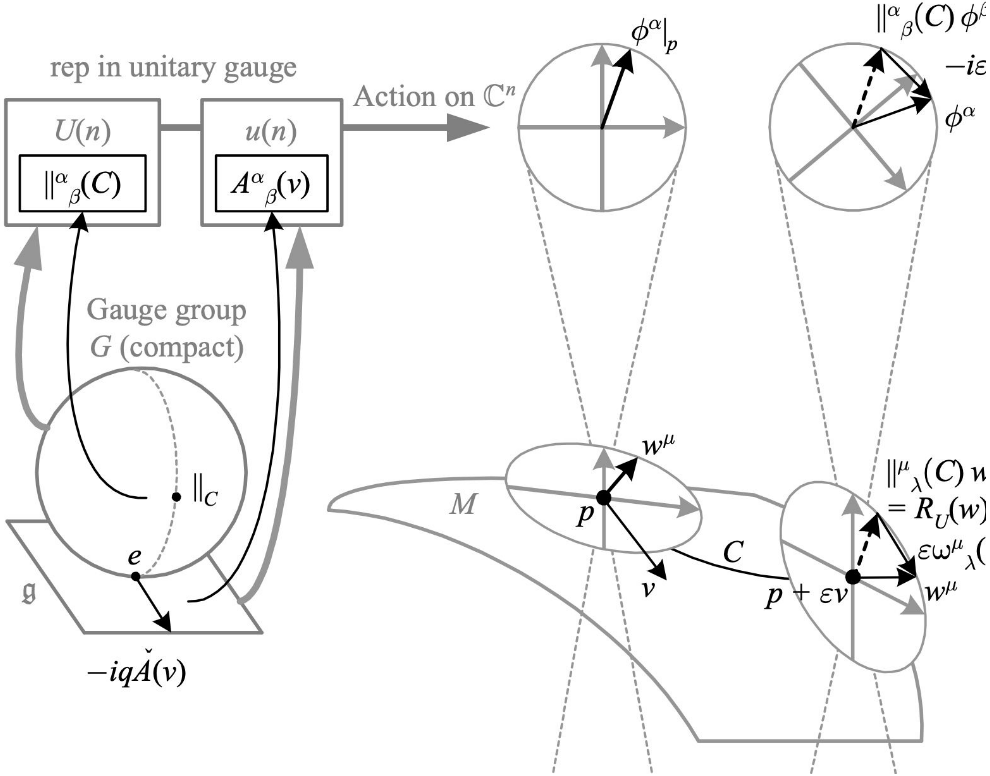


Image credit: Gauge Theories and Fiber Bundles: Definitions, Pictures, and Results, Adam Marsh

Convnets on manifolds and fiber bundles: Gauge symmetries, a concrete example.

Matrix-conjugation equivariant convnets

$$T_\Omega W_{x,i} = \Omega_x W_{x,i} \Omega_x^\dagger$$

Parallel transport of W

$$W_{x,i} \to \sum_{j,\mu,k} \omega_{i,j,\mu,k} \boxed{U_{x,k\cdot\mu} W_{x+k\cdot\mu,j} U_{x,k\cdot\mu}^\dagger}$$

Examples of non-linearities that preserve equivariance

$$W_{x,i} \to \sum_{j,k} \alpha_{i,j,k} W_{x,j} W'_{x,k} \qquad W_{x,i} \to g_{x,i}(\mathcal{U},\mathcal{W}) W_{x,i}$$

Favoni, M., Ipp, A., Müller, D.I. and Schuh, D., 2022. Lattice gauge equivariant convolutional neural networks. Physical Review Letters, 128(3), p.032003.
Gerken, J.E., Aronsson, J., Carlsson, O., Linander, H., Ohlsson, F., Petersson, C. and Persson, D., 2021. Geometric deep learning and equivariant neural networks. arXiv preprint arXiv:2105.13926.

Building Invariant Densities: General principle

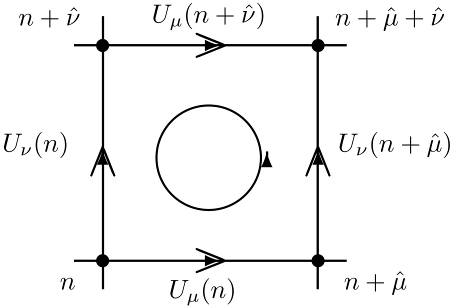
Lemma 1 (Equivariant flows) *Let* $p_{\mathrm{x}}(\mathbf{x})$ *be the density function of a flow-based model with transformation* $T : \mathbb{R}^D \to \mathbb{R}^D$ *and base density* $p_{\mathrm{u}}(\mathbf{u})$*. If* T *is equivariant with respect to* G *and* $p_{\mathrm{u}}(\mathbf{u})$ *is invariant with respect to* G*, then* $p_{\mathrm{x}}(\mathbf{x})$ *is invariant with respect to* G*.*

Rezende et al., *Equivariant Hamiltonian Flows*, arXiv (2019)
Köhler, Klein and Noe, *Equivariant Flows: Exact Likelihood Generative Learning for Symmetric Densities* ICML (2020)
Papamakarios *et al.*, *Normalizing flows for probabilistic modeling and inference*, JMLR *(2021)*

DeepMind

4 Special Manifolds: U(N), SU(N)

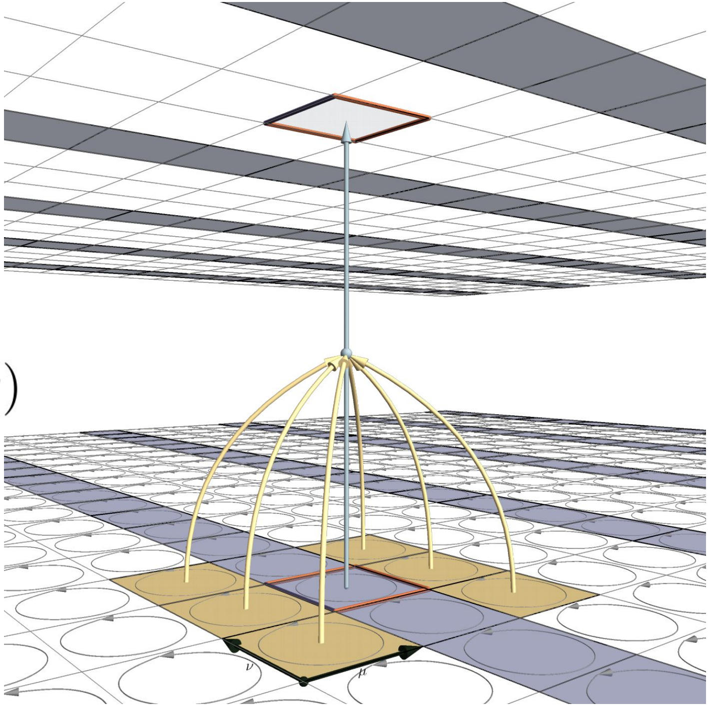
Lattice Quantum Chromodynamics



$$P_{\mu\nu}(x) := U(x,\mu)U(x+\hat{\mu},\nu)U^\dagger(x+\hat{\nu},\mu)U^\dagger(x,\nu)$$

$$S := -\beta \sum_{\text{sites } x} \sum_{\mu=1}^{D} \sum_{\nu=\mu+1}^{D} \mathrm{Re}\left[\frac{1}{N}\mathrm{Tr}\left(P_{\mu\nu}(x)\right)\right]$$

$$p(U) \propto e^{-\beta S[U]}$$



Motivation: Gauge Equivariance

$$Y_\mu(x) = f(U_\mu(x); \theta)$$

$$U_\mu(x) \rightarrow \Omega(x) U_\mu(x) \Omega(x + \hat{\mu})^\dagger$$

$$Y_\mu(x) \rightarrow \Omega(x) Y_\mu(x) \Omega(x + \hat{\mu})^\dagger$$

Motivation: Gauge Equivariance

Let h be an invertible map such that

$$h : \mathrm{SU}(N) \to \mathrm{SU}(N)$$

$$h(\Omega_\mu(x) X_\mu(x) \Omega_\mu(x)^\dagger) = \Omega_\mu(x) h(X_\mu(x)) \Omega_\mu(x)^\dagger$$

Then the map f,

$$f(X_\mu(x)) = h(P_{\mu\nu}(x)) S_{\mu\nu}(x)^\dagger$$

where $$S_{\mu\nu}(x) = X_\mu(x)^\dagger P_{\mu\nu}(x)$$

is equivariant to Gauge transformations

Building Equivariance: Matrix Conjugation Equivariance

$$X \to UXU^{\star}$$

$$U \in SU(n)$$

Proposition 1. *Let $f : G \to G$ be a matrix conjugation equivariant diffeomorphism. Then f restricted to T is a diffeomorphism of T that is equivariant under the action of the Weyl group.*

T is the maximal torus of G

Boyda, D., Kanwar, G., Racanière, S., Rezende, D.J., Albergo, M.S., Cranmer, K., Hackett, D.C. and Shanahan, P.E., 2020. Sampling using $ SU (N) $ gauge equivariant flows. arXiv preprint arXiv:2008.05456.

Building Equivariance: Matrix Conjugation Equivariance

In the case of $G = \mathrm{SU}(N)$ or $G = \mathrm{U}(N)$, a maximal torus is given by the subgroup of diagonal matrices, and the Weyl group is isomorphic to the group of permutations

Matrix Conjugation Equivariance ⇔ Permutation Equivariance of eigenvalues

Boyda, D., Kanwar, G., Racanière, S., Rezende, D.J., Albergo, M.S., Cranmer, K., Hackett, D.C. and Shanahan, P.E., 2020. Sampling using SU(N) gauge equivariant flows. arXiv preprint arXiv:2008.05456.

Building Equivariance: Matrix Conjugation Equivariance

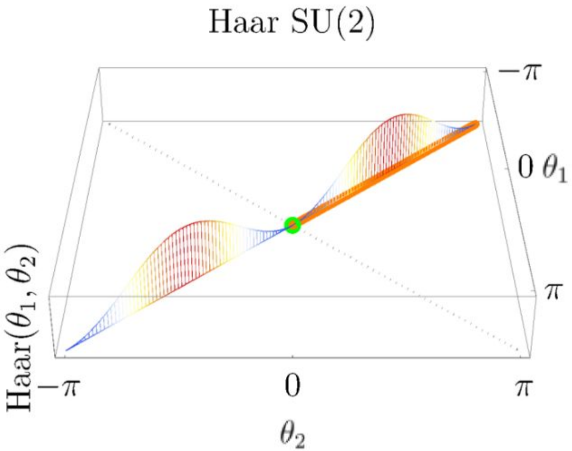
Matrix-conjugation diffeomorphisms on SU(N) are generated by permutation-equivariant diffeomorphisms on eigenvalues

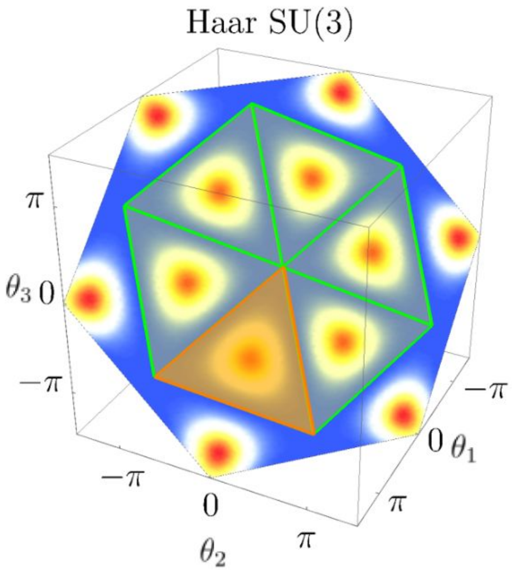
$$(X, D = \mathrm{diag}(w)) = \mathrm{eigen}(U)$$

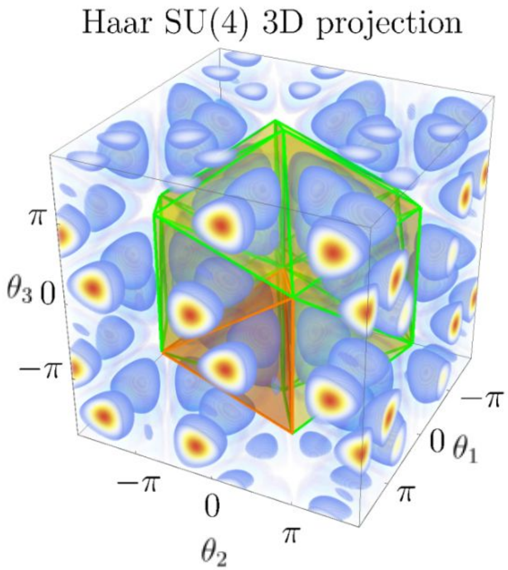
$$Y = X \mathrm{diag}(g(w)) X^\dagger$$

If g is a permutation-equivariant flow that preserves unitarity (\prod g(w) = 1)

Haar measure on the maximal torus of SU(N)







DeepMind

Alternative constructions for SU(N) Gauge-equivariant maps

Alternative Gauge-equivariant map: Exp-product map

Projects to the Lie algebra

A collection of staples

$$Y_\mu(x) = e^{\sum_k w_k \mathcal{P}[U_\mu(x) S_\mu^k(x)]} U_\mu(x)$$

Invertibility requires bounded coefficients

Alternative Gauge-equivariant map: SU(N) ODE flow, trivializing flows

Gauge-invariant scalar

Lie algebra generators

$$Y_\mu(x) = e^{\epsilon T^a \partial_{\mu,x,a}\phi(U)} U_\mu(x)$$

Invertibility requires bounded step-size

Right-invariant derivative

Trivializing maps, the Wilson flow and the HMC algorithm, Martin Luscher

5 Applications

DeepMind

5.1 Application: Free energy of solids

Collaborators

Peter Wirnsberger

George Papamakarios

Borja Ibarz

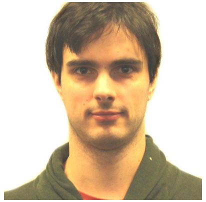
Andy Ballard

Stuart Abercrombie

Sébastien Racanière

Alexander Pritzel

Danilo Rezende



Charles Blundell

Wirnsberger, Ballard *et al.*, ***Targeted free energy estimation via learned mappings***, JCP (2020).

Wirnsberger, Papamakarios, Ibarz *et al.*, ***Normalizing flows for atomic solids***, MLST (2022).

Free energy

$$F = -\beta^{-1} \ln Z$$

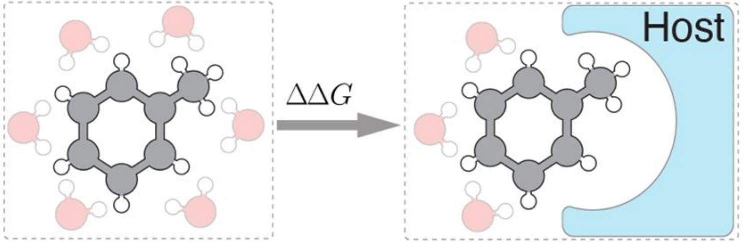


Image credit: Mey *et al.*, *Living J Comput Mol Sci.* (2021)

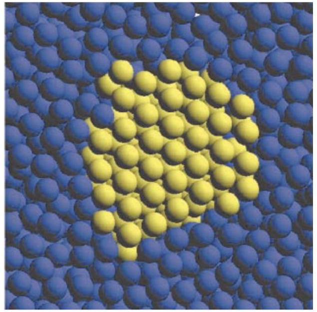
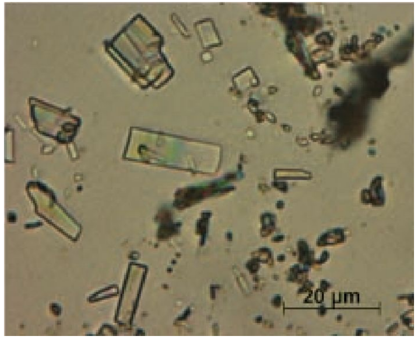


Image credit: Auer and Frenkel, *Nature* (2001)

Related to:

- Phase transitions
- Molecular stability
- Drug binding and solubility
- ...



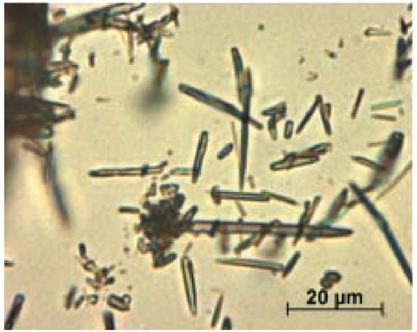
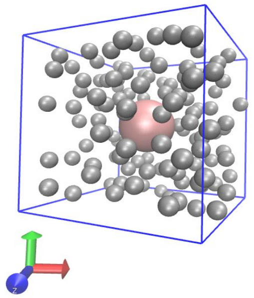


Image credit: Morissette *et al.*, *PNAS 100*

Problem definition

Estimate **free energy changes** between two states.



$$p_\alpha(x) = \frac{1}{Z_\alpha} e^{-\beta U_\alpha(x)}$$

state *A* or *B*

$$F_\alpha = -\beta^{-1} \log Z_\alpha$$

$$\Delta F = F_B - F_A$$

Estimators

Many specialised estimation techniques have been developed:

- Thermodynamic integration
- Free energy perturbation (FEP)
- Bennetts acceptance ratio (BAR)
- Jarzynski method / Annealed Importance Sampling
- Weighted histogram analysis method (WHAM)
- Multistate BAR (MBAR)
- Metadynamics…

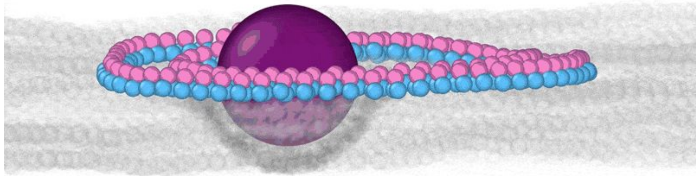
Can we use ML to improve them?

Frenkel and Smit (2002), *Understanding Molecular Simulation,* 2nd edn (San Diego), 2002.

Traditional approaches

- **Molecular Dynamics (MD)**
- **Markov Chain Monte Carlo (MCMC)**
  - Hamiltonian Monte Carlo
  - Langevin dynamics

Animations credit: Šarić Lab, andelasaric.com



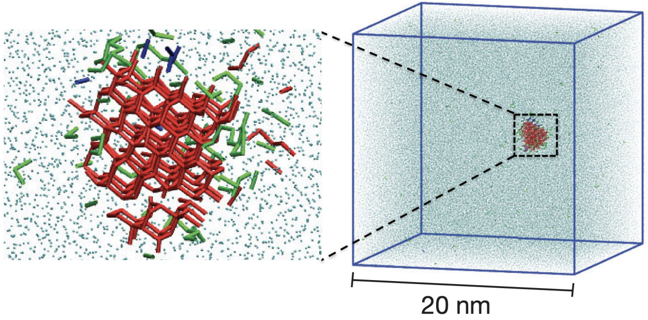
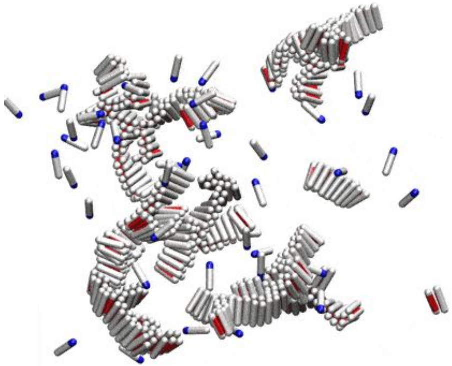


Image credit: Lupi *et al.*, *Nature 501*



Traditional approaches

Sampling & expectations

1. Burn-in period

2. Collecting samples

Samples directly from target distribution $p(x)$

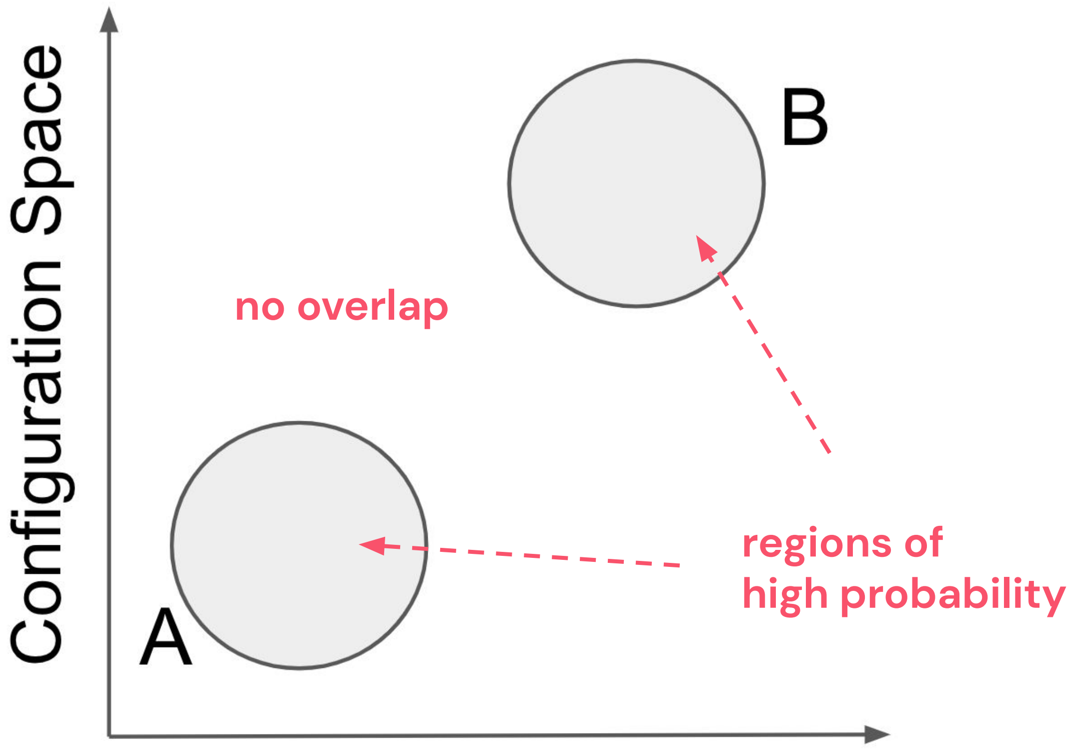
$$\{x_i\}_{1 \leq i \leq N}$$

3. MC estimate

No unbiasing required

$$\langle O \rangle \approx \hat{O} = N^{-1} \sum_{i=1}^{N} O(x_i)$$

The “overlap problem”



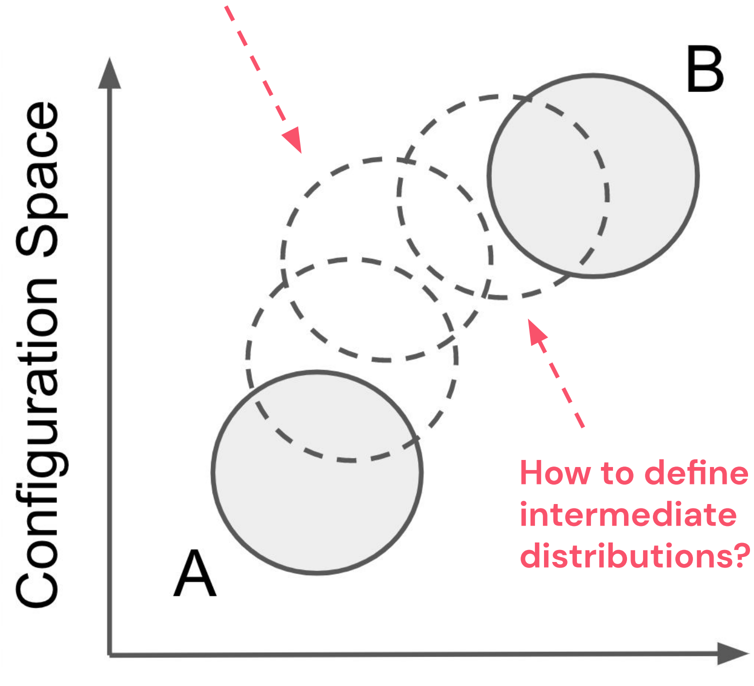
Multistate methods

Introduce **intermediate distributions**:

- Thermodynamic integration
- Multistep FEP
- WHAM
- MBAR, ...

Works well but is **expensive**.

Many simulations



How to define good intermediate distributions?

Image credit: Wirnsberger, Ballard, et al., *JCP* (2020).

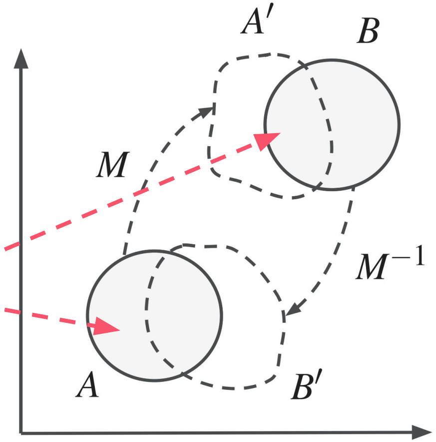
Learned estimators

ML

Free energy estimation as a learning problem: M_θ

Setting 1

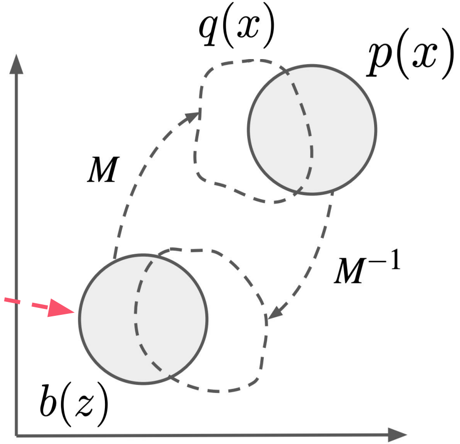
Train on datasets



$$p^*_{A'}(M(x)) = p^*_A(x)|\det J_M(x)|^{-1}$$

Setting 2

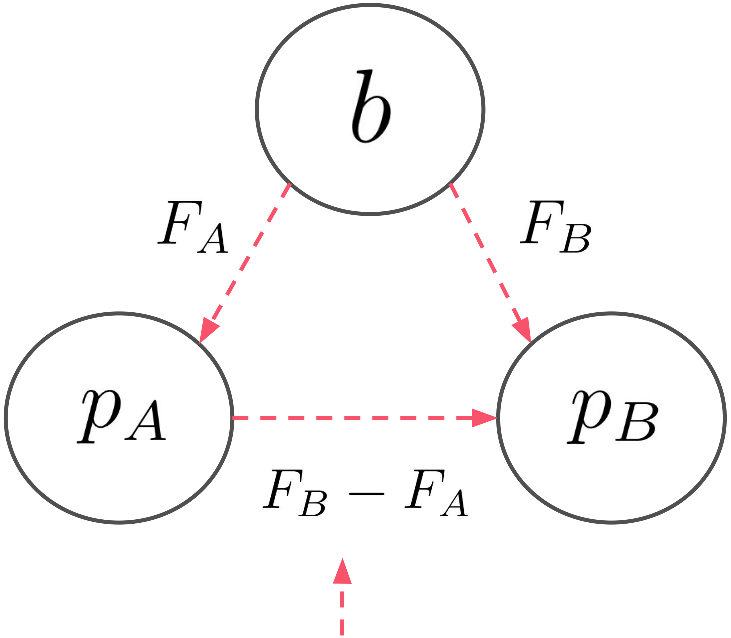
Sample from a tractable base



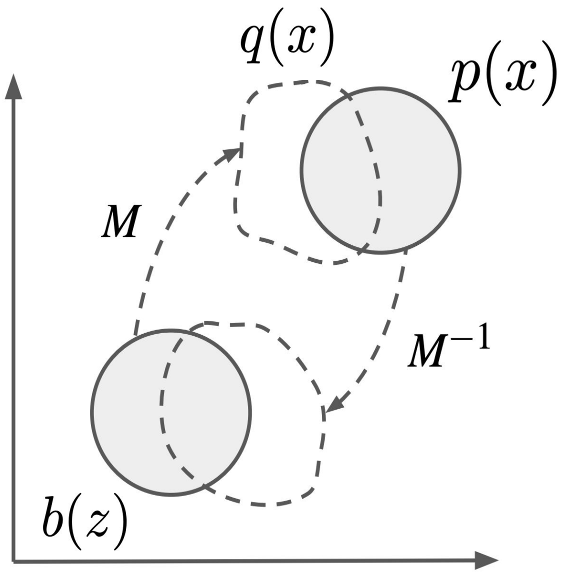
$$q(x) = b(z)|\det J_M(z)|^{-1}$$

Image credit: Wirnsberger, Ballard *et al.*, [J. Chem. Phys.](#) (2020).

Solids: Problem setup



Requires two experiments.



$$q(x) = b(z)|\det J_M(z)|^{-1}$$

Wirnsberger, Papamakarios, Ibarz *et al.*, *Normalizing flows for atomic solids*, MLST (2022).

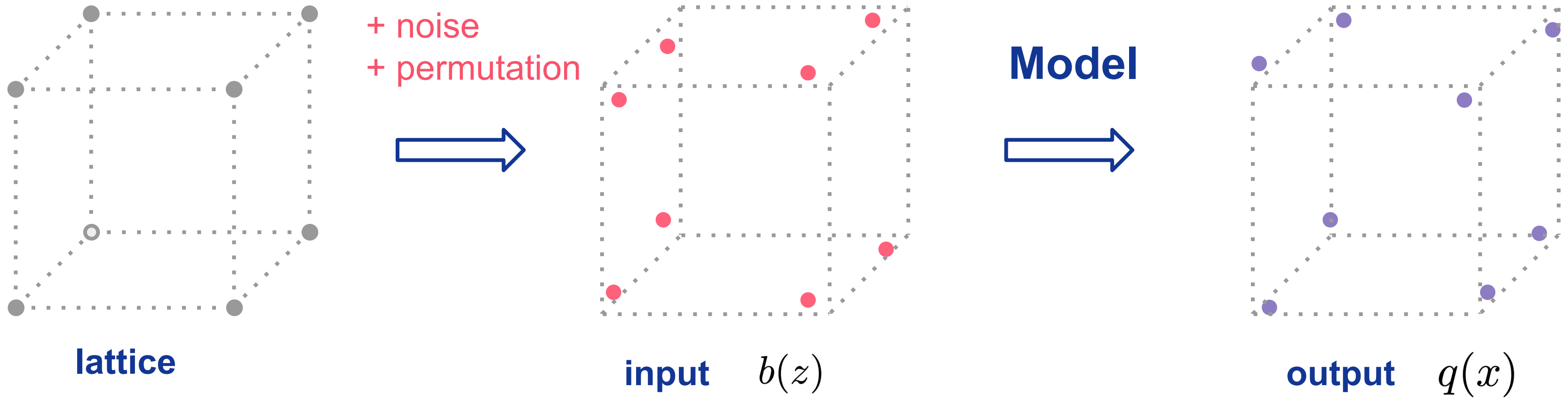
Atomic solids: permutation equivariance

Invariant Density
(permutation)

Equivariant Transformation

Complex Invariant Density

$$q(x) = b(z)\,|\det J_f(z)|^{-1}$$



Wirnsberger, Papamakarios, Ibarz *et al.*, *Normalizing flows for atomic solids*, MLST (2022).

Permutation-equivariant coupling layer

Split across particle coordinates:

$$\mathbf{x}_n = (\mathbf{x}_n^A, \mathbf{x}_n^B)$$

Coupling layer is
permutation-equivariant if C is.

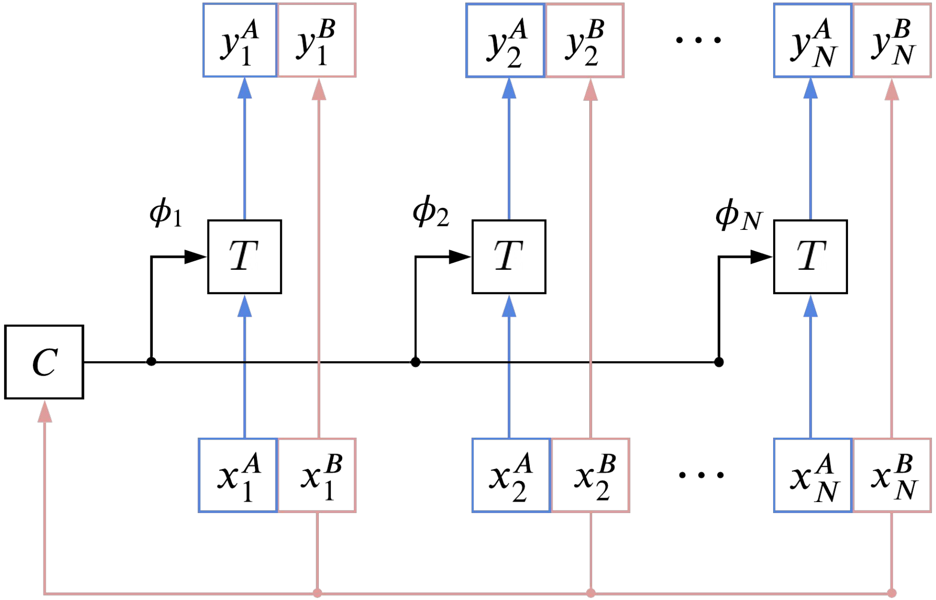


Image credit: Wirnsberger et al., *Targeted free energy estimation via normalizing flows*, JCP (2020).

Permutation-equivariant conditioner

Transformer architecture

(without positional embeddings)

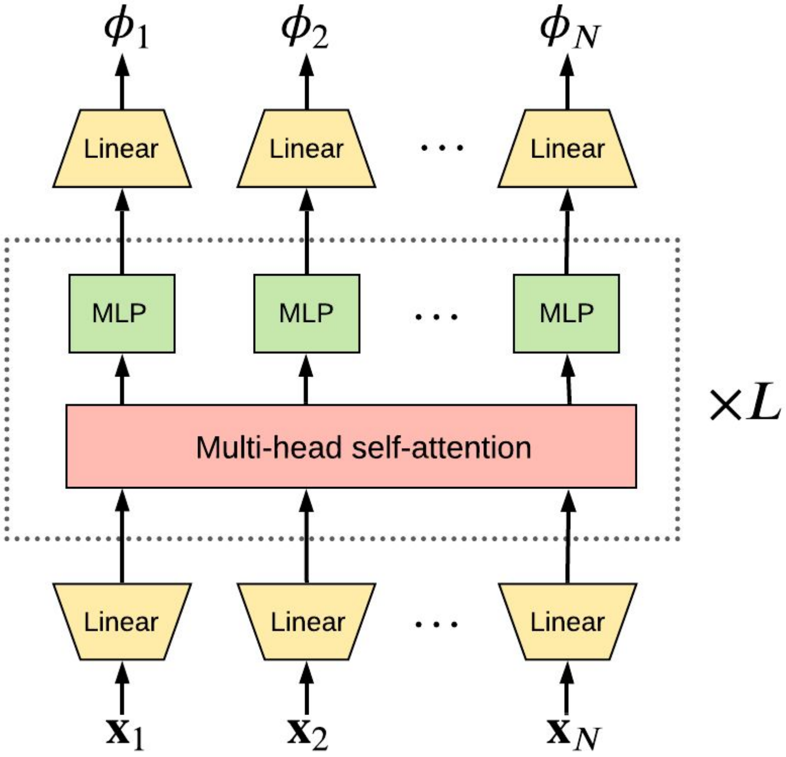
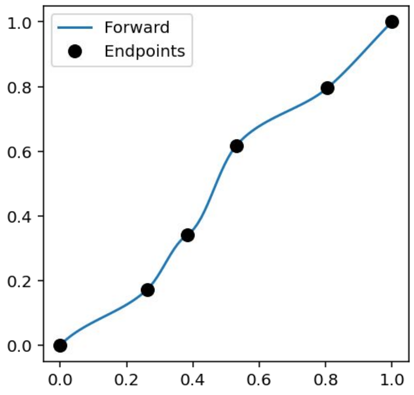
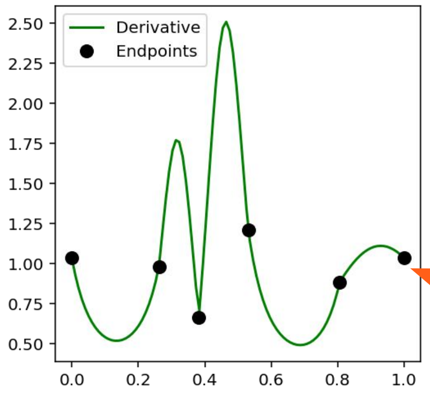


Image credit: Vaswani et al., *Attention is all you need*, NeurIPS (2017).

Coupling flow on tori: Periodic boundary conditions

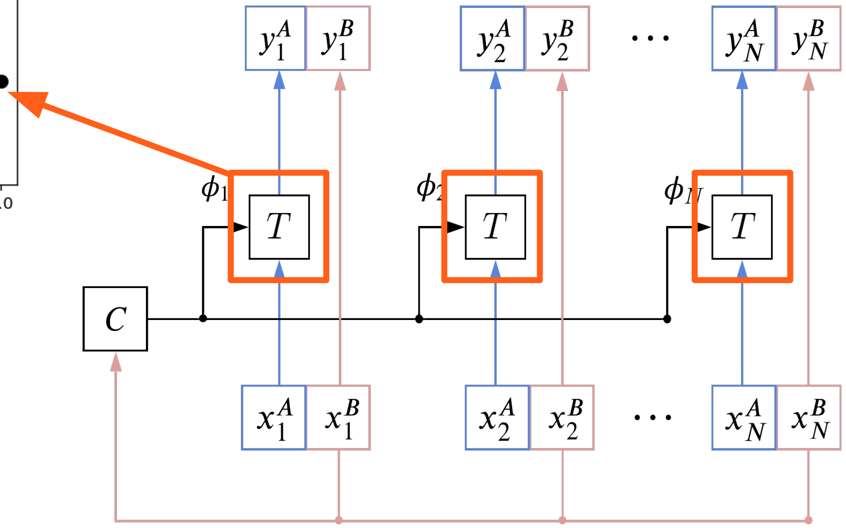




$$T(0) = 0$$

$$T(L) = L$$

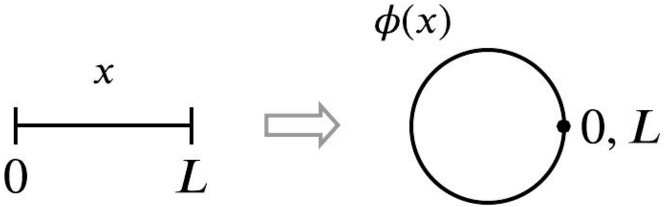
$$T'(0) = T'(L)$$



Slide credit: George Papamakarios

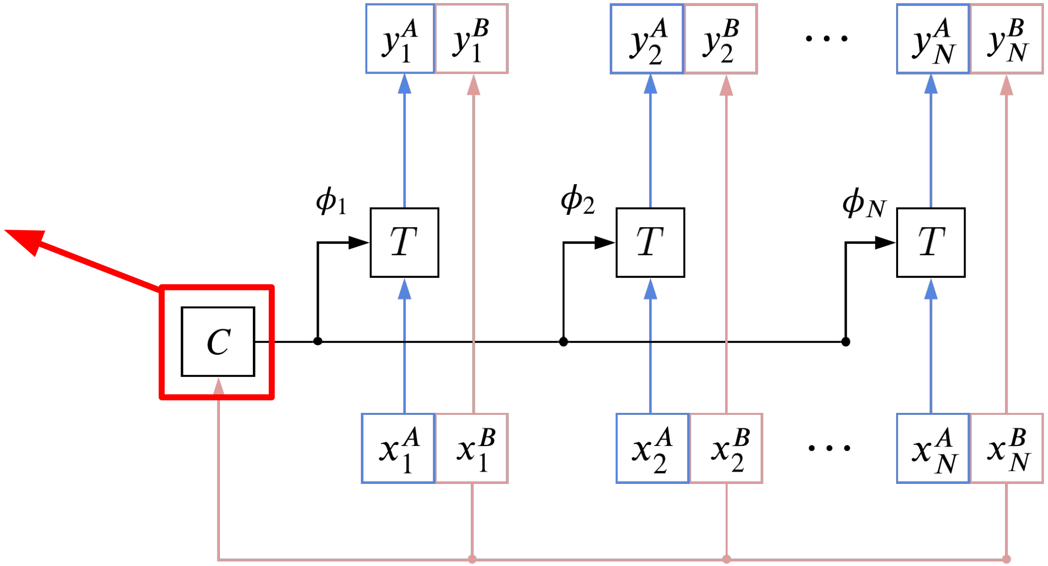
Rezende, Papamakarios, Racanière et al., *Normalizing flows on tori and spheres*, ICML (2020).

Coupling flow on tori: Circular embedding



$$\phi(\mathbf{x}) = \big(\cos(\omega\mathbf{x}), \sin(\omega\mathbf{x})\big)$$

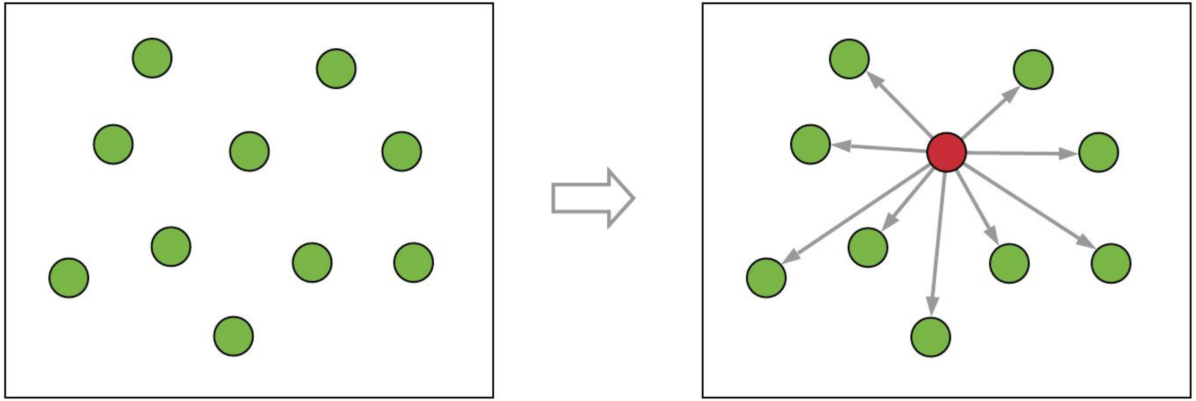
$$\omega = \frac{2\pi}{L}$$ + integer multiples



Slide credit: George Papamakarios

Image credit: Wirnsberger et al., *Targeted free energy estimation via normalizing flows*, JCP (2020).

Global translation symmetry



- Choose a particle as reference
- Place it randomly
- Flow generates *N-1* other particles relative to reference

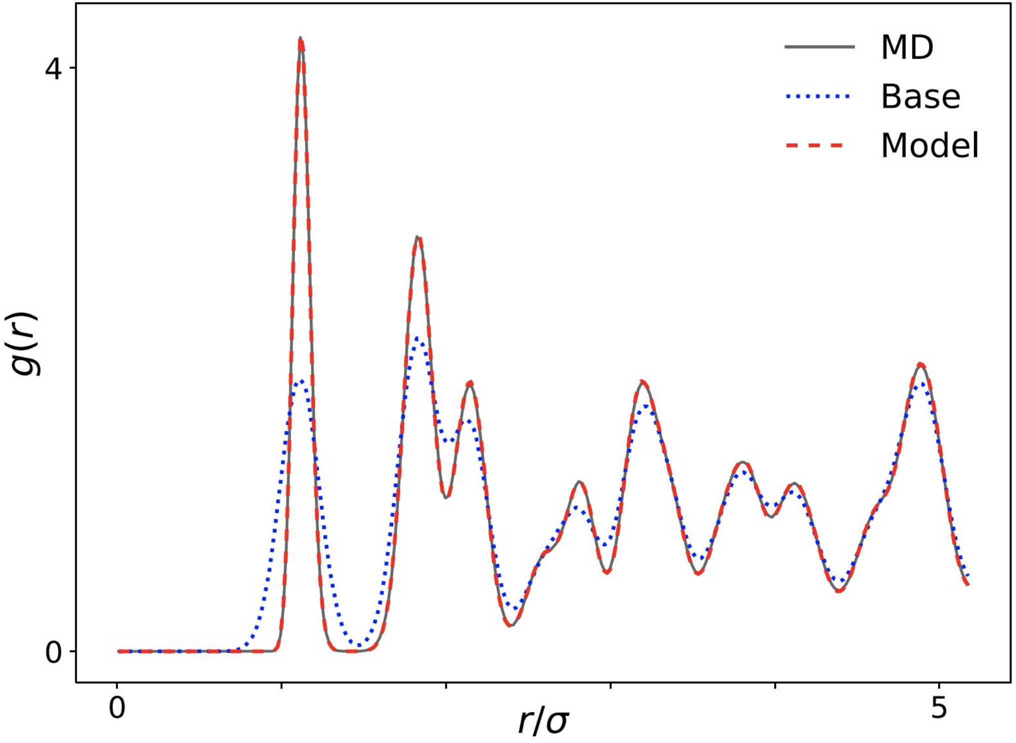
Slide credit: George Papamakarios

DeepMind

Results

Results: Radial distribution function

- 512 particles
- Cubic ice
- No unbiasing.

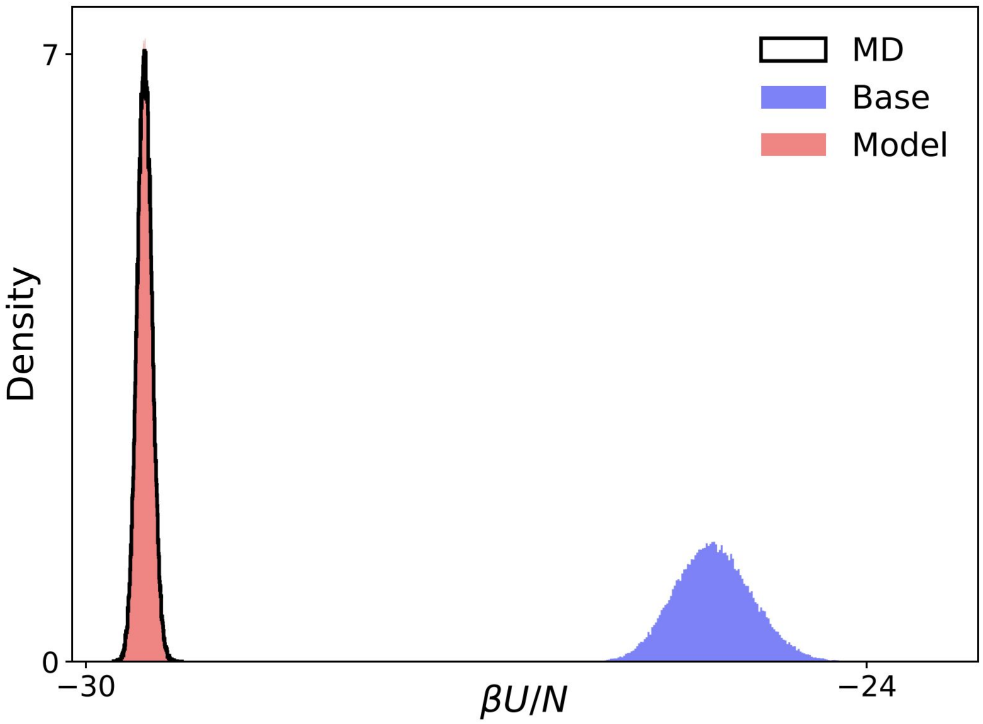


Wirnsberger, Papamakarios, Ibarz *et al.*, *Normalizing flows for atomic solids*, MLST (2022).

Solids: Energy histogram

Energies computed from the **base and the model differ significantly.**

No appreciable difference between model and MD.



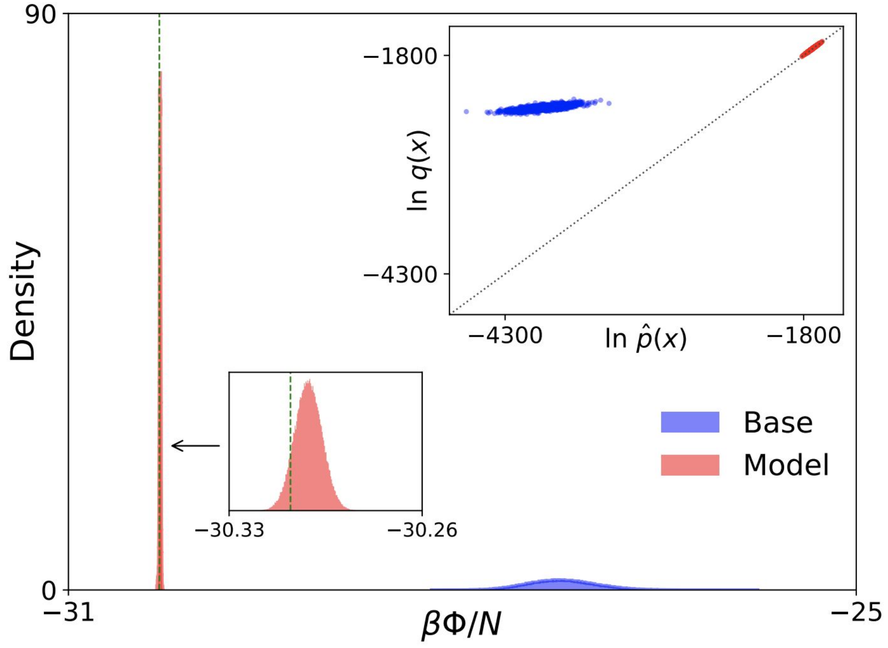
Wirnsberger, Papamakarios, Ibarz *et al.*, *Normalizing flows for atomic solids*, MLST (2022).

Solids: Histogram of work values

The **distribution of work values** exhibits a sharp peak.

$$\beta\Phi(x) = \beta U(x) + \ln q(x)$$

$$\ln Z = \ln \langle \exp(-\beta\Phi(x)) \rangle_q$$



Wirnsberger, Papamakarios, Ibarz *et al.*, *Normalizing flows for atomic solids*, MLST (2022).

Solids: Free energies

$$F = -\beta^{-1} (\ln Z - \ln N!)$$

no MD data → LFEP

Model + MD data (from target) → LBAR

100-200 MD runs (multistate) → MBAR

System	N	LFEP	LBAR	MBAR
LJ	256	3.10800(28)	3.10797(1)	3.10798(9)
LJ	500	3.12300(41)	3.12264(2)	3.12262(10)
Ice Ic	64	−25.16311(3)	−25.16312(1)	−25.16306(20)
Ice Ic	216	−25.08234(7)	−25.08238(1)	−25.08234(5)
Ice Ic	512	−25.06163(35)	−25.06161(1)	−25.06156(3)
Ice Ih	64	−25.18671(3)	−25.18672(2)	−25.18687(26)
Ice Ih	216	−25.08980(3)	−25.08979(1)	−25.08975(14)
Ice Ih	512	−25.06478(9)	−25.06479(1)	−25.06480(4)

Wirnsberger, Papamakarios, Ibarz *et al.*, *Normalizing flows for atomic solids*, MLST (2022).

5.2 Application: Lattice quantum chromodynamics

The team

Center for Theoretical Physics, MIT

Gurtej Kanwar

Phiala Shanahan

Denis Boyda

Dan Hackett

NYU

Center for Cosmology and Particle Physics, NYU

Michael Albergo

Kyle Cranmer

Julian Urban (work on fermions)

DeepMind

Sébastien Racanière

Danilo Rezende

Ali Razavi

Alex Matthews

Alex Botev

What is Lattice QCD?

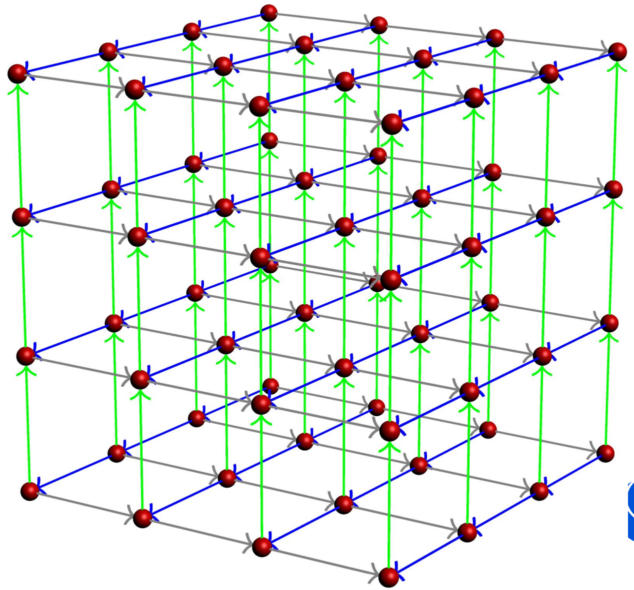
- Lattice quantum chromodynamics (LQCD) is a subfield of computational physics which aims to simulate elementary particle fields involved in the "strong interaction" called quarks and gluons.

- These simulations involve discretising space-time using a lattice and simulating quantum fluctuations of the particle fields; typically using HMC.

The problem space: The Standard Model of Particle Physics in a box

Three axes of model complexity:

- dimension of space-time: 2D, 3D and 4D;
- lattice size (discretisation of space-time): Eg from L=8 to L=32;
- features of the theory:
  - Gauge fields: photons, gluons
    - no force (ϕ^4)
    - electromagnetism with U(1)
    - weak nuclear force with ~SU(2)
    - strong nuclear force with SU(3)
  - Fermion fields: electrons, quarks



Scale Enables Impact: Larger lattices allow for ab-initio study of a larger number of problems

Lattice size = L
Volume = L^4
Beta >= 6

L >= 16

- Baryon spectroscopy (i.e. derive bound state energies / masses)

L >= 32

- Study nuclear fusion
- Big Bang nucleosynthesis

L > 96
(exascale compute)

- muon magnetic moment
- Study dark matter
- Study the interior of neutron stars

Flows for Scalar Fields

Modelling scalar fields with flows

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

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

[1]*Perimeter Institute for Theoretical Physics, Waterloo, Ontario N2L 2Y5, Canada*
[2]*Cavendish Laboratories, University of Cambridge, Cambridge CB3 0HE, United Kingdom*
[3]*University of Waterloo, Waterloo, Ontario N2L 3G1, Canada*
[4]*Center for Theoretical Physics, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA*

(Received 17 May 2019; published 22 August 2019; corrected 21 November 2019)

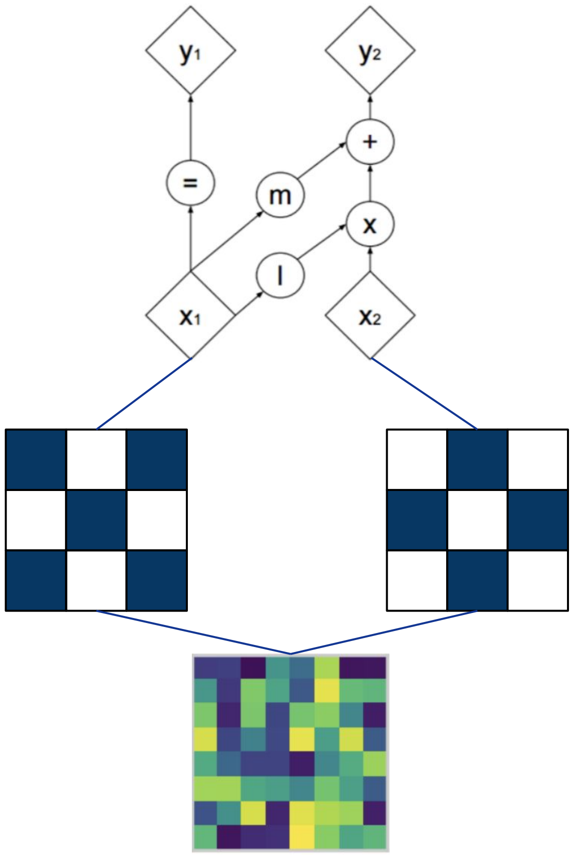
A Markov chain update scheme using a machine-learned flow-based generative model is proposed for Monte Carlo sampling in lattice field theories. The generative model may be optimized (trained) to produce samples from a distribution approximating the desired Boltzmann distribution determined by the lattice action of the theory being studied. Training the model systematically improves autocorrelation times in the Markov chain, even in regions of parameter space where standard Markov chain Monte Carlo algorithms exhibit critical slowing down in producing decorrelated updates. Moreover, the model may be trained without existing samples from the desired distribution. The algorithm is compared with HMC and local Metropolis sampling for ϕ^4 theory in two dimensions.

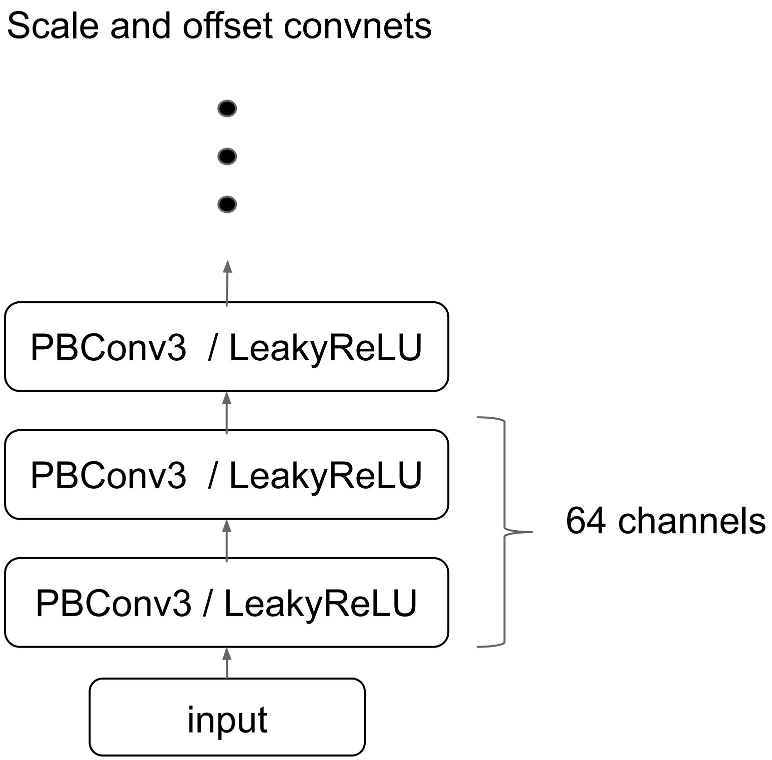
Modelling scalar fields with flows

$$S^E_{\text{latt}}(\phi) = \sum_{\vec{n}} \phi(\vec{n}) \left[\sum_{\mu \in \{1,2\}} 2\phi(\vec{n}) - \phi(\vec{n}+\hat{\mu}) - \phi(\vec{n}-\hat{\mu}) \right] + m^2 \phi(\vec{n})^2 + \lambda \phi(\vec{n})^4$$

$$p(\phi) = \frac{1}{Z} e^{-S(\phi)}, \qquad Z \equiv \int \prod_{\vec{n}} d\phi(\vec{n})\; e^{-S(\phi)}$$

Stack of masked flows





The learned model replicates HMC two-point functions

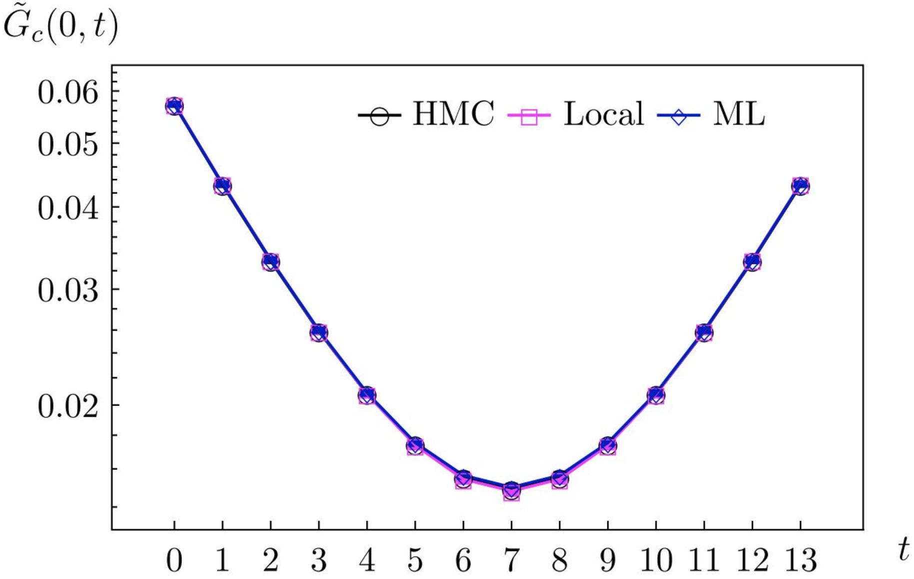


FIG. 3. Zero-momentum Green's functions evaluated for parameter set E5. Results computed using 10^6 configurations from the HMC, local Metropolis, and ML ensembles are consistent within statistical errors. Error bars indicate 68% confidence intervals estimated using bootstrap resampling with bins of size 100.

DeepMind

The Yukawa model: scalar fields + fermions

Modelling scalar and fermion fields with flows

Flow-based sampling for fermionic lattice field theories

Michael S. Albergo,[1,*] Gurtej Kanwar,[2,3,†] Sébastien Racanière,[4,‡] Danilo J. Rezende,[4,§] Julian M. Urban,[5,¶] Denis Boyda,[6,2,3] Kyle Cranmer,[1] Daniel C. Hackett,[2,3] and Phiala E. Shanahan[2,3]

[1] *Center for Cosmology and Particle Physics, New York University, New York, NY 10003, US*
[2] *Center for Theoretical Physics, Massachusetts Institute of Technology, Cambridge, MA 02139, U.S.A.*
[3] *The NSF AI Institute for Artificial Intelligence and Fundamental Interactions*
[4] *DeepMind, London, UK*
[5] *Institut für Theoretische Physik, Universität Heidelberg, Philosophenweg 16, 69120 Heidelberg, Germany*
[6] *Argonne Leadership Computing Facility, Argonne National Laboratory, Lemont IL-60439, USA*

Algorithms based on normalizing flows are emerging as promising machine learning approaches to sampling complicated probability distributions in a way that can be made asymptotically exact. In the context of lattice field theory, proof-of-principle studies have demonstrated the effectiveness of this approach for scalar theories, gauge theories, and statistical systems. This work develops approaches that enable flow-based sampling of theories with dynamical fermions, which is necessary for the technique to be applied to lattice field theory studies of the Standard Model of particle physics and many condensed matter systems. As a practical demonstration, these methods are applied to the sampling of field configurations for a two-dimensional theory of massless staggered fermions coupled to a scalar field via a Yukawa interaction.

Yukawa model

$$\mathcal{L} = \mathcal{L}_{\text{scalar}} + \sum_f \psi_f^\dagger D_f \psi_f$$

$$D_f = i\partial\!\!\!/ - m_f - g\phi$$

Discussion: Yukawa model

$$\int \mathcal{L}_{\text{eff}}(\phi) := -\log \int \mathcal{D}\psi^\dagger \mathcal{D}\psi e^{-\int dx \mathcal{L}(\phi,\psi)}$$

$$\int dx \mathcal{L}_{\text{eff}}(\phi) = \int dx \mathcal{L}_{\text{scalar}}(\phi) + \log \prod_f \boxed{\det D_f} + \text{cst}$$

When Nf = 2 and m1 = m2

$$\begin{aligned} \det D_{f_1} \det D_{f_2} &= \det D_{f_1} \det(\gamma_5 D_{f_2} \gamma_5) \\ &= \det D_{f_1} \det D_{f_2}^\dagger \\ &= \det D D^\dagger \end{aligned}$$

Christof Gattringer and Christian B. Lang.Quan-tum chromodynamics on the lattice.Lect. Notes Phys.,788:1–343, 2010.

Pseudo-fermions

$$(\det DD^\dagger)^{1/2} \propto \int \mathcal{D}\chi^\dagger \mathcal{D}\chi e^{-\chi^T (DD^\dagger)^{-1}\chi}$$

$$\mathcal{L}_{\text{eff}}(\phi, \chi) = \mathcal{L}_{\text{scalar}}(\phi) - \sum_f \chi_f^\dagger (DD^\dagger)^{-1} \chi_f$$

Christof Gattringer and Christian B. Lang.Quan-tum chromodynamics on the lattice.Lect. Notes Phys.,788:1–343, 2010.

Considered many combinations of target density

Name	Probability density	Use case
Joint[A]	$p(\phi, \varphi) = \frac{1}{Z} \exp(-S_B(\phi) - \varphi^\dagger [\mathcal{M}(\phi)]^{-1} \varphi)$	Section III D
ϕ-marginal	$p(\phi) = \frac{Z_\mathcal{N}}{Z} \exp(-S_B(\phi)) \det \mathcal{M}(\phi)$	Sections III A and III C
φ-conditional[A,B]	$p(\varphi\|\phi) = \frac{1}{Z_\mathcal{N} \det \mathcal{M}(\phi)} \exp(-\varphi^\dagger [\mathcal{M}(\phi)]^{-1} \varphi)$	Sections III A, III B and III C
φ-marginal[C]	$p(\varphi) = \frac{1}{Z} \int d\phi \, \exp(-S_B(\phi) - \varphi^\dagger [\mathcal{M}(\phi)]^{-1} \varphi)$	–
ϕ-conditional[A]	$p(\phi\|\varphi) = \frac{\exp(-S_B(\phi) - \varphi^\dagger [\mathcal{M}(\phi)]^{-1} \varphi)}{\int d\phi \, \exp(-S_B(\phi) - \varphi^\dagger [\mathcal{M}(\phi)]^{-1} \varphi)}$	Section III B

TABLE I. List of possible distributions derived from the joint target density in Equation (14). The normalizing constant Z is given by Equation (4) and $Z_\mathcal{N}$ is defined in Equation (10). Notes: (A) Only the joint, φ-conditional, and ϕ-conditional densities can be efficiently computed (up to normalization). (B) The φ-conditional can be sampled exactly by the method specified in Equation (16). (C) A closed form for the φ-marginal density is not generally known (even unnormalized).

Various MCMC schemes

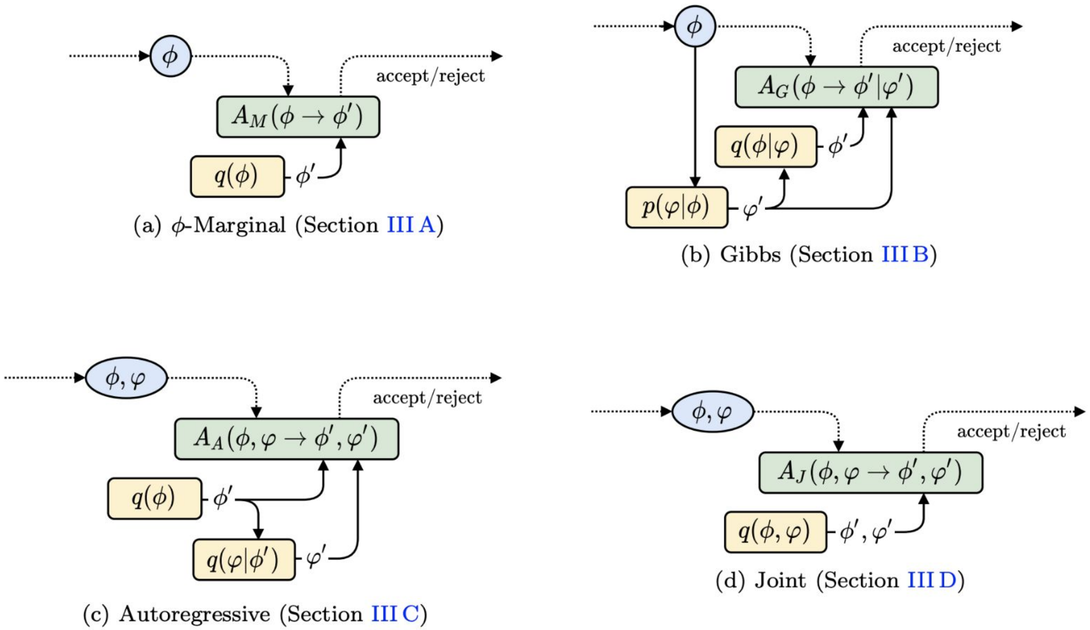
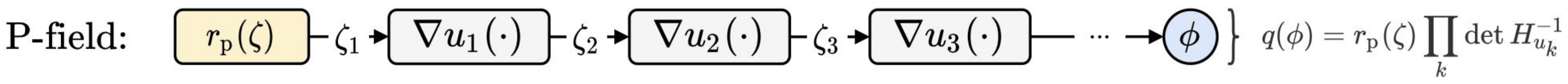


FIG. 1. Diagrams illustrating the four types of sampling schemes described in Section III. Blue circles/ellipses depict the current state of the Markov chain. Yellow boxes depict exactly sampleable densities either produced from generative models or by Equation (16). Green boxes correspond to Metropolis accept/reject steps using the acceptance probabilities defined in the text. Dotted lines indicate the Markov chain, whereas solid lines correspond to the internal operations of each Markov chain step.

The Convex Potential Yukawa Flow

$$p(\phi, \varphi) = \frac{1}{Z} \exp\left(-S_B(\phi) - \varphi^\dagger \left[\mathcal{M}(\phi)\right]^{-1} \varphi\right)$$



(a) ϕ-Marginal architecture based on convex potential flows (Section IV C 1).

$$q(\phi) = r_{\mathrm{p}}(\zeta) \prod_k \det H_{u_k}^{-1}$$

$$\varphi = \mathcal{A}(\phi)\chi, \quad \text{where} \quad \chi \sim \frac{1}{Z_{\mathcal{N}}} e^{-\chi^\dagger \chi}$$

Key challenge: scalable gradient estimation

$$\phi, \chi \sim \mathbb{N}(0, 1)$$

$$\phi' = \nabla H(\phi)$$

$$\chi' = D_{\phi'}(\chi)$$

Key challenge: scalable gradient estimation

Scalar flow grad LDJ

$$\begin{aligned}\nabla_\theta \mathrm{LDJ} &= \nabla_\theta \mathbb{E}\left[\mathrm{stopgrad}(z^T J_\theta^{-1}) J_\theta z\right] \\ &= \nabla_\theta \mathbb{E}\left[\mathrm{stopgrad}(\mathrm{CG}(J_\theta^T, z))^T J_\theta z\right]\end{aligned}$$

Fermion flow grad LDJ

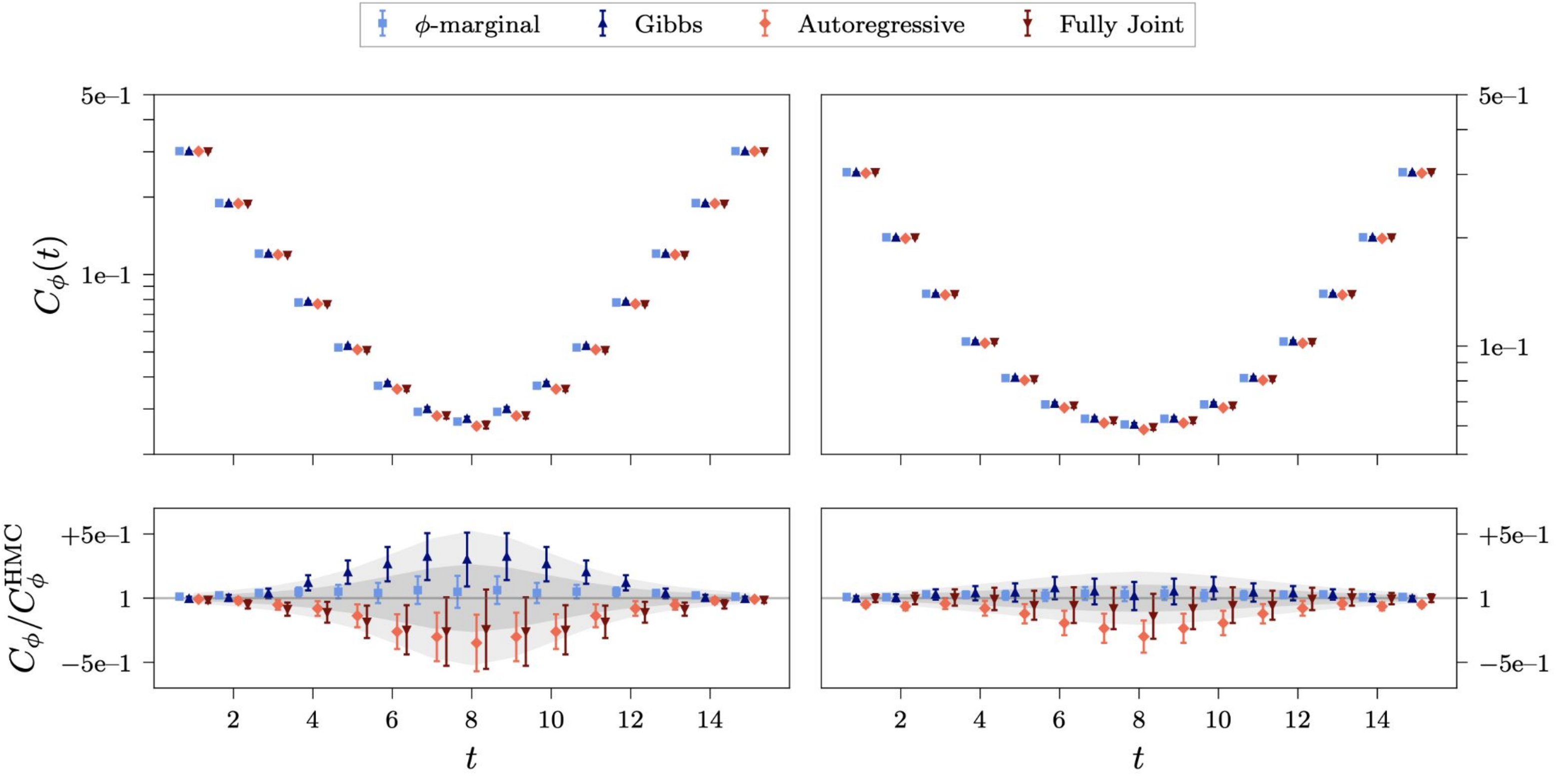
$$\begin{aligned}\nabla_\theta \mathrm{LDJ} &= \nabla_\theta \mathbb{E}\left[\mathrm{stopgrad}(J_\theta (J_\theta^T J_\theta + \kappa \mathbb{I})^{-1} z)^T J_\theta z\right] \\ &= \nabla_\theta \mathbb{E}\left[\mathrm{stopgrad}(J_\theta \mathrm{CG}(J_\theta^T J_\theta + \kappa \mathbb{I}, z))^T J_\theta z\right]\end{aligned}$$

Main Results: MCMC Acceptance rates

MCMC Approach	Modeled targets	Flow model	Parameters	Acc. rate	$\langle \lvert M \rvert \rangle$	$\langle \lvert \bar{\psi}\psi \rvert \rangle$	τ_M^{int}	$\tau_{\bar{\psi}\psi}^{\text{int}}$
ϕ-Marginal (III A)	$p(\phi)$	IV C 1	V B 1	92%	0.0734(1)	0.0159(1)	0.72(1)	0.71(1)
				92%	0.0792(1)	0.0491(1)	0.67(1)	0.67(1)
Gibbs (III B)	$p(\phi \vert \varphi)$	IV C 2	V B 2	60%	0.0735(1)	0.0160(1)	2.02(4)	2.02(3)
				44%	0.0792(1)	0.0490(1)	2.74(4)	2.73(4)
Autoregressive (III C)	$p(\phi), p(\varphi \vert \phi)$	IV C 3	V B 3	53%	0.0731(1)	0.0159(1)	2.16(3)	2.16(3)
				43%	0.0790(1)	0.0489(1)	3.62(7)	3.60(7)
Fully Joint (III D)	$p(\phi, \varphi)$	IV C 4	V B 4	37%	0.0733(1)	0.0159(1)	4.98(11)	4.98(11)
				31%	0.0791(1)	0.0490(1)	8.73(30)	8.67(30)

TABLE III. Sampling performance metrics and observables for all approaches, computed from 100 Markov chains with 10k proposals each, where the first 1k are discarded for thermalization. For each model, the first row shows results obtained for $g = 0.1$ and the second row for $g = 0.3$, respectively. For comparison, the values obtained with HMC listed in Table II are consistent with the measurements from our models. Autocorrelation times τ^{int} are computed for each of the 100 chains and then averaged, and errors are obtained with statistical jackknife. The results are discussed in more detail in Section V C. All models except the autoregressive make use of even-odd preconditioning of the action.

Main Results: Bias analysis



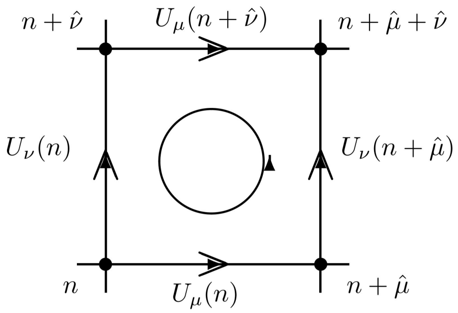
Summary

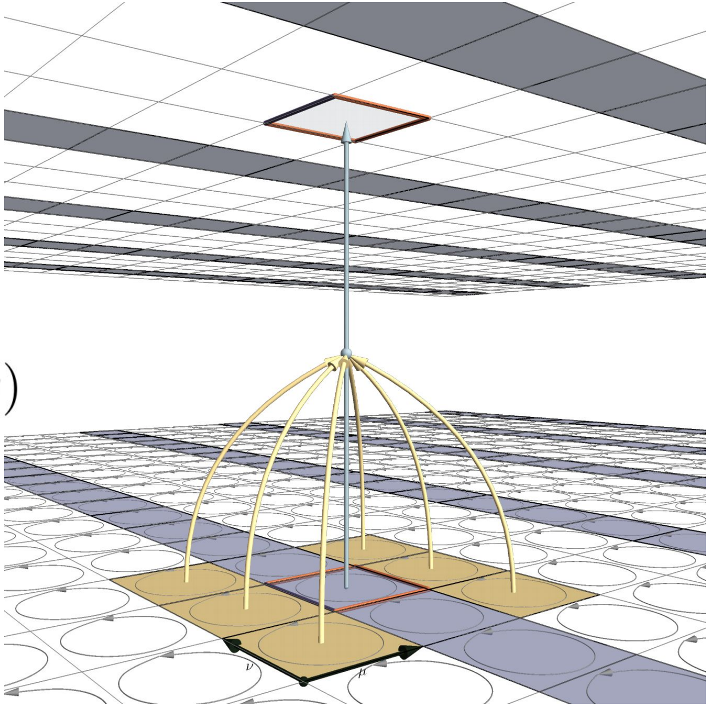
- Masked normalizing flows are a good family of models for 2D scalar fields
- They can incorporate translational symmetry and boundary conditions
- Introducing fermions add substantial complexity:
    - Requires working with scalar-pseudo-fermion effective action
    - Requires inversion and gradients of the operator DD* (expensive, can have large condition number)

DeepMind

U(N) and SU(N) equivariant flows: Sampling gauge and fermion fields at criticality

Lattice Quantum Chromodynamics





$$P_{\mu\nu}(x) := U(x,\mu)U(x+\hat{\mu},\nu)U^\dagger(x+\hat{\nu},\mu)U^\dagger(x,\nu)$$

$$S := -\beta \sum_{\text{sites } x} \sum_{\mu=1}^{D} \sum_{\nu=\mu+1}^{D} \mathrm{Re}\left[\frac{1}{N}\mathrm{Tr}\left(P_{\mu\nu}(x)\right)\right]$$

$$p(U) \propto e^{-\beta S[U]}$$

DeepMind

Abelian Gauge: U(1)

Modelling Gauge fields with flows

Equivariant Flow-Based Sampling for Lattice Gauge Theory

Gurtej Kanwar,[1] Michael S. Albergo,[2] Denis Boyda,[1] Kyle Cranmer,[2] Daniel C. Hackett,[1] Sébastien Racanière,[3] Danilo Jimenez Rezende,[3] and Phiala E. Shanahan[1]

[1]*Center for Theoretical Physics, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA*
[2]*Center for Cosmology and Particle Physics, New York University, New York, New York 10003, USA*
[3]*DeepMind Technologies Limited, 5 New Street Square, London EC4A 3TW, United Kingdom*

(Received 1 April 2020; revised 14 August 2020; accepted 24 August 2020; published 15 September 2020)

We define a class of machine-learned flow-based sampling algorithms for lattice gauge theories that are gauge invariant by construction. We demonstrate the application of this framework to U(1) gauge theory in two spacetime dimensions, and find that, at small bare coupling, the approach is orders of magnitude more efficient at sampling topological quantities than more traditional sampling procedures such as hybrid Monte Carlo and heat bath.

DOI: 10.1103/PhysRevLett.125.121601

DeepMind

SU(N) Yang-Mills Theory

Modelling Gauge fields with flows

PHYSICAL REVIEW D **103**, 074504 (2021)

Sampling using SU(N) gauge equivariant flows

Denis Boyda,[1,*] Gurtej Kanwar,[1,†] Sébastien Racanière,[2,‡] Danilo Jimenez Rezende,[2,§] Michael S. Albergo,[3]
Kyle Cranmer,[3] Daniel C. Hackett,[1] and Phiala E. Shanahan[1]

[1]*Center for Theoretical Physics, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA*

[2]*DeepMind, London N1C 4AG, United Kingdom*

[3]*Center for Cosmology and Particle Physics, New York University, New York, New York 10003, USA*

(Received 24 September 2020; accepted 16 March 2021; published 20 April 2021)

We develop a flow-based sampling algorithm for SU(N) lattice gauge theories that is gauge invariant by construction. Our key contribution is constructing a class of flows on an SU(N) variable [or on a U(N) variable by a simple alternative] that respects matrix conjugation symmetry. We apply this technique to sample distributions of single SU(N) variables and to construct flow-based samplers for SU(2) and SU(3) lattice gauge theory in two dimensions.

DOI: 10.1103/PhysRevD.103.074504

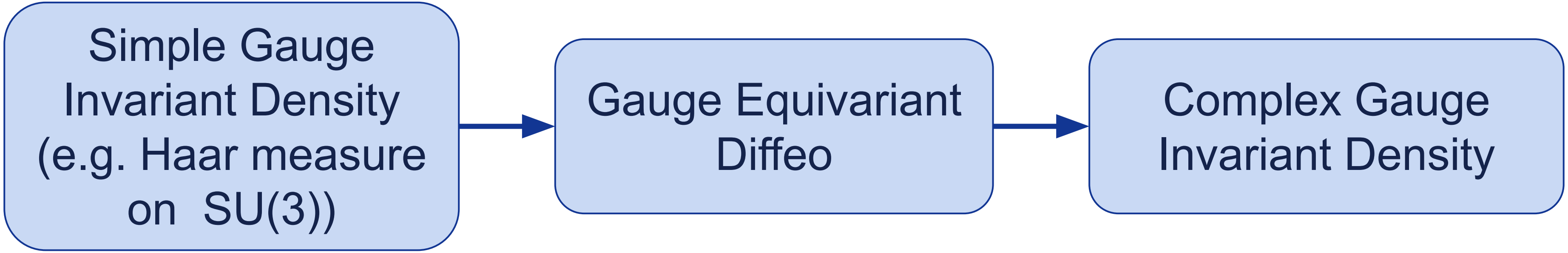
Continuous symmetries: Gauge transformations

$$U_\mu(x) \to \Omega(x) U_\mu(x) \Omega(x+\hat{\mu})^\dagger$$

$$P_{\mu\nu}(x) \to \Omega(x) P_{\mu\nu}(x) \Omega(x)^\dagger$$

$$\begin{aligned}\mathrm{Tr} P_{\mu\nu}(x) \to \mathrm{Tr}\,\Omega(x) P_{\mu\nu}(x) \Omega(x)^\dagger \\ = \mathrm{Tr} P_{\mu\nu}(x) \Omega(x)^\dagger \Omega(x) \\ = \mathrm{Tr} P_{\mu\nu}(x)\end{aligned}$$

General architecture: Pure-Gauge equivariant flow

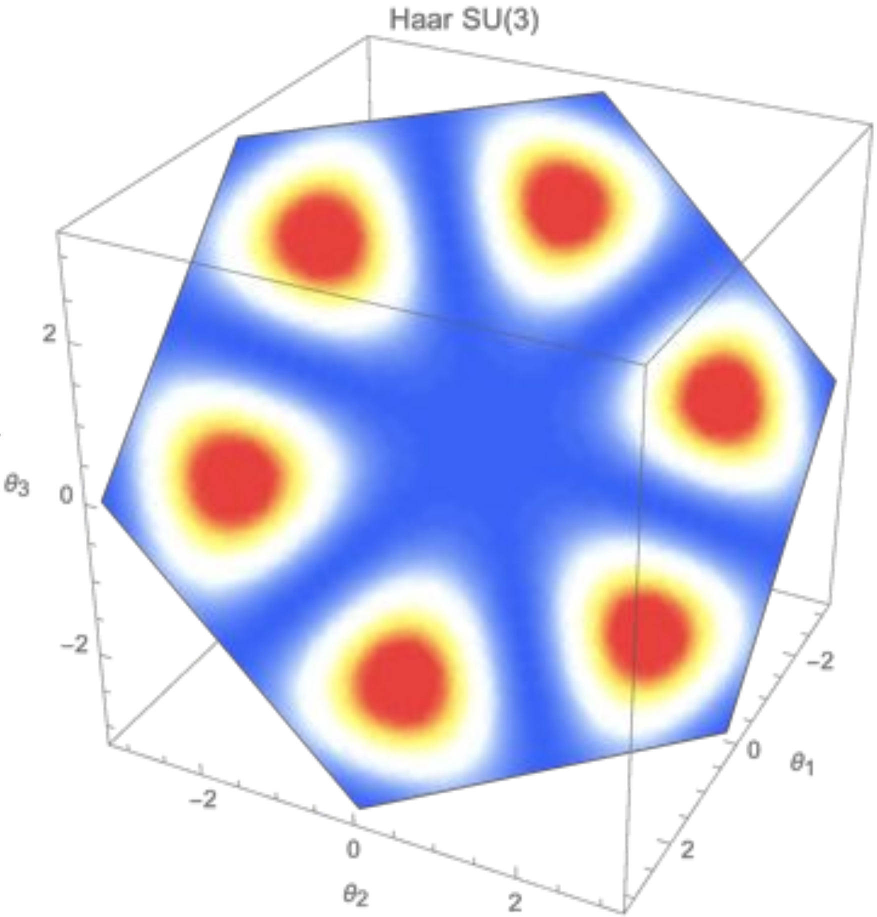


Haar measure on SU(3)

$$X \in \mathrm{SU}(3)$$

$$X = A \begin{pmatrix} e^{i\theta_1} & 0 & 0 \\ 0 & e^{i\theta_2} & 0 \\ 0 & 0 & e^{i\theta_3} \end{pmatrix} A^\dagger$$

$$\mathrm{Haar}(X) \propto \prod_{i>j} |e^{i\theta_i} - e^{i\theta_j}|^2$$



Gauge Equivariant Flow

$$Y_\mu(x) = f(U_\mu(x); \theta)$$

$$U_\mu(x) \to \Omega(x) U_\mu(x) \Omega(x + \hat{\mu})^\dagger$$

$$Y_\mu(x) \to \Omega(x) Y_\mu(x) \Omega(x + \hat{\mu})^\dagger$$

Gauge Equivariant Flow

Let h be an invertible map such that

$$h : \mathrm{SU}(N) \to \mathrm{SU}(N)$$

$$h(\Omega_\mu(x) X_\mu(x) \Omega_\mu(x)^\dagger) = \Omega_\mu(x) h(X_\mu(x)) \Omega_\mu(x)^\dagger$$

Then the map f,

$$f(X_\mu(x)) = h(P_{\mu\nu}(x)) S_{\mu\nu}(x)^\dagger$$

where $$S_{\mu\nu}(x) = X_\mu(x)^\dagger P_{\mu\nu}(x)$$

is equivariant to Gauge transformations

Gauge Equivariant Flow

This reduces the problem to finding a flow h such that

$$h : SU(N) \to SU(N)$$

$$h(\Omega_\mu(x) X_\mu(x) \Omega_\mu(x)^\dagger; \theta) = \Omega_\mu(x) h(X_\mu(x); \theta) \Omega_\mu(x)^\dagger$$

This is a flow equivariant to matrix conjugation transformations

Matrix-conjugation equivariant flows on SU(N) and U(N)

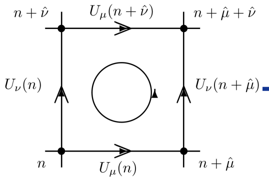
This flow is equivariant to matrix-conjugation transformations

$$(X, D = \mathrm{diag}(w)) = \mathrm{eigen}(U)$$

$$Y = X \mathrm{diag}(g(w)) X^{\dagger}$$

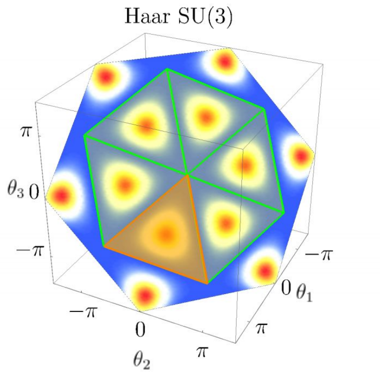
If g is a permutation-equivariant flow that preserves unitarity (\prod g(w) = 1)

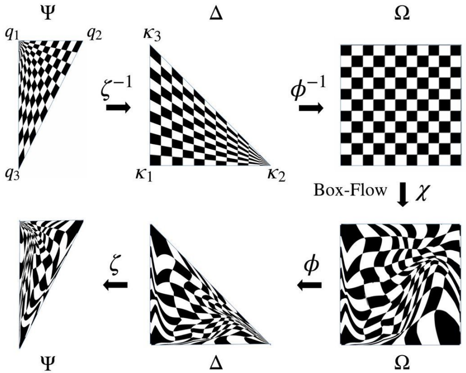
Our approach: an onion flow



Conjugation equivariant flow on U(N), SU(N). Low dimensional group of symmetries. → Flow on a torus with a finite symmetry group → Flow on a simplex → Flow on a box

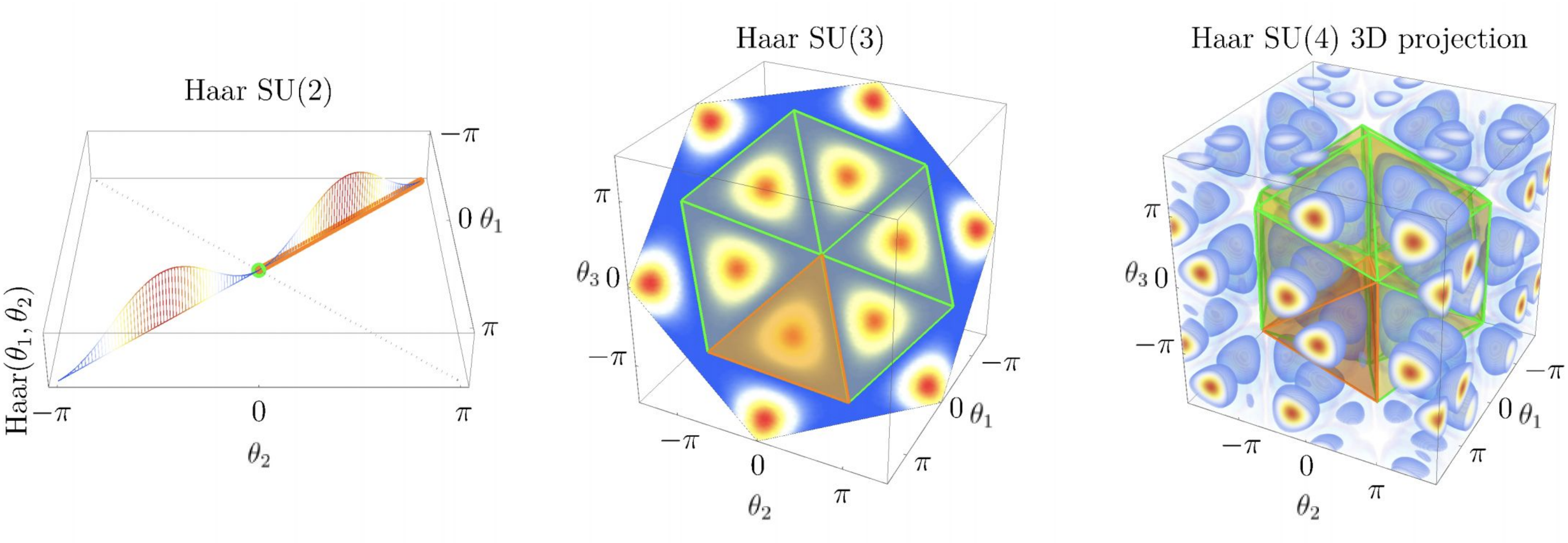
$$\begin{pmatrix} e^{i\theta_1} & 0 & \cdots & 0 \\ 0 & e^{i\theta_2} & & \vdots \\ \vdots & & \ddots & 0 \\ 0 & \cdots & 0 & e^{i\theta_N} \end{pmatrix}$$





Building Equivariant flows: Permutation Equivariant Flows on maximal toruses

Canonicalize -> Flow on cell -> Uncanonicalize

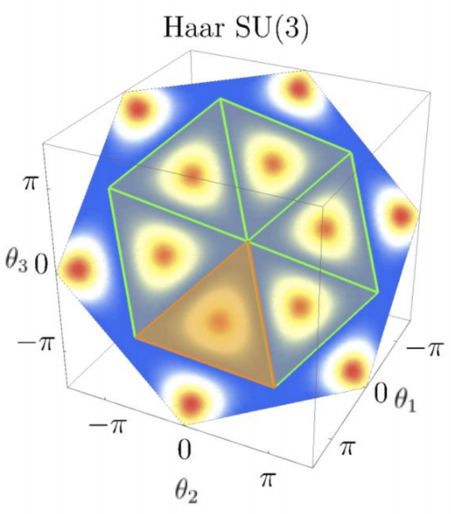


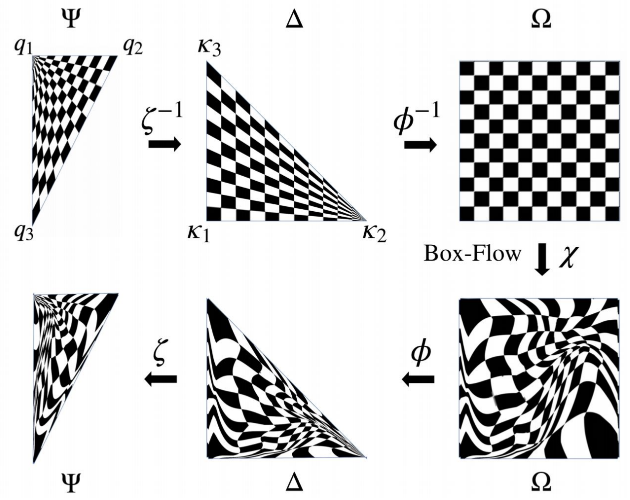
Boyda, D., Kanwar, G., Racanière, S., Rezende, D.J., Albergo, M.S., Cranmer, K., Hackett, D.C. and Shanahan, P.E., 2020. Sampling using SU(N) gauge equivariant flows. arXiv preprint arXiv:2008.05456.

Bender, C., O'Connor, K., Li, Y., Garcia, J.J., Zaheer, M. and Oliva, J., 2019. Exchangeable Generative Models with Flow Scans. arXiv preprint arXiv:1902.01967.

Building Equivariant flows: Permutation Equivariant Flows

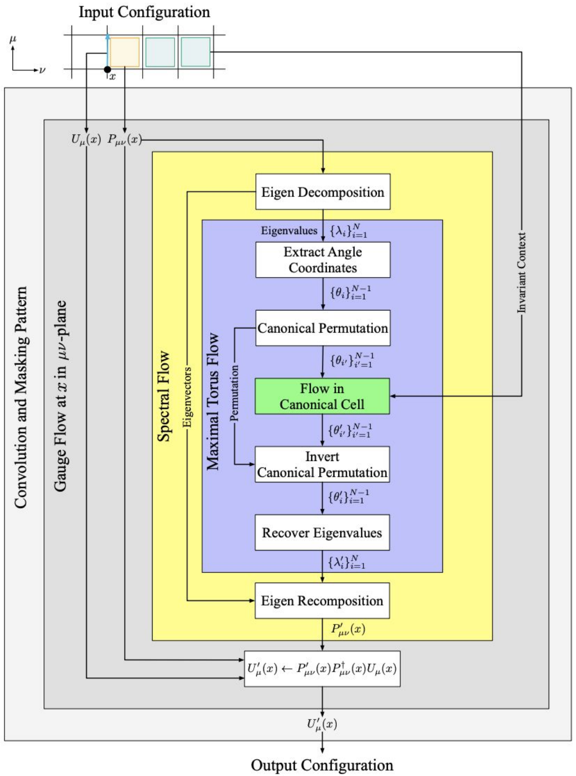
For special unitary groups permutation/Weyl equivariant flows reduces to a flow on a N-simplex

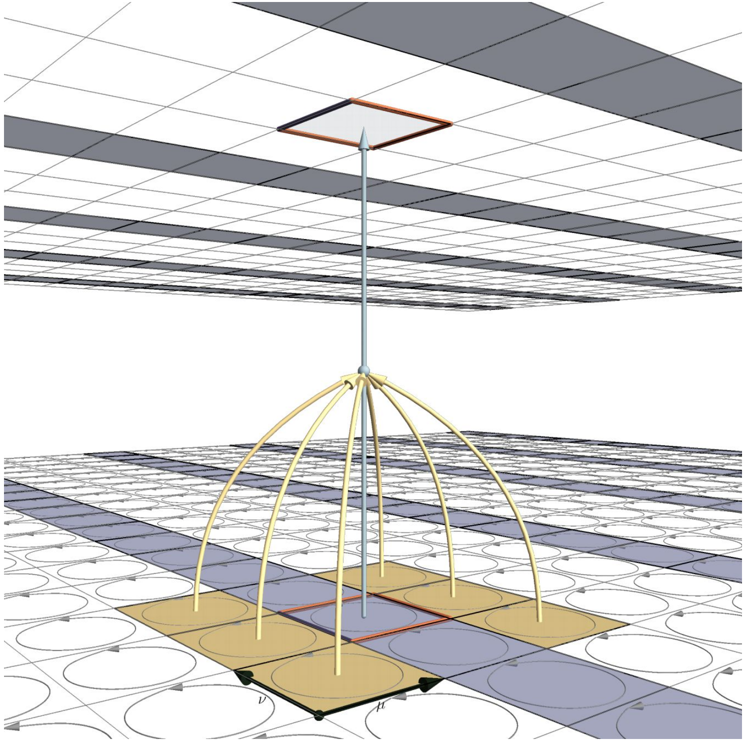




Boyda, D., Kanwar, G., Racanière, S., Rezende, D.J., Albergo, M.S., Cranmer, K., Hackett, D.C. and Shanahan, P.E., 2020. Sampling using SU (N) gauge equivariant flows. arXiv preprint arXiv:2008.05456.

SU(3) Gauge equivariant flow





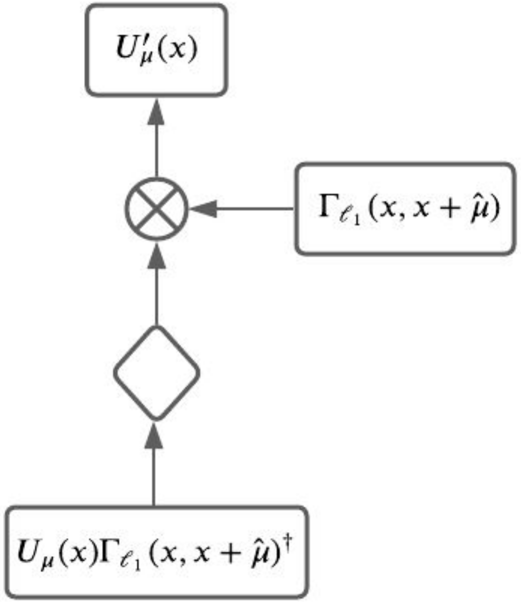
Boyda, D., Kanwar, G., Racanière, S., Rezende, D.J., Albergo, M.S., Cranmer, K., Hackett, D.C. and Shanahan, P.E., 2020. Sampling using $ SU (N) $ gauge equivariant flows. arXiv preprint arXiv:2008.05456.

TL;DR Gauge equivariant Flows

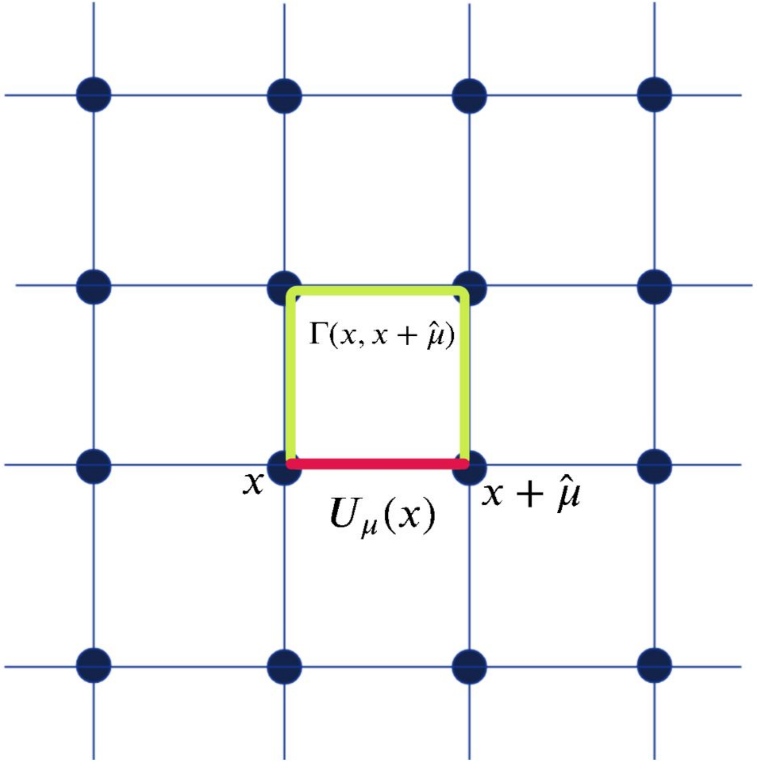
$\diamond$ Matrix conjugation equivariant map

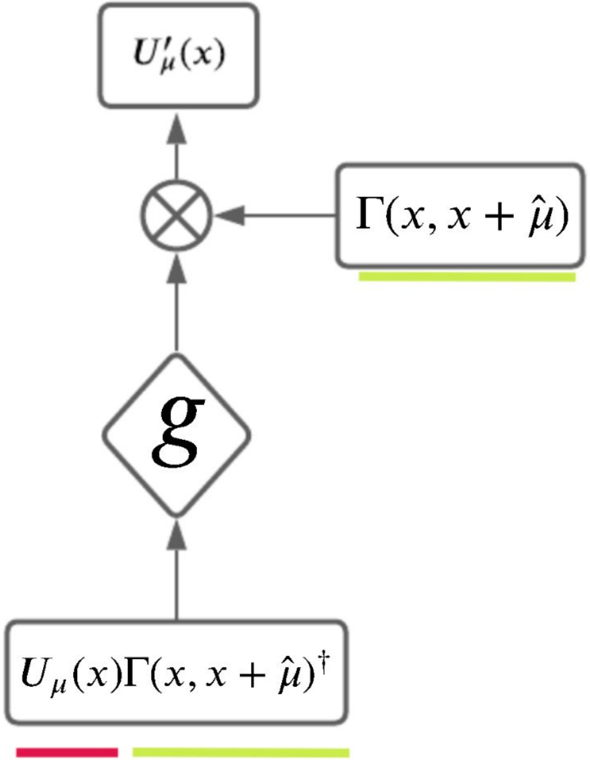
$\otimes$ Matrix product

$$\diamond(\Omega X \Omega^\dagger) = \Omega \diamond(X) \Omega^\dagger$$

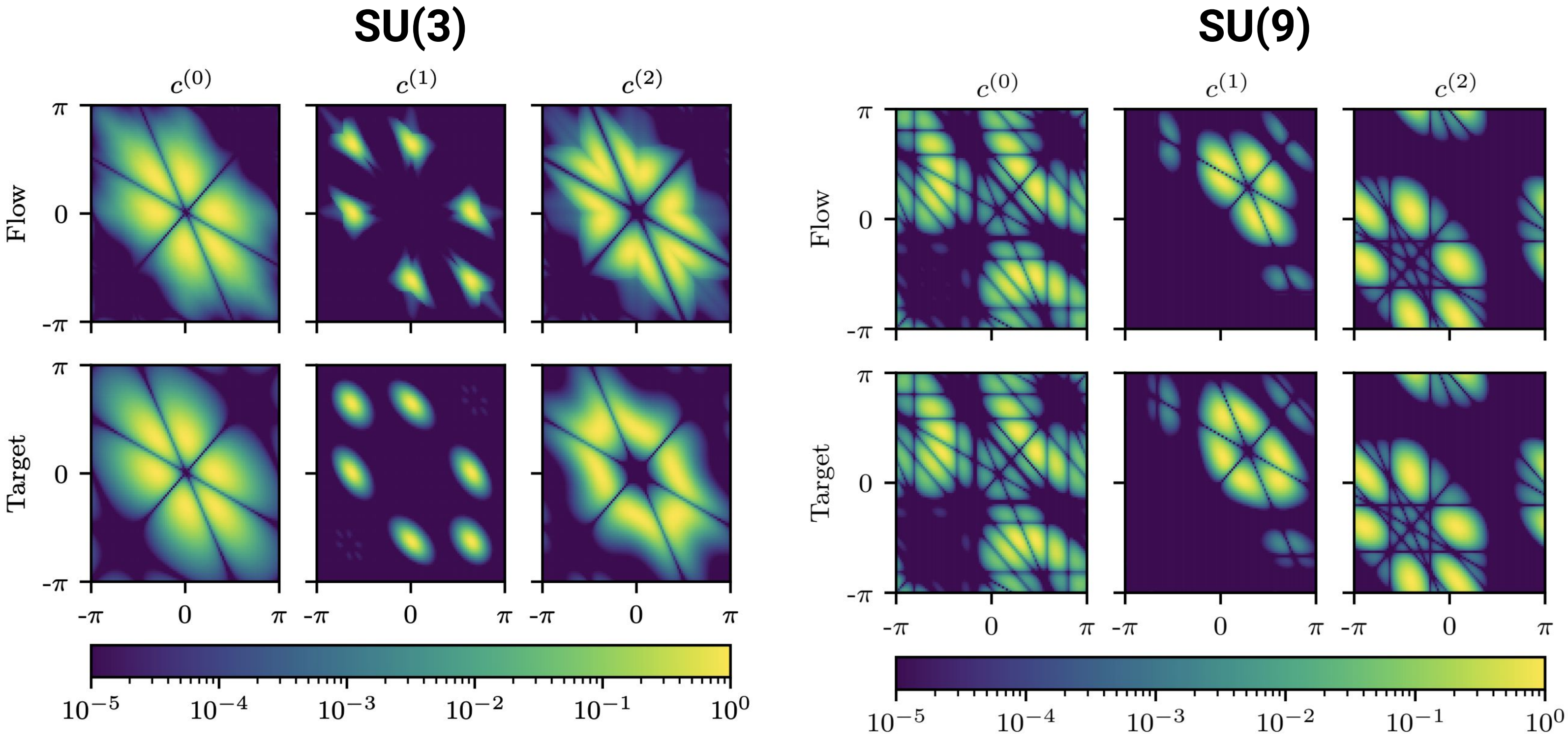


High-level pure Gauge flow



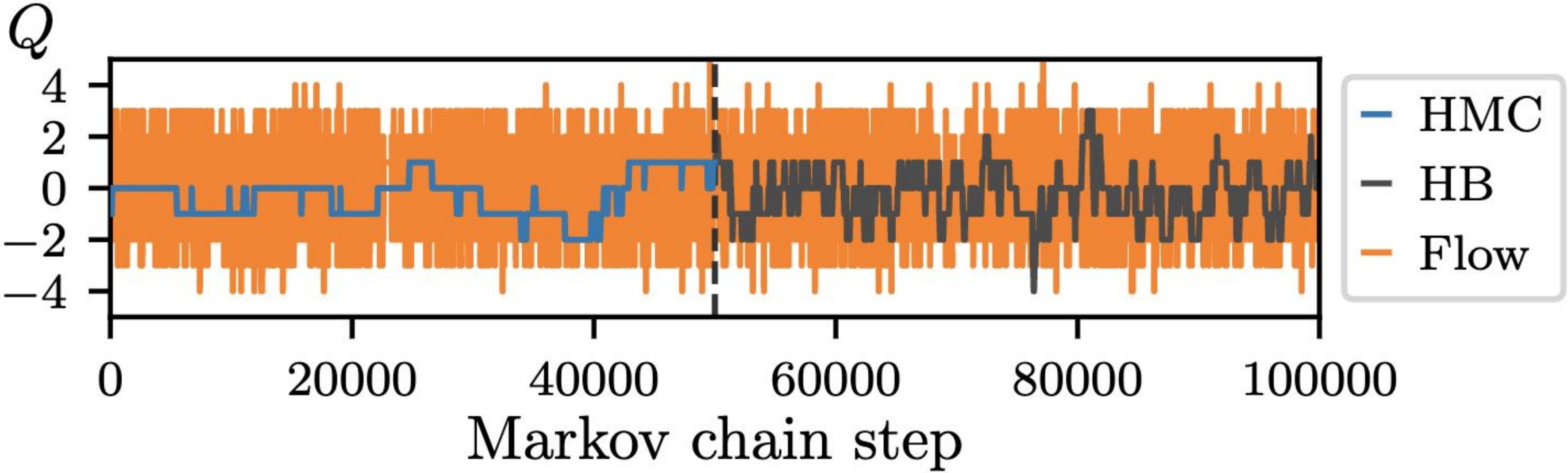


Building Gauge Equivariant flows: SU(N>3) Gauge equivariant flows: Simulating pure Gauge QCD



Boyda, D., Kanwar, G., Racanière, S., Rezende, D.J., Albergo, M.S., Cranmer, K., Hackett, D.C. and Shanahan, P.E., 2020. Sampling using SU(N) gauge equivariant flows. arXiv preprint arXiv:2008.05456.

Critical slowdown regime in 2D for U(1): Evidence of faster mixing rates with flow-based MCMC



DeepMind

The Schwinger model: U(1) Gauge + fermions in 2D

Modelling Gauge & fermion fields with flows

Flow-based sampling in the lattice Schwinger model at criticality

Michael S. Albergo,[1] Denis Boyda,[2,3,4] Kyle Cranmer,[1] Daniel C. Hackett,[3,4] Gurtej Kanwar,[5,3,4] Sébastien Racanière,[6] Danilo J. Rezende,[6] Fernando Romero-López,[3,4] Phiala E. Shanahan,[3,4] and Julian M. Urban[7]

[1] *Center for Cosmology and Particle Physics, New York University, New York, NY 10003, USA*
[2] *Argonne Leadership Computing Facility, Argonne National Laboratory, Lemont IL-60439, USA*
[3] *Center for Theoretical Physics, Massachusetts Institute of Technology, Cambridge, MA 02139, USA*
[4] *The NSF AI Institute for Artificial Intelligence and Fundamental Interactions*
[5] *Albert Einstein Center, Institute for Theoretical Physics, University of Bern, 3012 Bern, Switzerland*
[6] *DeepMind, London, UK*
[7] *Institut für Theoretische Physik, Universität Heidelberg, Philosophenweg 16, 69120 Heidelberg, Germany*

Recent results suggest that flow-based algorithms may provide efficient sampling of field distributions for lattice field theory applications, such as studies of quantum chromodynamics and the Schwinger model. In this work, we provide a numerical demonstration of robust flow-based sampling in the Schwinger model at the critical value of the fermion mass. In contrast, at the same parameters, conventional methods fail to sample all parts of configuration space, leading to severely underestimated uncertainties.

Schwinger model at criticality

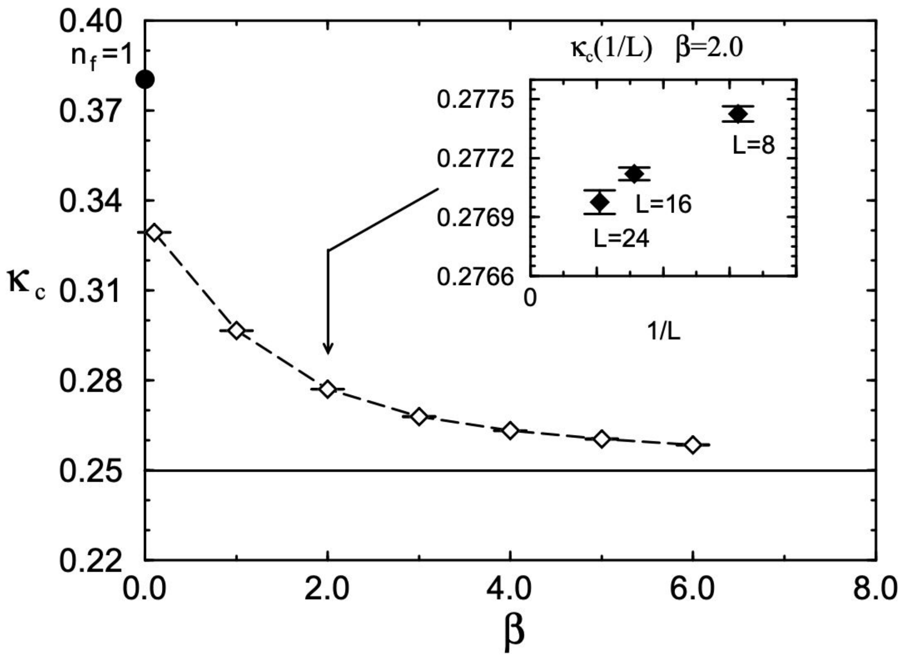
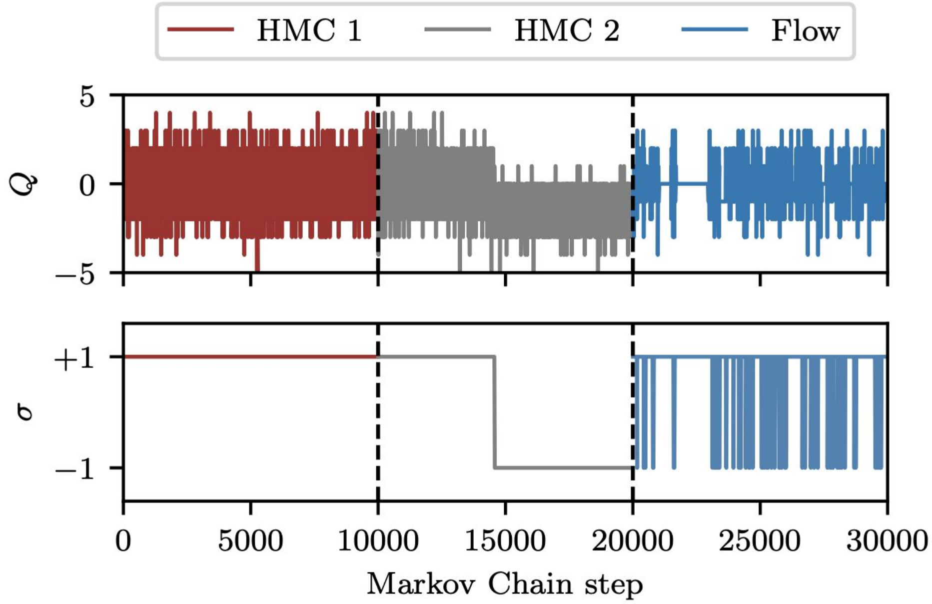
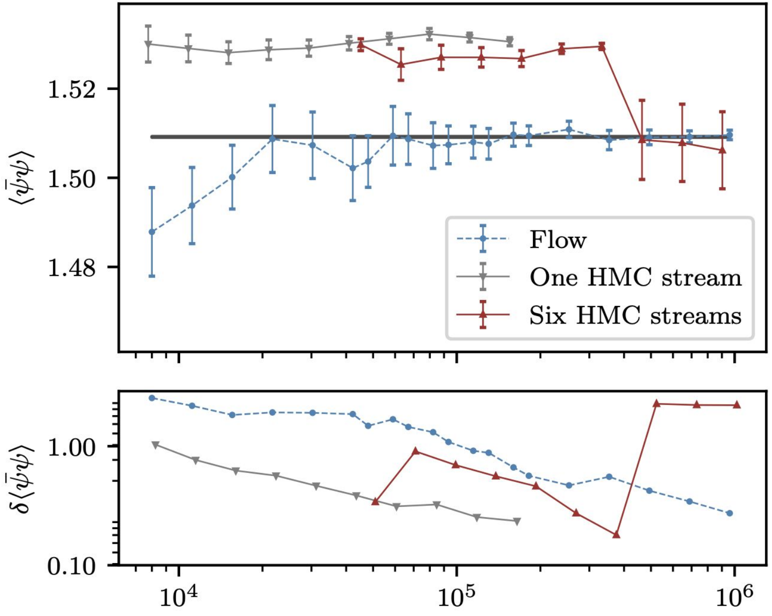


Figure 3. Phase diagram for the 2-flavour model for the 16 × 16 lattice (dashed lines to guide the eye); the value for the 1-flavour model is from [7].

Schwinger model at critical mass: Evidence of faster mixing rates with flow-based MCMC





DeepMind

2D QCD: SU(3) Gauge + Quarks

Modelling Gauge & fermion fields with flows

Gauge-equivariant flow models for sampling in lattice field theories with pseudofermions

Ryan Abbott,[1,2] Michael S. Albergo,[3] Denis Boyda,[4,1,2] Kyle Cranmer,[3]
Daniel C. Hackett,[1,2] Gurtej Kanwar,[5,1,2] Sébastien Racanière,[6] Danilo J. Rezende,[6]
Fernando Romero-López,[1,2] Phiala E. Shanahan,[1,2] Betsy Tian,[1] and Julian M. Urban[7]

[1] *Center for Theoretical Physics, Massachusetts Institute of Technology, Cambridge, MA 02139, USA*
[2] *The NSF AI Institute for Artificial Intelligence and Fundamental Interactions*
[3] *Center for Cosmology and Particle Physics, New York University, New York, NY 10003, USA*
[4] *Argonne Leadership Computing Facility, Argonne National Laboratory, Lemont IL-60439, USA*
[5] *Albert Einstein Center, Institute for Theoretical Physics, University of Bern, 3012 Bern, Switzerland*
[6] *DeepMind, London, UK*
[7] *Institut für Theoretische Physik, Universität Heidelberg, Philosophenweg 16, 69120 Heidelberg, Germany*

This work presents gauge-equivariant architectures for flow-based sampling in fermionic lattice field theories using pseudofermions as stochastic estimators for the fermionic determinant. This is the default approach in state-of-the-art lattice field theory calculations, making this development critical to the practical application of flow models to theories such as QCD. Methods by which flow-based sampling approaches can be improved via standard techniques such as even/odd preconditioning and the Hasenbusch factorization are also outlined. Numerical demonstrations in two-dimensional U(1) and SU(3) gauge theories with $N_f = 2$ flavors of fermions are provided.

Fermions?

$$\mathcal{L} = \mathcal{L}_{\text{Gauge}} + \sum_f \psi_f^\dagger D_f \psi_f$$

Grassmann fields

$$D_f = i\partial\!\!\!/ - m_f - g\phi$$

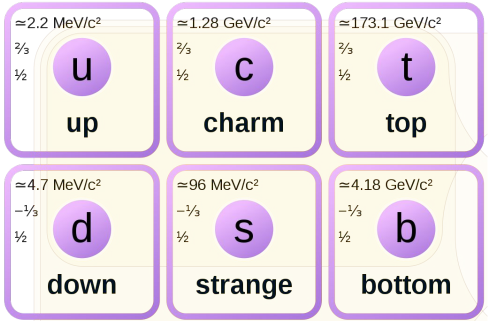
Commuting vector field

$$(\det DD^\dagger)^{1/2} \propto \int \mathcal{D}\chi^\dagger \mathcal{D}\chi e^{-\frac{1}{2}\chi^T (DD^\dagger)^{-1}\chi}$$

$$\mathcal{L}_{\text{eff}}(U,\chi) = \mathcal{L}_{\text{Gauge}}(U) - \sum_f \chi_f^\dagger (DD^\dagger)^{-1} \chi_f$$

Incorporating Quarks

$$S_f(\bar{q}, q, U) = a^4 \sum_{x,y} \bar{q}_f(x) D_f[U](x,y) q_f(y)$$



$$\mathcal{L}_{\text{eff}}(U, \chi) = \mathcal{L}_{\text{Gauge}}(U) - \sum_f \chi_f^\dagger (DD^\dagger)^{-1} \chi_f$$

Continuous symmetries: Gauge transformations

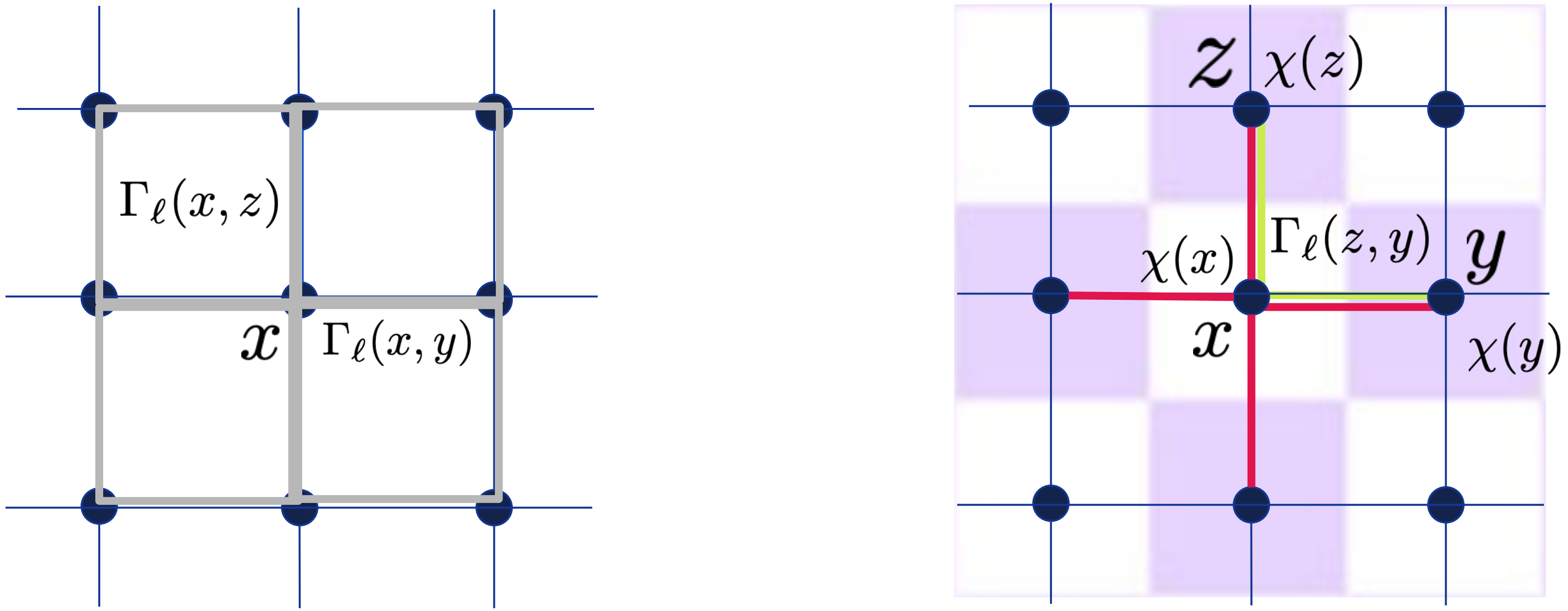
$$U_\mu(x) \in \mathrm{SU}(N) \qquad \Omega(x) \in \mathrm{SU}(N) \qquad \chi(x) \in \mathbb{C}^N$$

$$\begin{aligned}
U_\mu(x) &\rightarrow \Omega(x) U_\mu(x) \Omega^\dagger(x+\hat{\mu}) \\
\psi(x) &\rightarrow \Omega(x)\psi(x) \\
\psi^\dagger(x) &\rightarrow \psi^\dagger(x)\Omega^\dagger(x) \\
\psi(x)\psi^\dagger(y) &\rightarrow \Omega(x)\psi(x)\psi^\dagger(y)\Omega^\dagger(y) \\
\Gamma_\ell(x,y) &\rightarrow \Omega(x)\Gamma_\ell(x,y)\Omega^\dagger(y) \\
\Gamma_\ell(x,y)\psi(y) &\rightarrow \Omega(x)\Gamma_\ell(x,y)\psi(y)
\end{aligned}$$

$$\alpha = f_\theta(\mathrm{Tr}\Gamma_\ell(x,x), \chi(z)\Gamma_\ell(z,y)\chi(y))$$

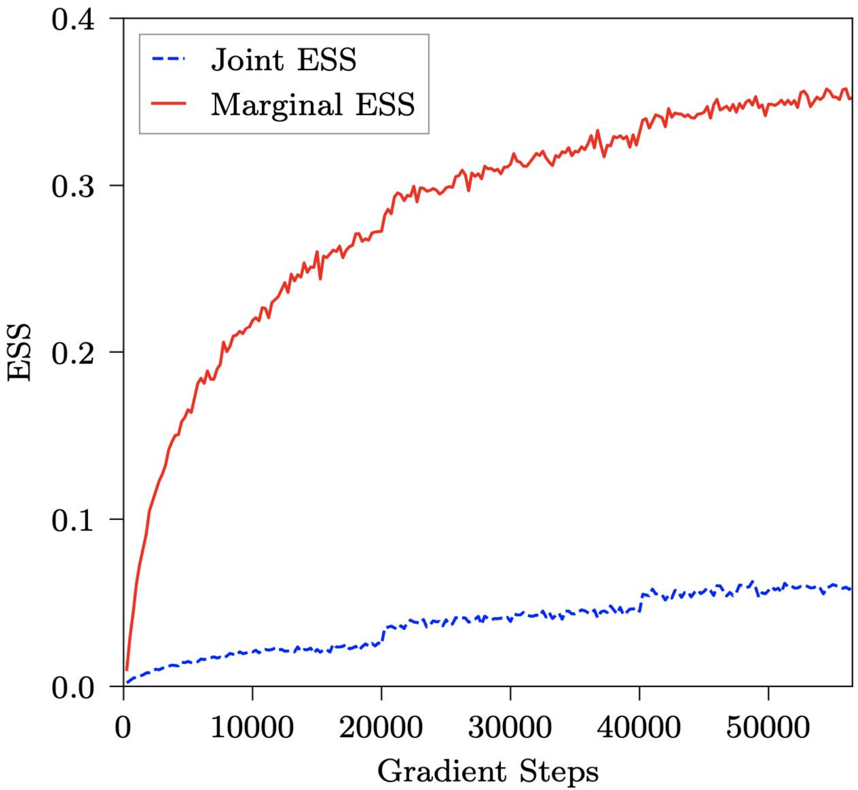
Parallel-transported fields

$$\chi(x)' = e^{s_\alpha}\chi(x) + \sum_{\ell \in F_o} w_\alpha(\ell)\Gamma_\ell(x,y)\chi(y)$$

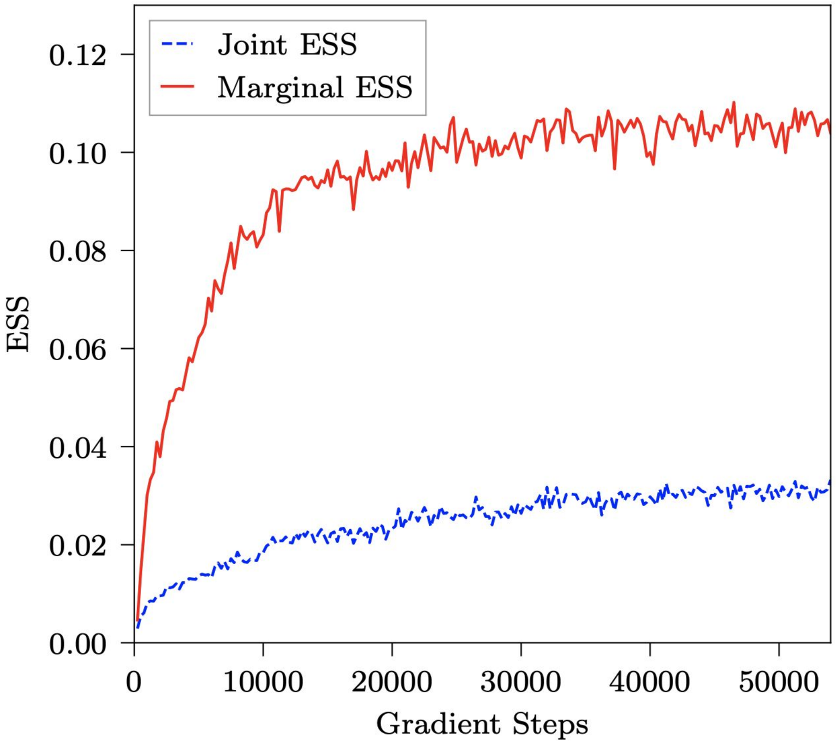


Performance results (L=16, U(1) / SU(3) + fermions)

U(1)



SU(3)



DeepMind

Towards 4D QCD: SU(3) Gauge + Quarks

Modelling Gauge & fermion fields with flows

Sampling QCD field configurations with gauge-equivariant flow models

Ryan Abbott,[a,b] Michael S. Albergo,[c] Aleksandar Botev,[g] Denis Boyda,[a,b,d] Kyle Cranmer,[c,e] Daniel C. Hackett,[a,b] Gurtej Kanwar,[a,b,f] Alexander G. D. G. Matthews,[g] Sébastien Racanière,[g] Ali Razavi,[g] Danilo J. Rezende,[g] Fernando Romero-López,[a,b] Phiala E. Shanahan[a,b,*] and Julian M. Urban[a,b,h]

[a] *Center for Theoretical Physics, Massachusetts Institute of Technology, Cambridge, MA 02139, USA*
[b] *The NSF AI Institute for Artificial Intelligence and Fundamental Interactions*
[c] *Center for Cosmology and Particle Physics, New York University, New York, NY 10003, USA*
[d] *Argonne Leadership Computing Facility, Argonne National Laboratory, Lemont IL 60439, USA*
[e] *Physics Department, University of Wisconsin-Madison, Madison, WI 53706, USA.*
[f] *Albert Einstein Center, Institute for Theoretical Physics, University of Bern, 3012 Bern, Switzerland*
[g] *DeepMind, London, UK*
[h] *Institut für Theoretische Physik, Universität Heidelberg, Philosophenweg 16, 69120 Heidelberg, Germany*

E-mail: phiala@mit.edu

Machine learning methods based on normalizing flows have been shown to address important challenges, such as critical slowing-down and topological freezing, in the sampling of gauge field configurations in simple lattice field theories. A critical question is whether this success will translate to studies of QCD. This Proceedings presents a status update on advances in this area. In particular, it is illustrated how recently developed algorithmic components may be combined to construct flow-based sampling algorithms for QCD in four dimensions. The prospects and challenges for future use of this approach in at-scale applications are summarized.

Full QCD experiments (4D, L=4)

- Plaquette:

$$P = \frac{1}{N_c}\frac{1}{L^4}\sum_x \sum_{\mu<\nu} \mathrm{Re\ tr}\, P_{\mu\nu}(x), \tag{2}$$

where $N_c = 3$ is the number of colors, $L = 4$ is the extent of the lattice geometry, and $P_{\mu\nu}$ denotes the 1×1 Wilson loop which extends in the μ and ν directions;

- Polyakov loop:

$$L = \frac{1}{L^3}\sum_{\vec{x}} \mathrm{tr} \prod_{x_0} U_0(x_0, \vec{x}), \tag{3}$$

where U_0 is the gauge link in the time direction;

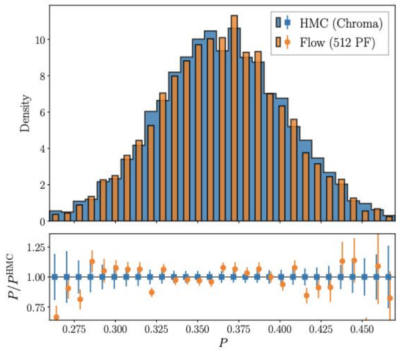
- Pion correlation function:

$$C_\pi(x_0) = -\sum_{\vec{x}} \langle [\bar{u}\gamma_5 d](x_0, \vec{x})[\bar{d}\gamma_5 u](0, \vec{0})\rangle, \tag{4}$$

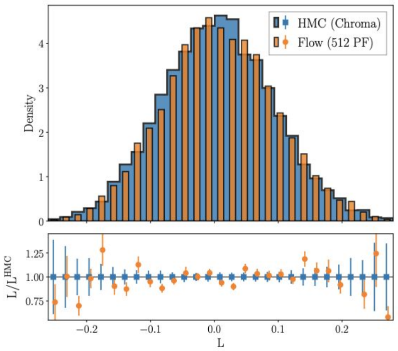
measured using point sources;

- Topological charge:

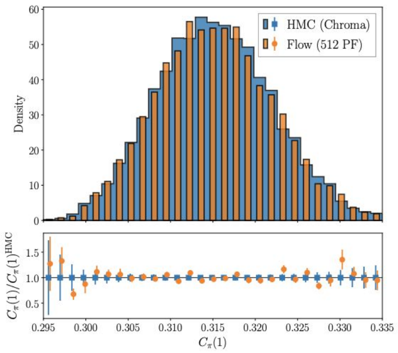
$$Q = \frac{1}{16\pi^2}\sum_x \epsilon_{\mu\nu\rho\sigma} F_{\mu\nu}(x) F_{\rho\sigma}(x), \tag{5}$$



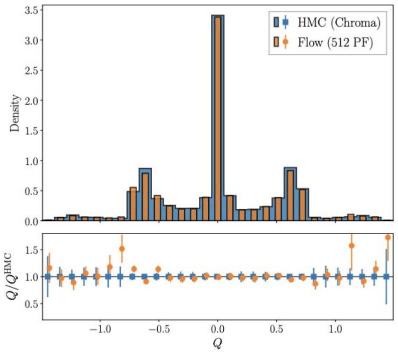
(a) Plaquette



(b) Polyakov loop



(c) Pion correlation function at $x_0 = 1$



(d) Topological charge at $t/a^2 = 4$

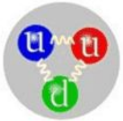
DeepMind

Towards physical calculations: Hadron Spectroscopy

Towards real calculations: Hadron Spectroscopy

Hadrons

Baryons are composed of three quarks



Baryons

Mesons are composed of one quark and one antiquark

Mesons

Nucleons

Particle	Mass (MeV/c^2)	τ (sec)
p	938.2	$>10^{11}$
n	939.5	10^3

Hyperons

Particle	Mass (MeV/c^2)	τ (sec)
Λ	1115	2.6×10^{-10}
Σ^+	1189	0.8×10^{-10}
Σ^0	1192	10^{-14}
Σ^-	1197	1.6×10^{-10}
Ξ^0	1314	3×10^{-10}
Ξ^-	1321	1.8×10^{-10}
Ω^-	1675	1.3×10^{-10}

Pions

Particle	Mass (MeV/c^2)	τ (sec)
π^-, π^+	139	2.5×10^{-8}
π^0	135	1.8×10^{-16}

Kaons

Particle	Mass (MeV/c^2)	τ (sec)
K^-, K^+	494	1.2×10^{-8}
K^0	498	
η	550	10^{-18}

Spectroscopy: From correlators to particle mass

Average over model samples

$$\left\langle \eta_\pi(\boldsymbol{p},t)\,\eta_\pi^\dagger(\boldsymbol{p},t')\right\rangle = -\frac{1}{V_\mathrm{s}}\sum_{\boldsymbol{x},\boldsymbol{x}'} e^{-\mathrm{i}\boldsymbol{p}(\boldsymbol{x}-\boldsymbol{x}')}\left\langle \mathrm{Tr}\left[\mathcal{D}_u^{-1}(x,x')\gamma_4\Gamma^\dagger\gamma_4\mathcal{D}_d^{-1}(x',x)\Gamma\right]\right\rangle_\mathrm{G}$$

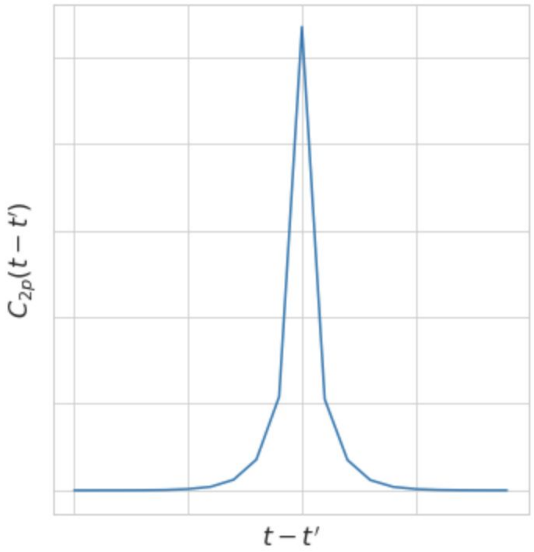
$$\left\langle \eta_\pi(\boldsymbol{p},t)\eta_\pi^\dagger(\boldsymbol{p},t')\right\rangle = e^{-E_\pi(\boldsymbol{p})T/2}\left|\langle 0|\,\eta_\pi(\boldsymbol{p})\,|\pi(\boldsymbol{p})\rangle\right|^2 2\cosh\left[\left(T/2-(t-t')\right)E_\pi(\boldsymbol{p})\right]+\ldots$$

Particle energy at momentum p

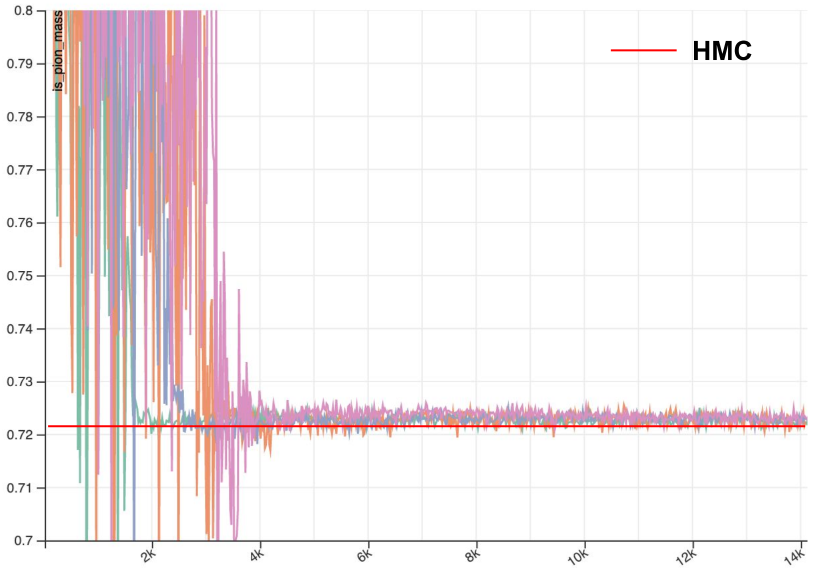
$$E^2(p) = m^2c^4 + p^2c^2$$

Spectroscopy: From correlators to particle mass

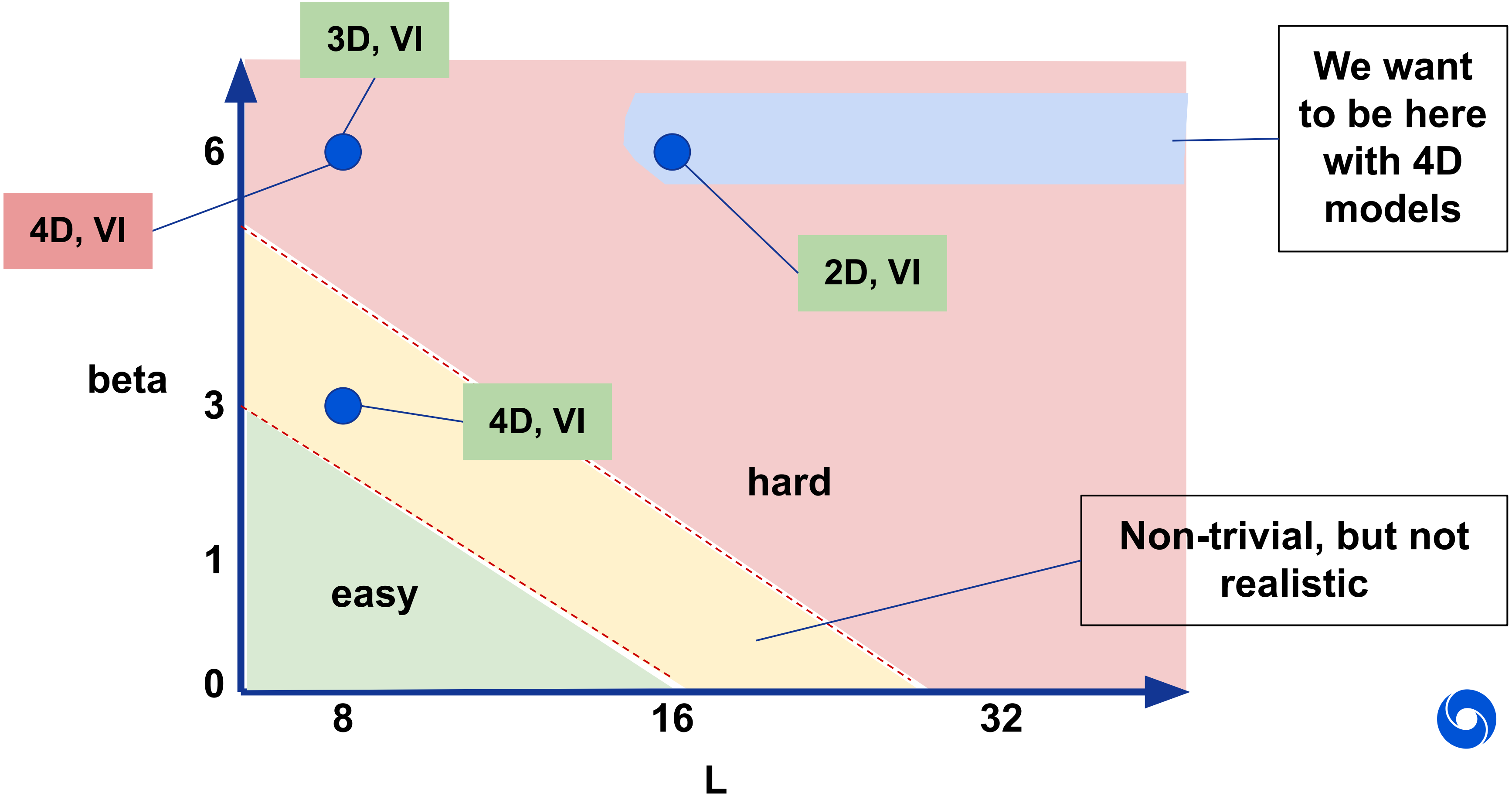
Pion correlator (p=0)



Pion mass vs model training



How hard is to reach physically meaningful settings?



Summary

- We can construct flows with U(N) and SU(N) Gauge symmetry
- In 2D results are quite promising
- They can also be extended to include pseudo-fermion transformations
- Based on Yukawa and Schwinger models, introducing fermions adds substantial complexity:
  - Require working with pseudo-fermion effective action
  - Require inversion of the operator DD* (expensive, can have very large condition number)
  - Increased combinatorics:
    - Much larger space of Gauge-invariant quantities to consider

DeepMind

Discussion

Summary

- **Remarkable progress in the development of NFs** for sampling and free energy estimation (from LQCD to molecular systems).

- NFs allow us to **address old problems in completely new ways** by leveraging the flexibility of neural networks.

- **Challenges and limitations**:
  - Training and evaluating models without ground-truth samples
  - Scaling up to larger and more complex systems
  - Need more general and robust mechanisms to correct for model bias and bound error of expectations