

Generative AI and Lattice Gauge Theory

Partially based on work in [2510.26081](#), [2512.19877](#)

Octavio Vega

University of Illinois Urbana-Champaign

May 18, 2026 – LITP Spring Symposium

Collaborators

Marina Marinkovic



ETH zürich

Javad Komijani



ETH zürich

Gurtej Kanwar



 THE UNIVERSITY
of EDINBURGH

Aida El-Khadra



I ILLINOIS

Octavio Vega



I ILLINOIS

Monte Carlo and Lattice QFT

Quantum Field Theory

Quantum field theory (QFT) describes many systems across physics

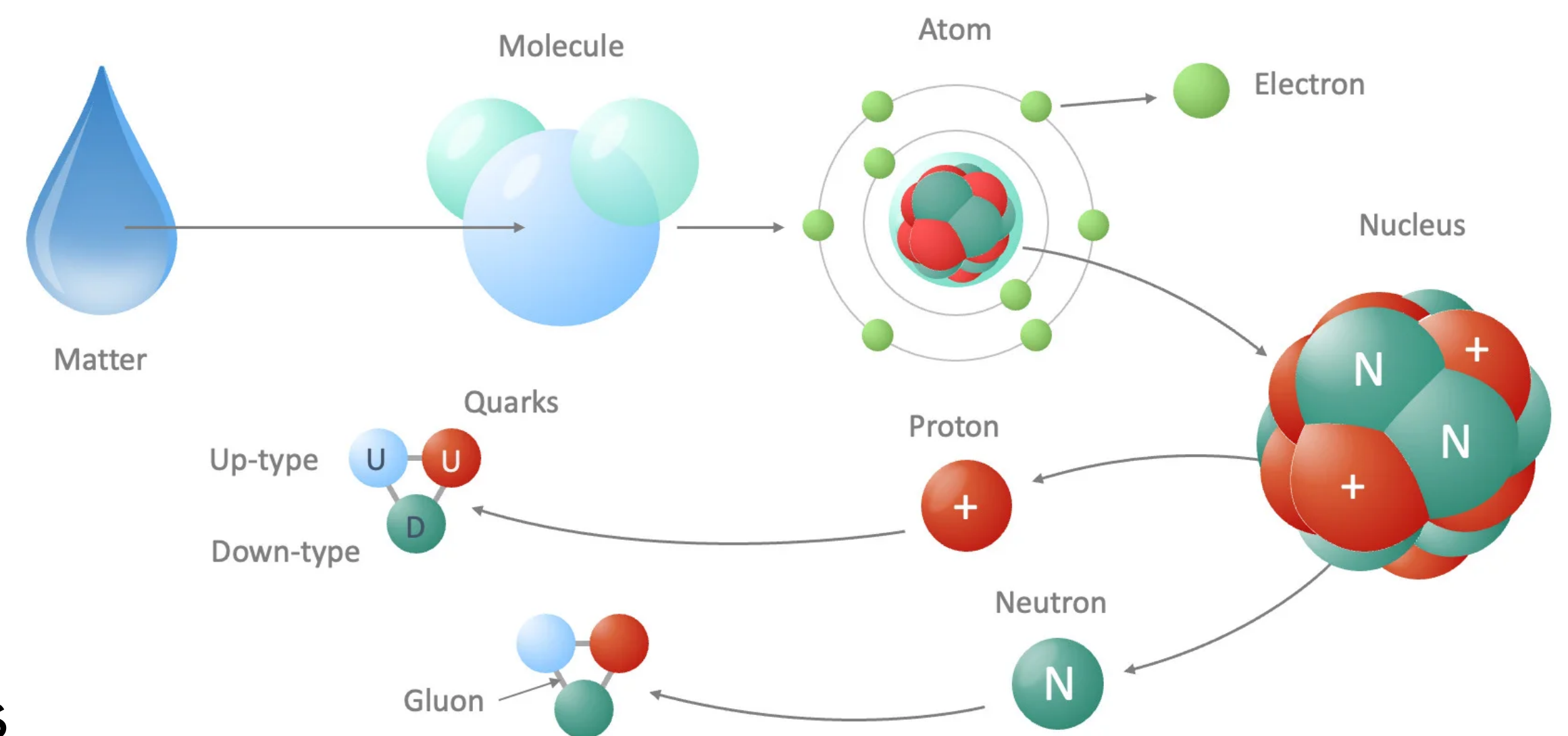
- Fundamental laws of nature and particle physics

Strong interactions described by **quantum chromodynamics (QCD)**

Some theories (e.g. QCD) **hard to solve**

→ perturbative at high energies

→ need computational methods at low energies



Why Lattice Field Theory?

Wilson, *PRD* **10** (1974)

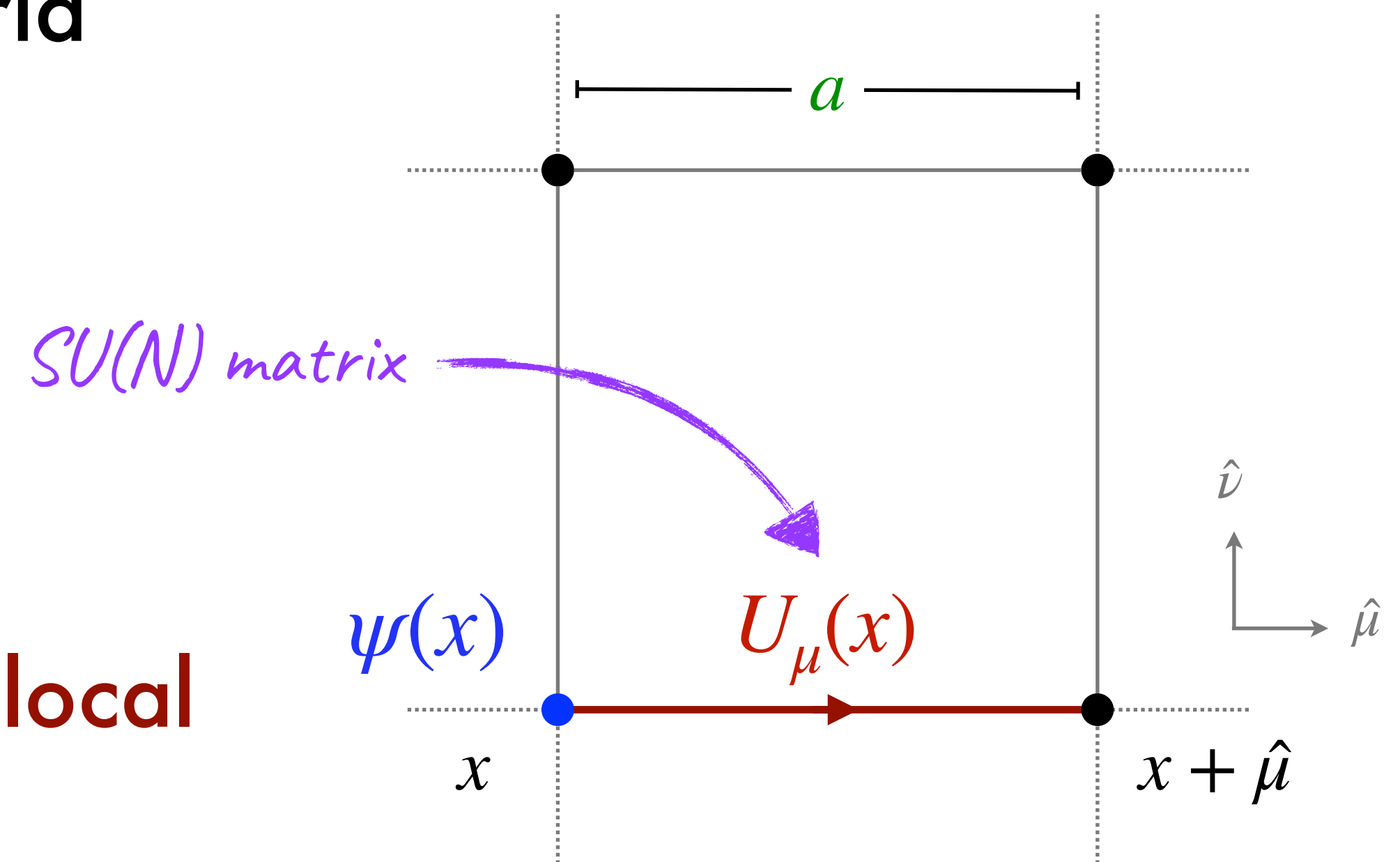
Replace continuum spacetime with **discrete grid**

- Finite spacing a between lattice points
- Matter fields $\psi(x)$ live on sites
- Gauge fields $U_\mu(x)$ live on links between sites

Gauge symmetry: physical invariance under **local transformations** by symmetry groups

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

- Physics encoded in energy functional $S[U, \psi]$ "action"



Ensemble Generation in LFT

Action defines Boltzmann **probability weights** $p(\phi) \propto e^{-S[\phi]}$

→ amenable to Monte Carlo simulation

Physical observables computed as path integral expectations

$$\langle \mathcal{O} \rangle = \frac{1}{\mathcal{Z}} \int [\mathcal{D}\phi] \mathcal{O}(\phi) e^{-S[\phi]} \approx \frac{1}{n} \sum_{i=1}^n \mathcal{O}(\phi_i)$$

integral over all possible configurations

empirical mean

⚠ Good estimates require **many correctly distributed** samples $\phi_i \sim p(\phi)$

Classical Approach: MCMC

Metropolis et al, *J. Chem. Phys.* **21** (1953)

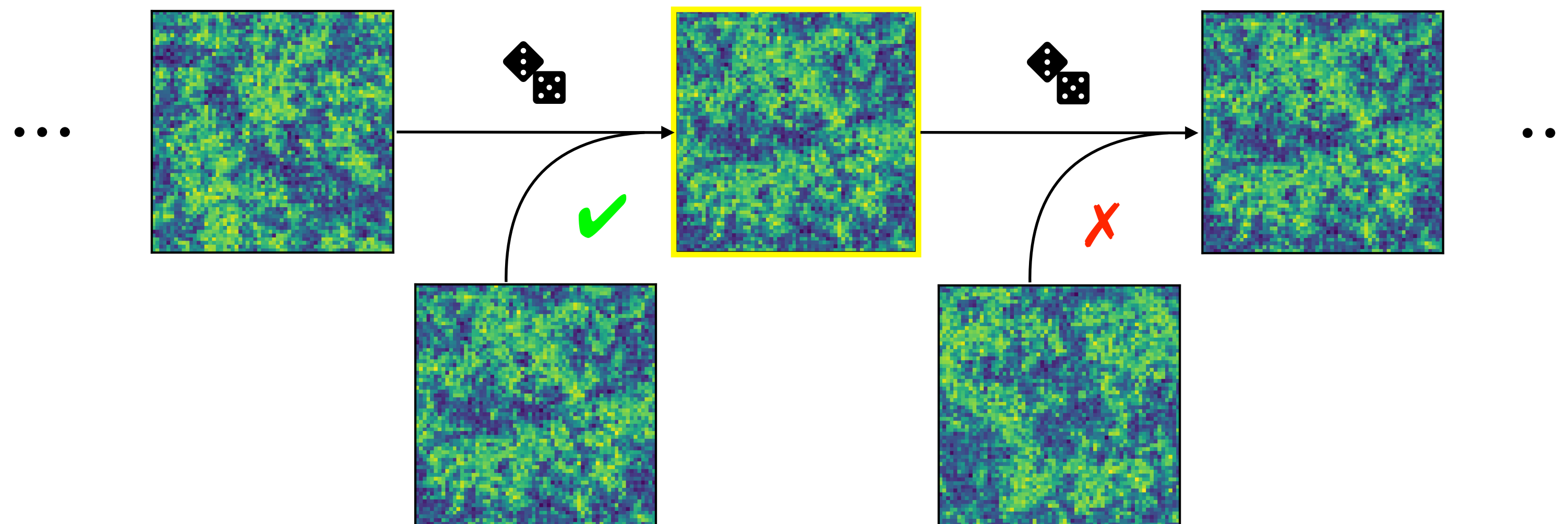
Duane et al, *Phys. Lett. B* **195** (1987)

Ergodic sampling with Markov chain Monte Carlo (MCMC)

- e.g. *Metropolis*, *Hybrid Monte Carlo (HMC)*

Rejections introduce autocorrelations that diverge with $a \rightarrow 0$

⚠ curse of dimensionality inhibits phase space exploration



Transport in Lattice Field Theory

Classical Algorithms: MCMC

- ⚠ Critical slowing down
- ⚠ Topological freezing

Theoretical Approach: trivializing maps

Lüscher, *Commun. Math. Phys.* **293** (2009)

Modern methods: deep learning

- Flow-based (continuous and discrete)
- **Diffusion-based** ← *This talk*

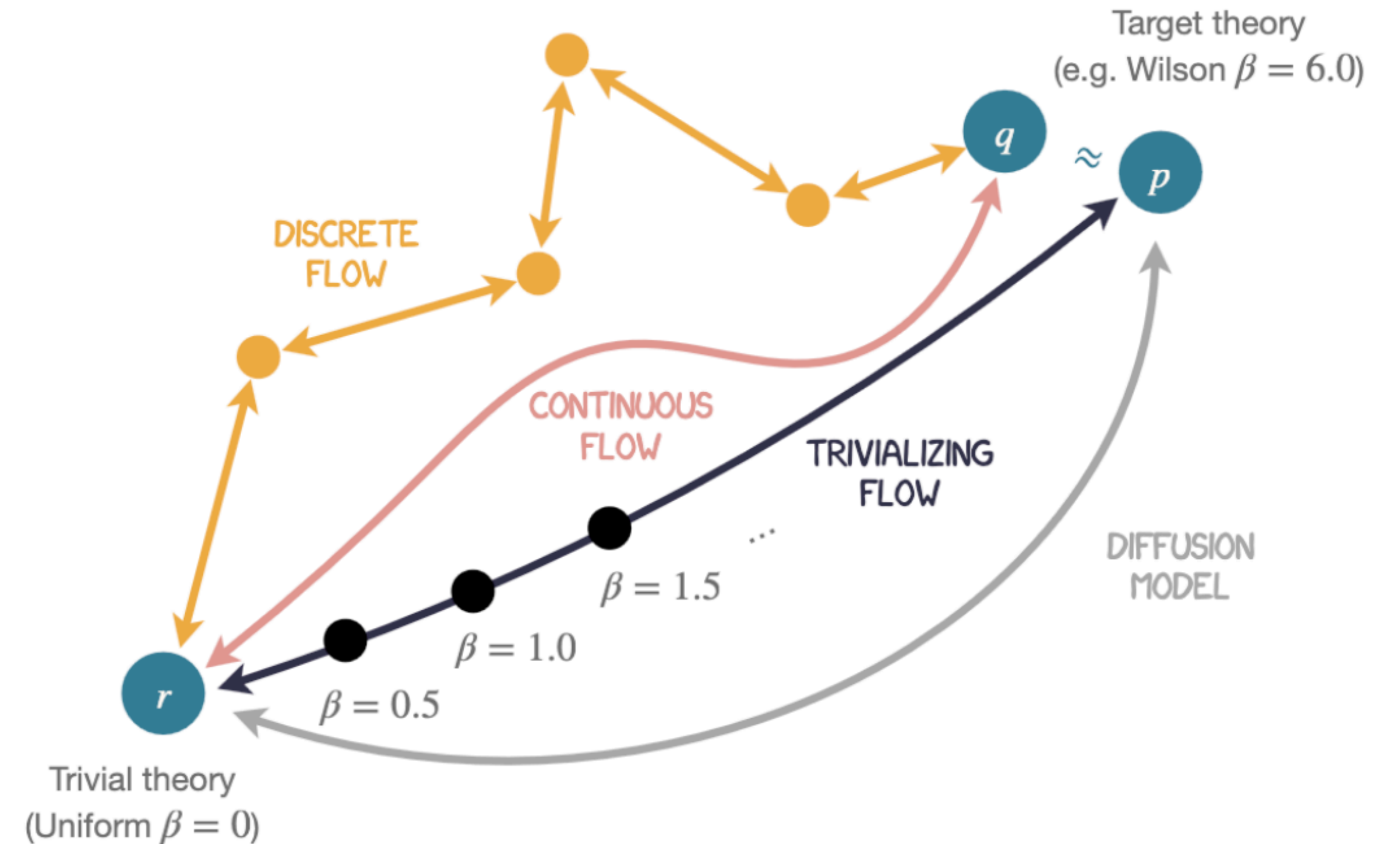
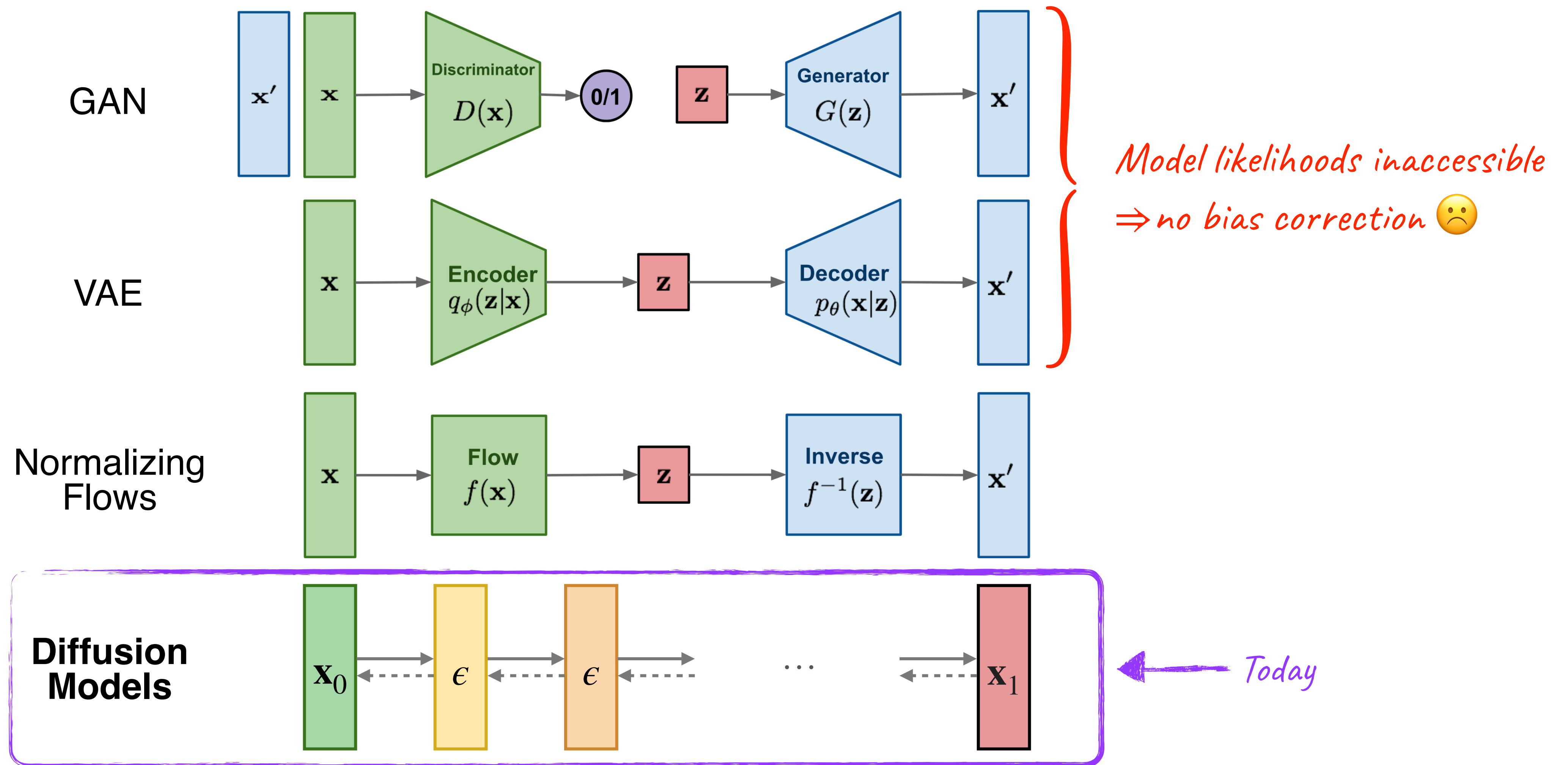


Image Credit: G. Kanwar

Taxonomy of Generative Models



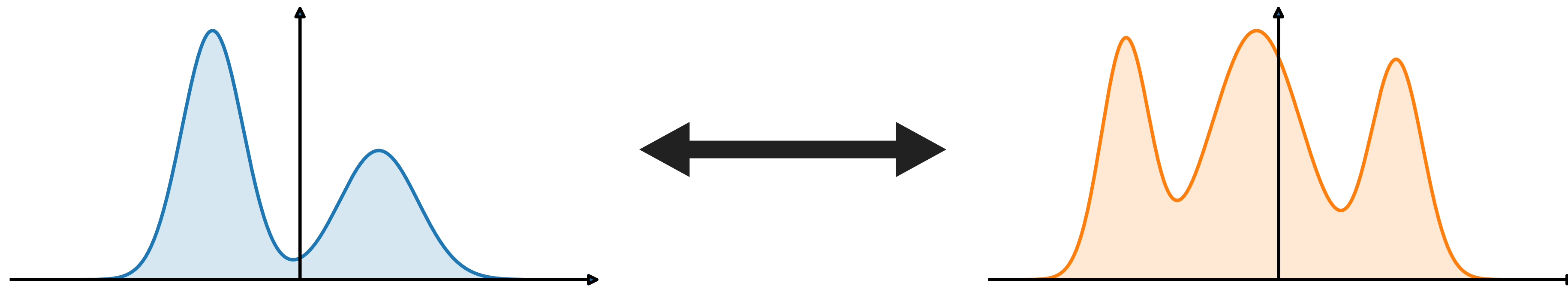
Adapted from: P. Lippe

Measure Transport

Dynamical Measure Transport

Density evolves *continuously* according to **Fokker-Planck Equation**:

$$\partial_t p_t(\phi) = -\nabla \cdot (p_t v_t)(\phi) + \frac{1}{2} g_t^2 \nabla^2 p_t(\phi)$$



Corresponding **SDE** gives motion of samples $\phi_t \sim p_t(\phi)$:

$$d\phi_t = \underbrace{v_t(\phi_t)dt}_{\text{advection}} + \underbrace{g_t dW_t}_{\text{diffusion}}$$

Connection to Generative AI Paradigms

Different **SDE** \leftrightarrow different **FPE**, thus different **generative models**

$$\partial_t p_t(\phi) = \underbrace{-\nabla \cdot (p_t v_t)(\phi)}_{\text{pure drift}} + \underbrace{\frac{1}{2} g_t^2 \Delta p_t(\phi)}_{\text{pure noise}}$$

Deterministic Transport

$$\partial_t p_t = -\nabla \cdot (p_t v_t)$$

continuity equation

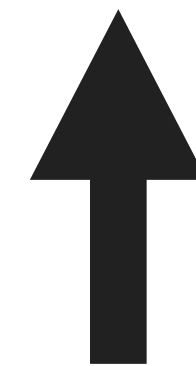
\rightarrow normalizing flows

Random Evolution

$$\partial_t p_t = \frac{1}{2} g_t^2 \Delta p_t$$

heat equation

\rightarrow diffusion models



How to Sample in Continuous Time?

Theorem: Every forward process $d\phi_t = v_t(\phi_t)dt + g_t dW_t$ has an associated family of **reverse processes**, given by

$$d\phi_t = \left[v_t(\phi_t) - \frac{g_t^2 + \tilde{g}_t^2}{2} \nabla \log p_t(\phi_t) \right] dt + \tilde{g}_t dW_t$$

score function

Integrate backwards to obtain **new samples** ϕ

- control reverse process noise with \tilde{g}_t
- ⚠ **Score function** $\nabla \log p_t(\phi_t)$ generally **unknown**

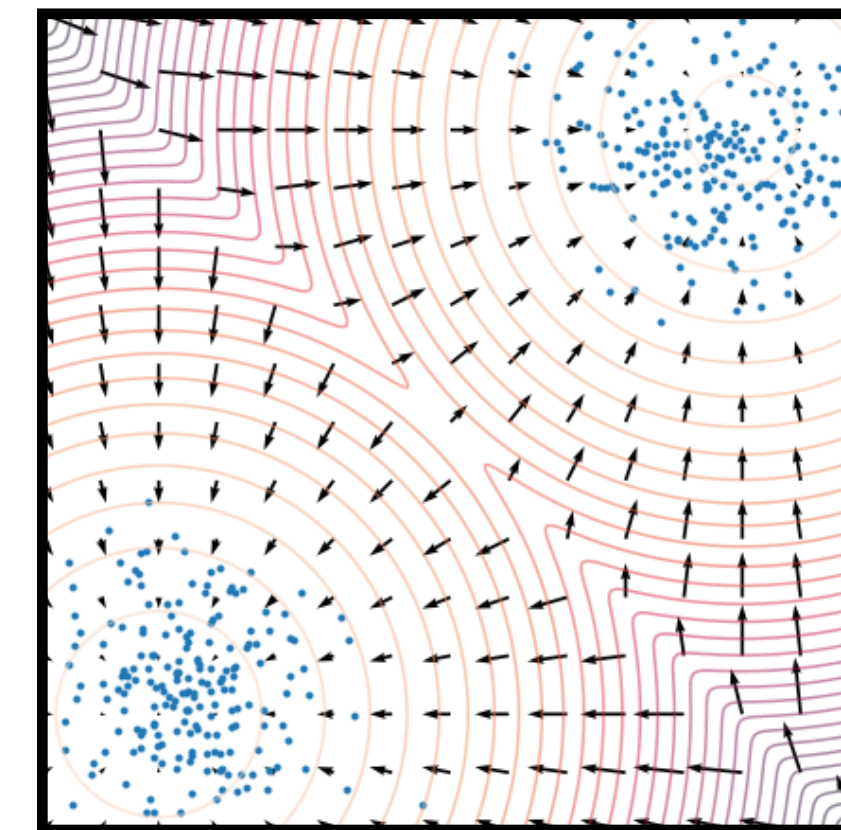


Image Credit: Song and Ermon, *NeurIPS* **33** (2019)

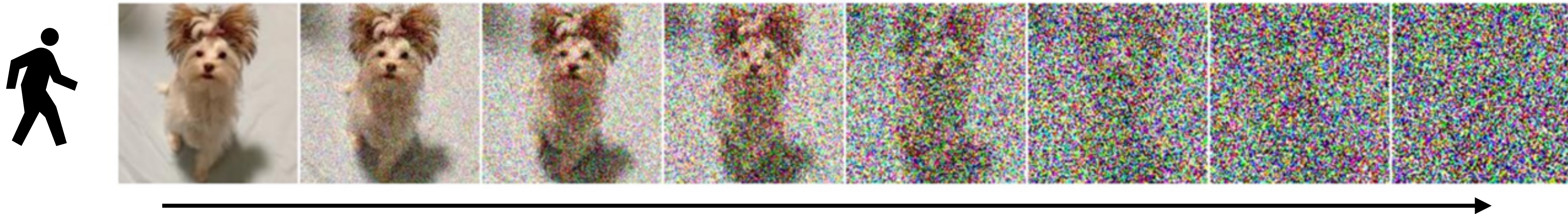
Generative Modeling

Diffusion-Based Sampling

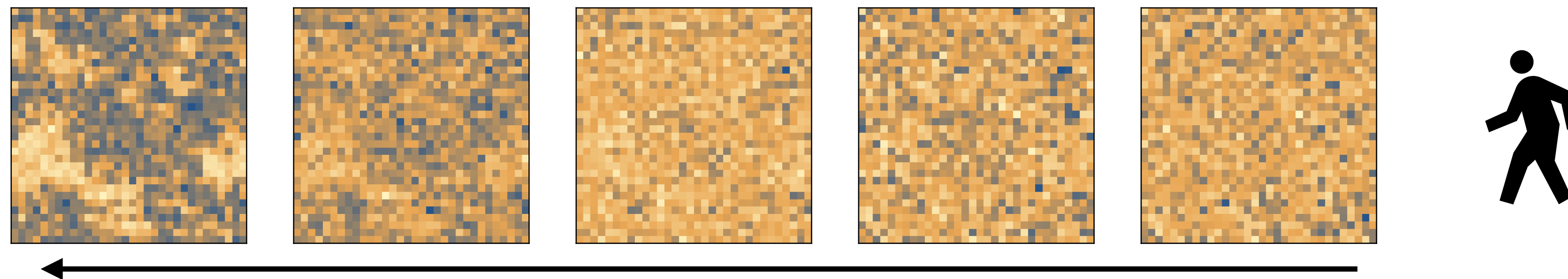
Idea: learn to reverse forward process

Diffusion process: data \rightarrow noise

Image credit: Song et al, *ICLR* (2021)



Denoising process: noise \rightarrow data



Requires the score function $\mathbf{b}_t(\phi) \propto \nabla \log p_t(\phi)$

Learning Score Functions

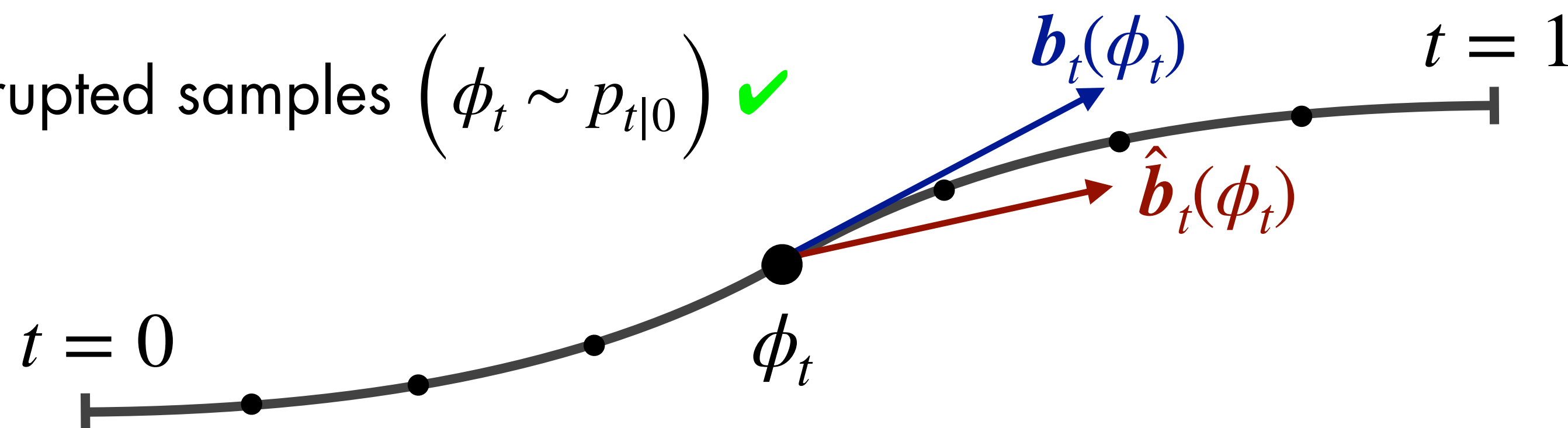
Train **neural net** to approximate b_t through *score matching*

$$\mathcal{L} [\hat{b}] = \int_0^1 \mathbb{E}_{\phi_0 \sim p_0} \mathbb{E}_{\phi_t \sim p_{t|0}} \left[\left\| \hat{b}_t(\phi_t) - \nabla \log p_{t|0}(\phi_t | \phi_0) \right\|^2 \right] dt$$

Computing \mathcal{L} requires sampling

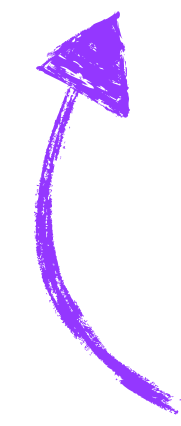
- **Target:** training data ($\phi_0 \sim p_0$) ✓

- **Transition:** corrupted samples ($\phi_t \sim p_{t|0}$) ✓



Simulation-Free Training

Heat kernel solves the diffusion equation for any **initial data**:



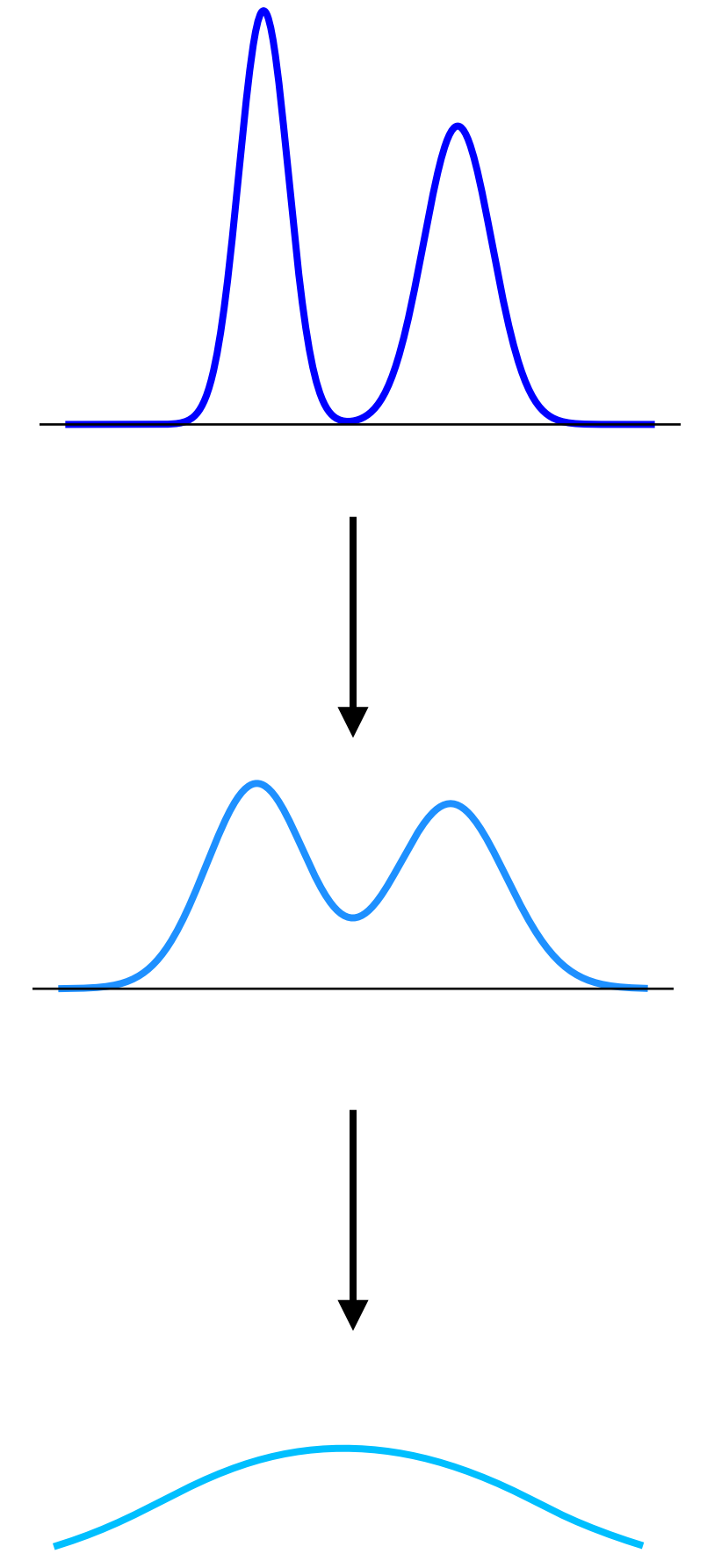
Green's function for Δ

$$p_t(\phi_t) = \int_{\mathbb{R}^n} K_t(\phi_t - \phi_0) p_0(\phi_0) d\phi_0$$

Simulate **Solve** forward SDE in a *single step* \rightarrow fast training

$$\phi_t = \phi_0 + \sigma_t \eta \quad \text{"one-shot" solution}$$

Sampling $K_t(\cdot)$ is trivial on \mathbb{R}^n ... just **Gaussian**: $\eta \sim \mathcal{N}(0, \mathbb{I})$



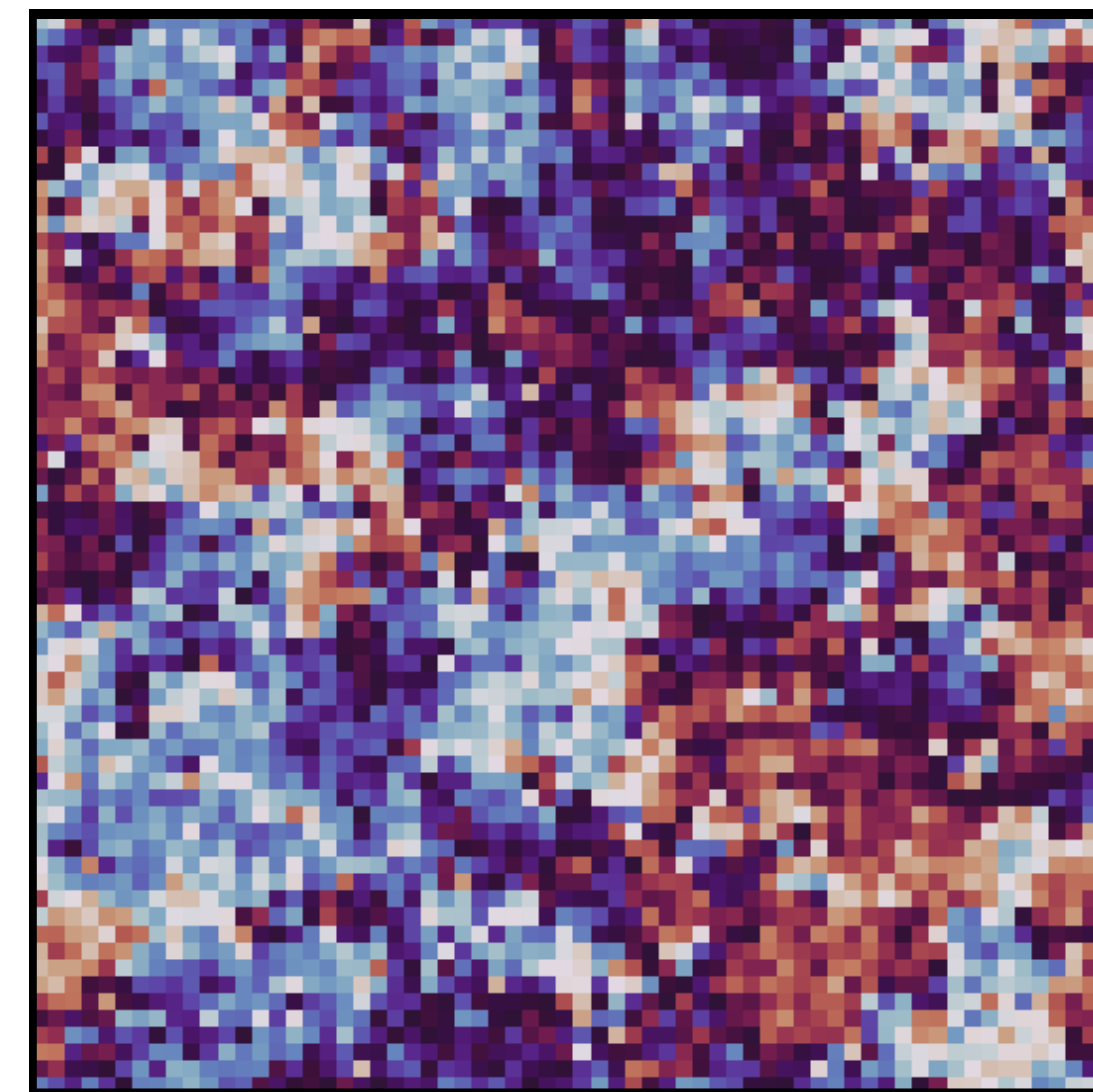
Exactness in Generative AI

LFT calculations must be **exact**... but machine learning models are **biased**

Image Credit: Saharia et al, *NeurIPS* **36** (2022)



$$S[\phi] = ?$$



$$S[\phi] = -7404.9$$

! Correcting bias requires target likelihood $\log p = -S$ and **model likelihood** $\log q$

Symmetries of LFT

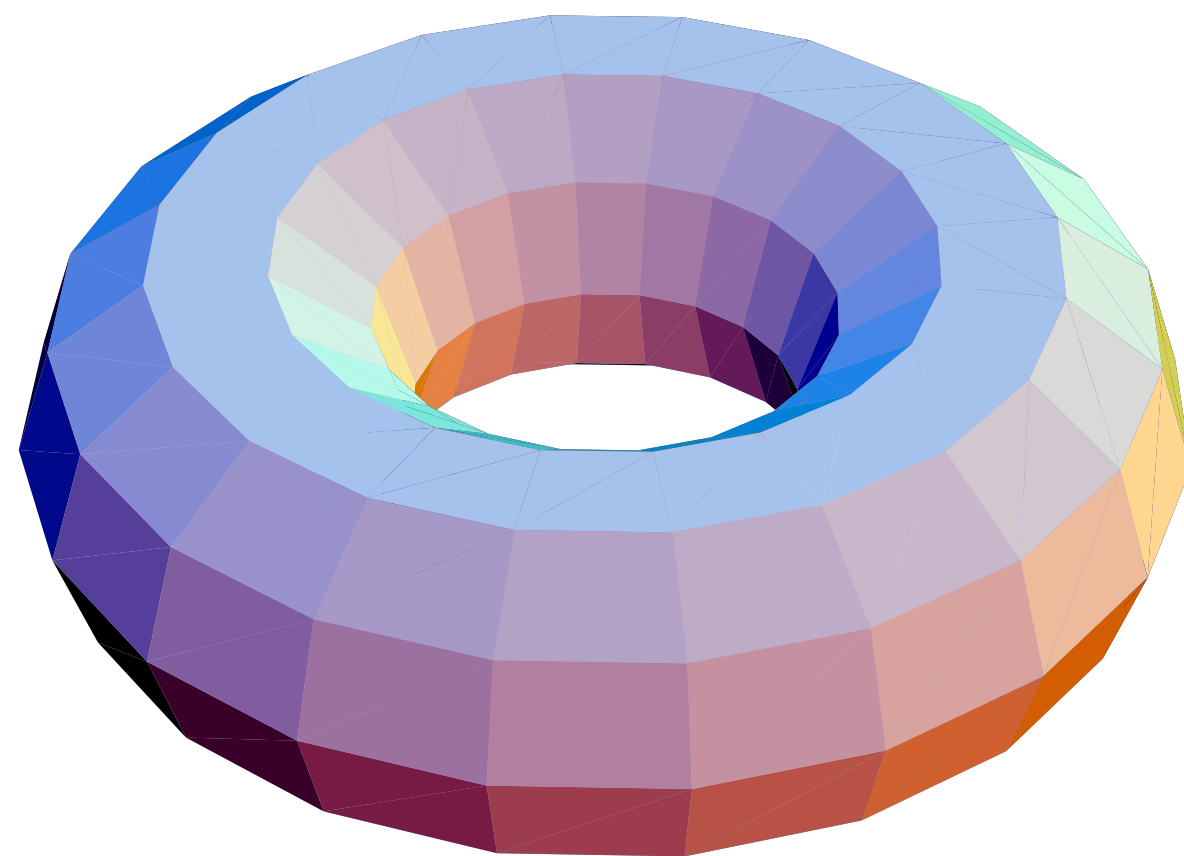
LFT probabilities often *invariant* under **group symmetries**: $p(\phi) \equiv p(G \cdot \phi)$

Internal Symmetries

- Global \mathbb{Z}_2 reflections $\phi \rightarrow -\phi$
- Local U(1) rotations $U_\mu(x) \rightarrow e^{i\alpha(x)}U_\mu(x)$

Spacetime Symmetries

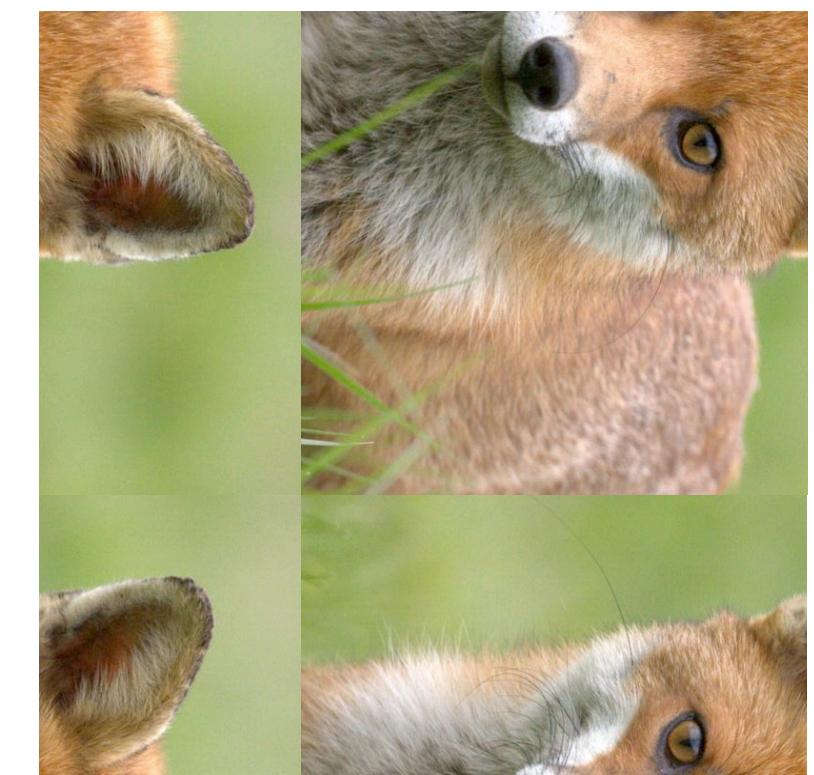
- Discrete translations \rightarrow convolutions ✓
- Periodic bounds \rightarrow circular padding ✓



Toroidal topology



=



Equivariance of the Score Function

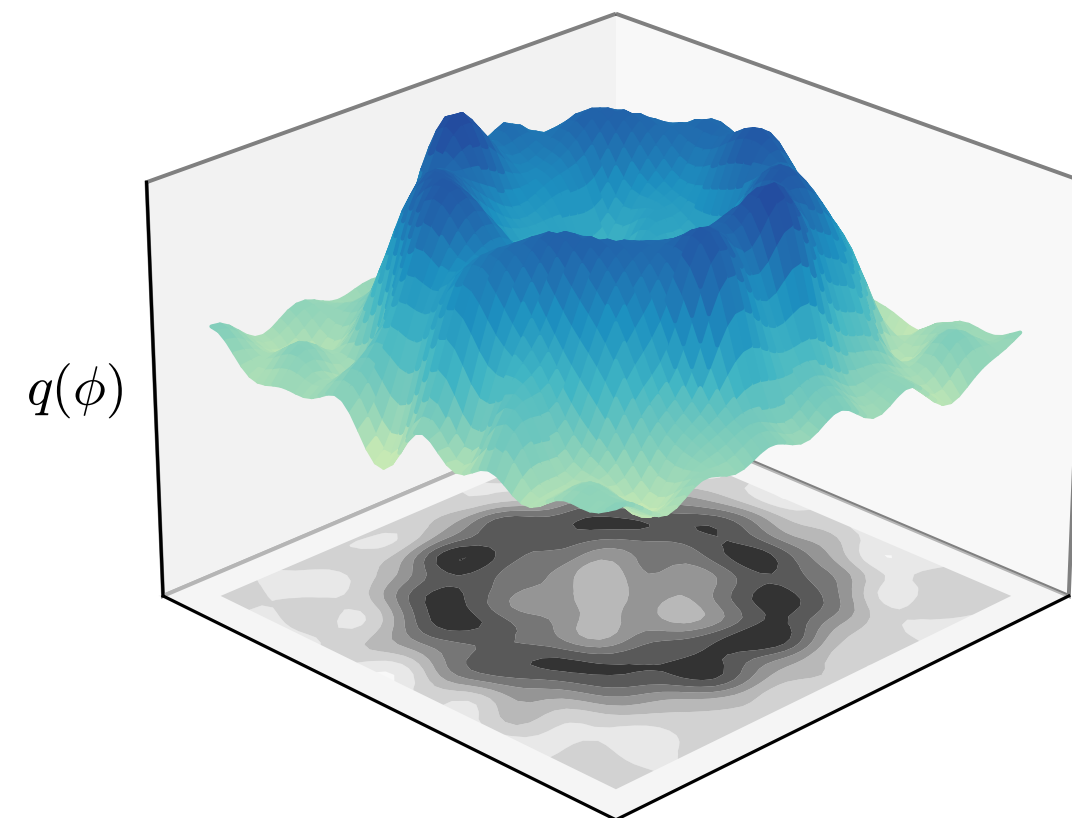
Theorem. If $\phi' = g \cdot \phi$ and $S[\phi'] \equiv S[\phi]$, then *Proof: Chain rule for grad(S)*

*invariant action
⇒ equivariant force*

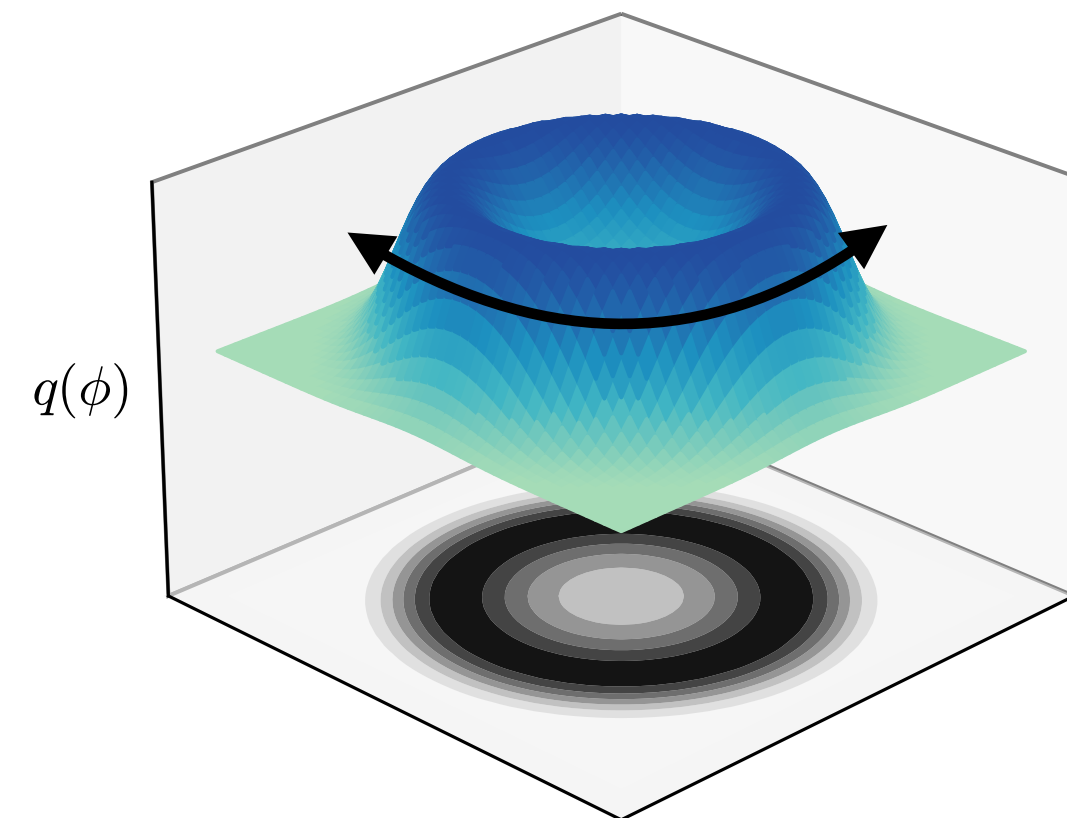
$$\mathbf{b}_t(\phi) \mapsto \left(\frac{\partial \phi'}{\partial \phi} \right)^{-T} \mathbf{b}_t(\phi)$$

invariant prior + **equivariant** map = **symmetric** model

Learned Symmetry



Exact Symmetry



Adapted from:
G. Kanwar

Application to Scalar Field Theory

OV, J. Kominani, A. El-Khadra, and M. Marinkovic, 2510.26081

Scalar ϕ^4 Theory in 2D

the "hello world" of QFT

Replace continuum action $S[\phi] = \int d^2x [-\phi(x)\partial_x^2\phi(x) + m^2\phi(x)^2 + \lambda\phi(x)^4]$ with

$$S[\phi] = \sum_x \left[- \sum_{\mu} \phi(x)\phi(x + \hat{\mu}) + (2 + m^2) \phi(x)^2 + \lambda\phi(x)^4 \right]$$

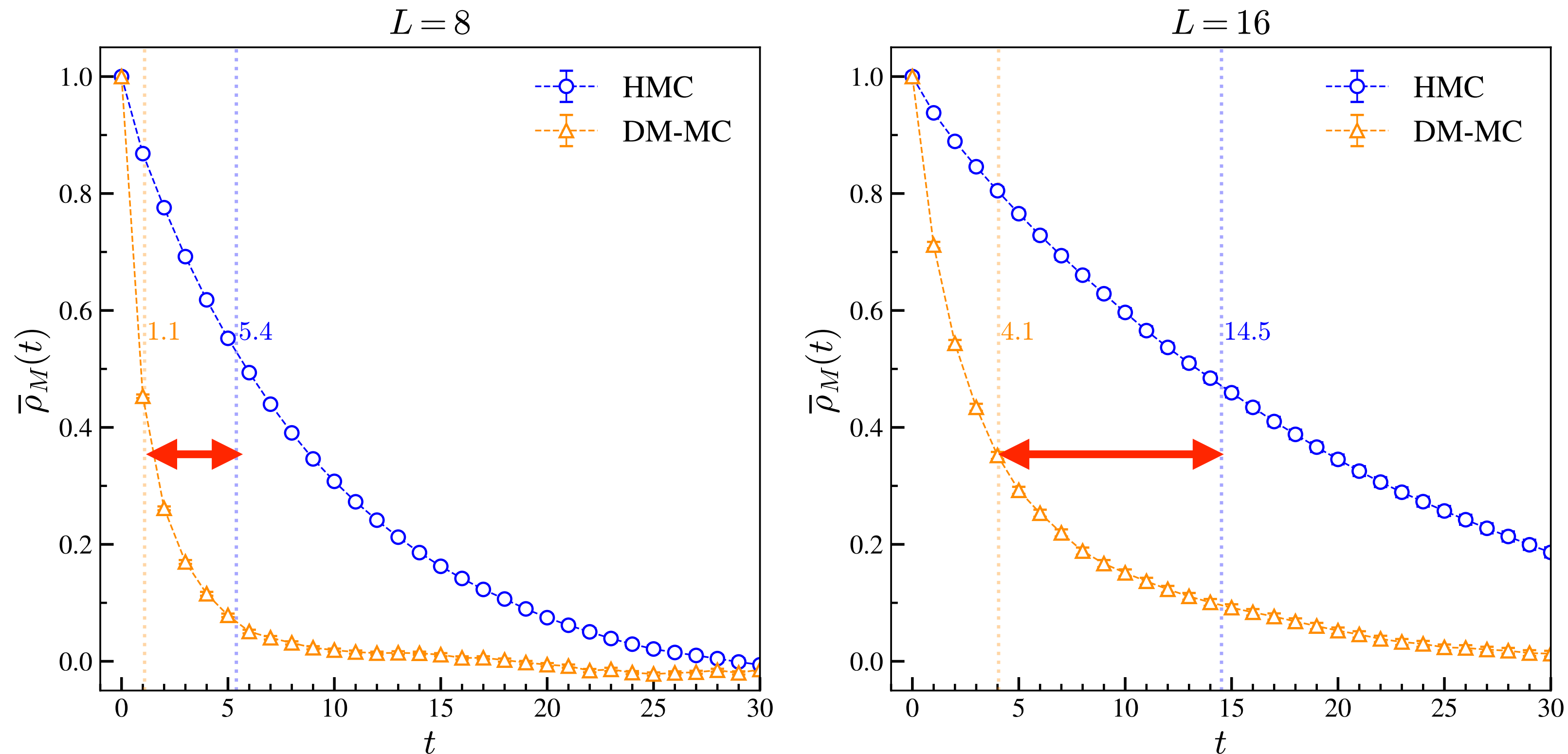
Symmetric under **global \mathbb{Z}_2 reflections**:

$$S[\phi] \equiv S[-\phi] \Rightarrow b_t(-\phi) = -b_t(\phi)$$

\therefore use an antisymmetric score network

A Solution to Critical Slowing Down?

Normalized **autocorrelations** for resampling from diffusion model vs. HMC



3-5x reduction in autocorrelation!

How Much Does Symmetry Matter?

Sym.	No Regularization				Force Reg. ($c_0 = 0.1$)			
	NFE (\downarrow)	ESS (\uparrow)	AR (\uparrow)	KL (\downarrow)	NFE (\downarrow)	ESS (\uparrow)	AR (\uparrow)	KL (\downarrow)
<i>Euler Integrator ($N_{\text{step}} = 500$)</i>								
$\{e\}$	500	0.103(15)	0.284(12)	-59.58(05)	500	0.282(23)	0.434(16)	-60.22(16)
\mathbb{Z}_2	1000	0.201(22)	0.382(13)	-60.07(04)	1000	0.558(26)	0.597(12)	-60.57(02)
\mathbb{T}	500	0.280(21)	0.462(14)	-60.33(03)	500	0.545(20)	0.602(10)	-60.59(01)
$\mathbb{Z}_2 \times \mathbb{T}$	1000	0.527(44)	0.593(22)	-60.60(03)	1000	0.672(21)	0.671(13)	-60.70(01)
<i>Runge-Kutta (RK4) Integrator ($N_{\text{step}} = 125$)</i>								
$\{e\}$	500	0.131(18)	0.304(15)	-59.85(10)	500	0.264(20)	0.430(14)	-60.77(14)
\mathbb{Z}_2	1000	0.209(18)	0.396(16)	-60.53(11)	1000	0.531(24)	0.573(12)	-61.16(16)
\mathbb{T}	500	0.289(28)	0.450(19)	-60.63(08)	500	0.522(33)	0.602(11)	-61.03(10)
$\mathbb{Z}_2 \times \mathbb{T}$	1000	0.579(34)	0.613(20)	-61.09(08)	1000	0.597(28)	0.653(08)	-61.24(18)

more symmetry = better quality

Application to $U(1)$ Gauge Theory

OV, J. Kominani, A. El-Khadra, and M. Marinkovic, 2510.26081

Abelian Gauge Theory on the Lattice

Gauge-invariant action built from *plaquettes*

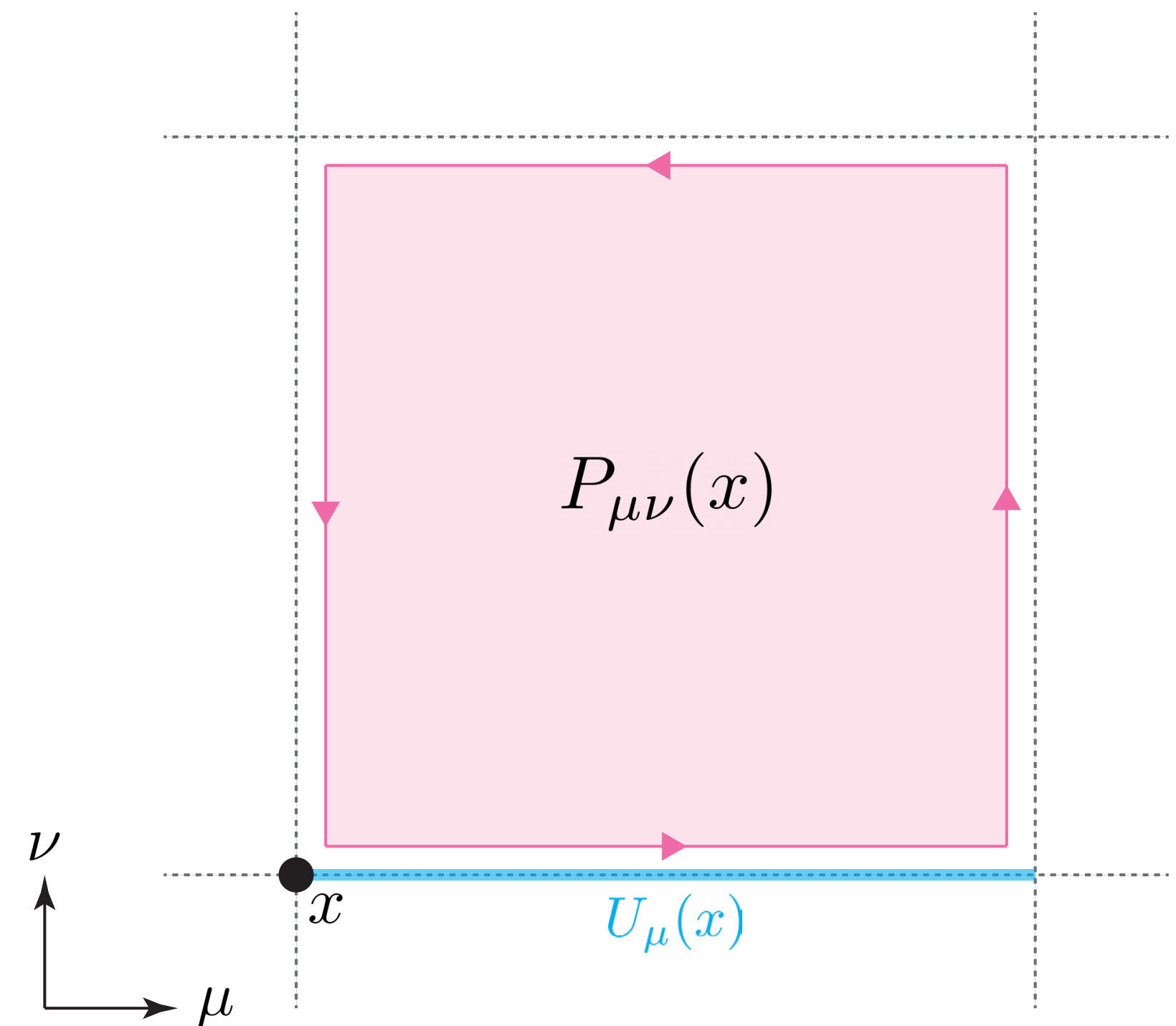
$$S[U] = -\beta \sum_x \sum_{\mu < \nu} \text{Re} P_{\mu\nu}(x)$$

Links are complex numbers $U_\mu(x) \in \mathbb{C}$

- Work with their phases
- $A_\mu(x) = \arg U_\mu(x)$, $\phi_{\mu\nu}(x) = \arg P_{\mu\nu}(x)$

$$S[A] = -\beta \sum_x \sum_{\mu < \nu} \cos \phi_{\mu\nu}(x)$$

$$P_{\mu\nu}(x) = U_\mu(x)U_\nu(x + \hat{\mu})U_\mu^*(x + \hat{\nu})U_\nu^*(x)$$



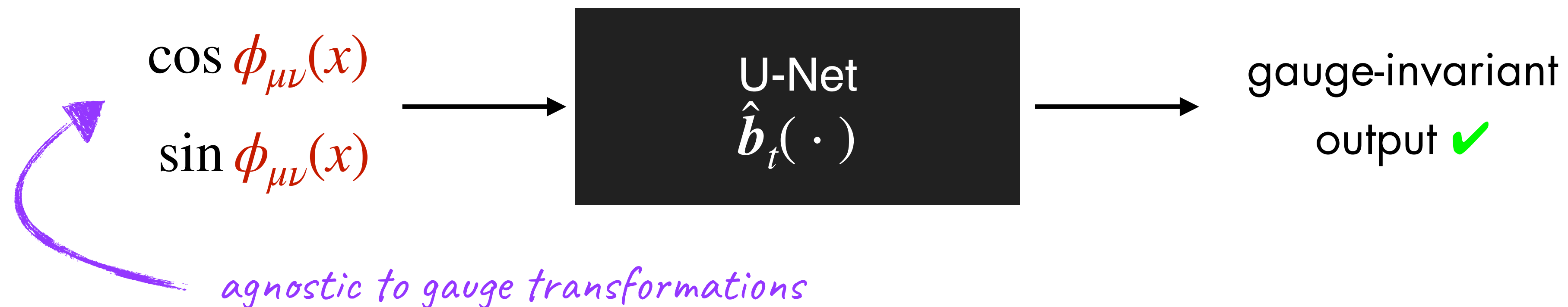
Gauge-Equivariant Score Network

⚠ The U(1) force (score) is gauge-**invariant*** 🤔

$$\nabla_{A_\mu} S[A] = \beta \sum_{\nu \neq \mu} \left[\sin \phi_{\mu\nu}(x) - \sin \phi_{\mu\nu}(x - \hat{\nu}) \right]$$

- Abelian group \Rightarrow simple adjoint rep \Rightarrow equivariance = invariance

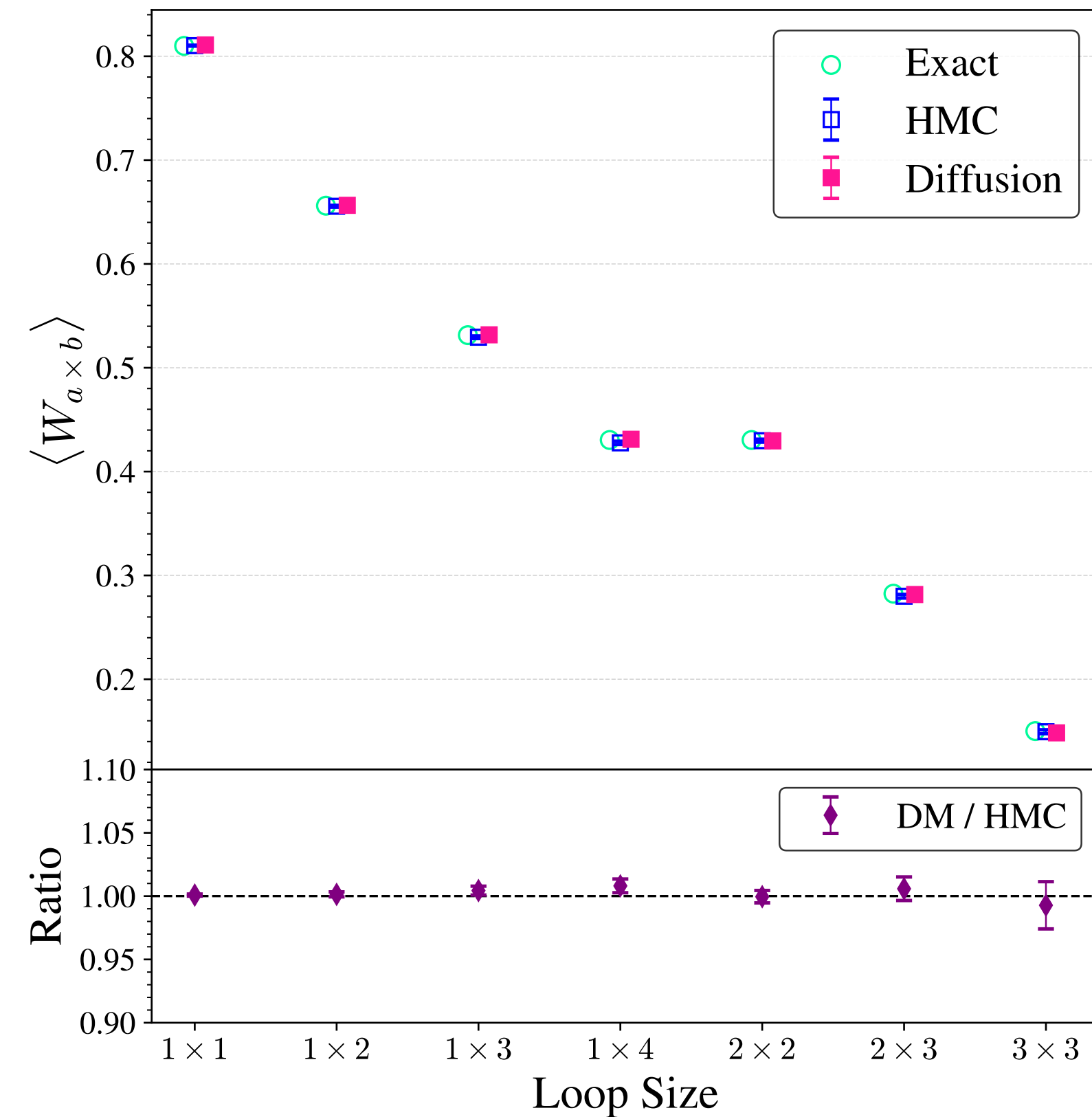
Idea: input **gauge-invariant data** to the score network



Observables

Bias-correction closes the gap in $\langle \mathcal{O} \rangle$

β	Ensemble	$\langle P \rangle$	χ_Q	ESS
1.0	HMC	0.4460(8)	0.0406(6)	—
	Diffusion	0.4450(8)	0.0406(6)	93.2%
2.0	HMC	0.6971(5)	0.0197(3)	—
	Diffusion	0.6967(6)	0.0194(3)	81.0%
3.0	HMC	0.8098(4)	0.0111(1)	—
	Diffusion	0.8085(4)	0.0110(1)	63.1%
4.0	HMC	0.8632(3)	0.0075(1)	—
	Diffusion	0.8626(3)	0.0075(1)	45.6%
5.0	HMC	0.8935(2)	0.0056(1)	—
	Diffusion	0.8923(2)	0.0057(1)	40.6%



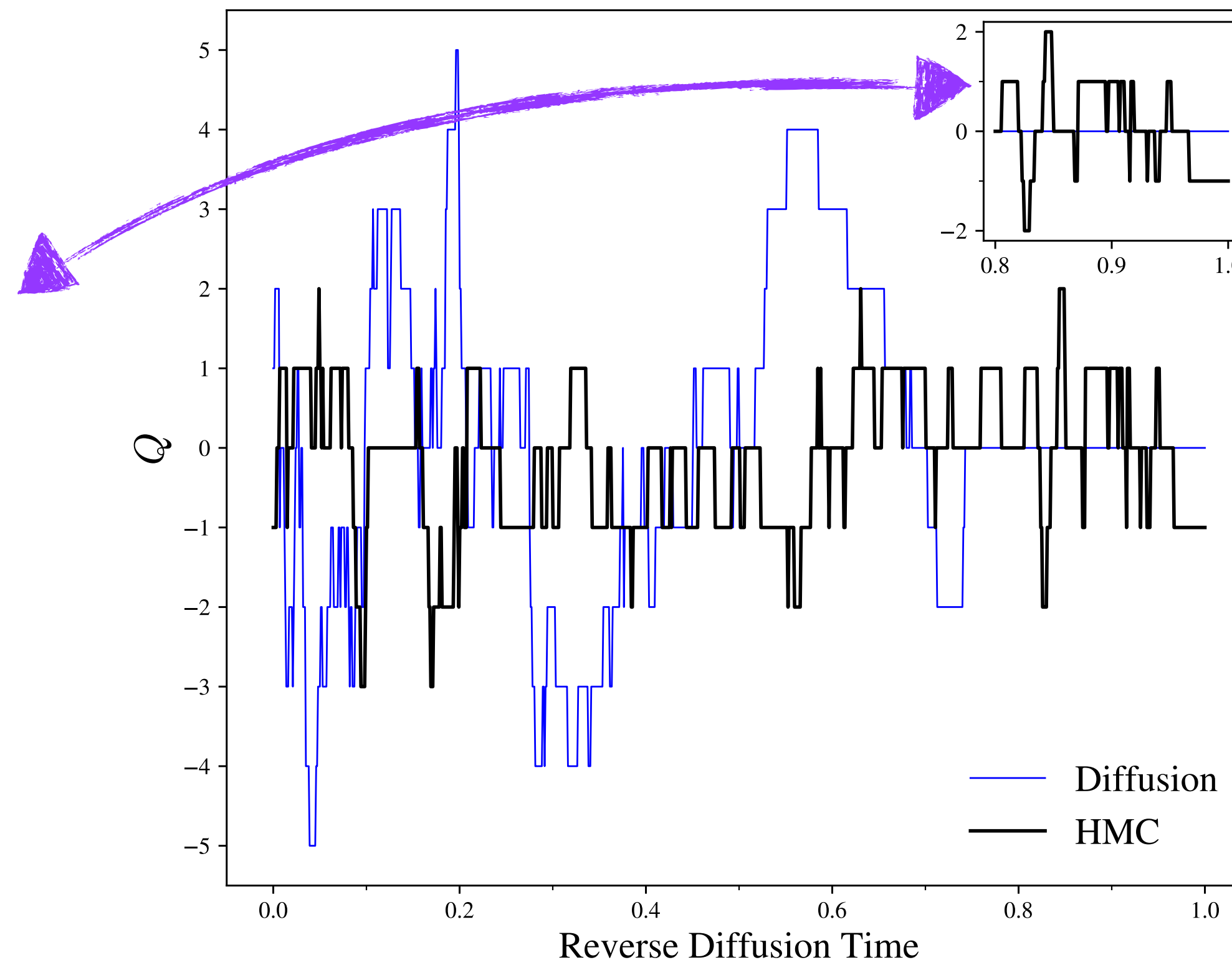
What About Topological Freezing?

Label U(1) configs by their **topological charge** $Q = \frac{1}{2\pi} \sum_x \arg P_{01}(x) \in \mathbb{Z}$

- Q is a **global quantity**, like **gender**
- high-level attributes decided earlier



Ho et al., *NeurIPS* **34** (2020)



less freezing!

Diffusion on $SU(N)$

G. Kanwar and OV, PoS **LATTICE2025** (2026)

Diffusion on $SU(N)$

⚠ curved manifold

⚠ non-Abelian Lie group $\rightarrow \exp(A)\exp(B) \neq \exp(A + B)$

\Rightarrow *a priori* training not simulation-free 😡

A **one-shot** solution to the $SU(N)$ heat equation $\partial_t p_t(U) = \Delta_U p_t(U)$?

$$p_t(U_t) = \int_{SU(N)} dU_0 K_t \left(U_t U_0^\dagger \right) p_0(U_0), \quad U_t = V U_0?$$

Key idea: Laplace-Beltrami operator Δ_U is **translation-invariant**

Heat Evolution on the Maximal Torus

Conjugation-invariance: $\Delta_U \equiv \Delta_{P^\dagger U P} \Rightarrow K_t(U) \equiv K_t(P^\dagger U P)$

Diagonalization on $SU(N)$ is equivalent to unitary conjugation

$$U = V^\dagger \Lambda V, \Lambda = \text{diag} (e^{i\theta_1}, \dots, e^{i\theta_N})$$

$\rightarrow K_t$ reduces to subgroup of **diagonal matrices** (maximal torus $\mathbb{T} \leq SU(N)$)

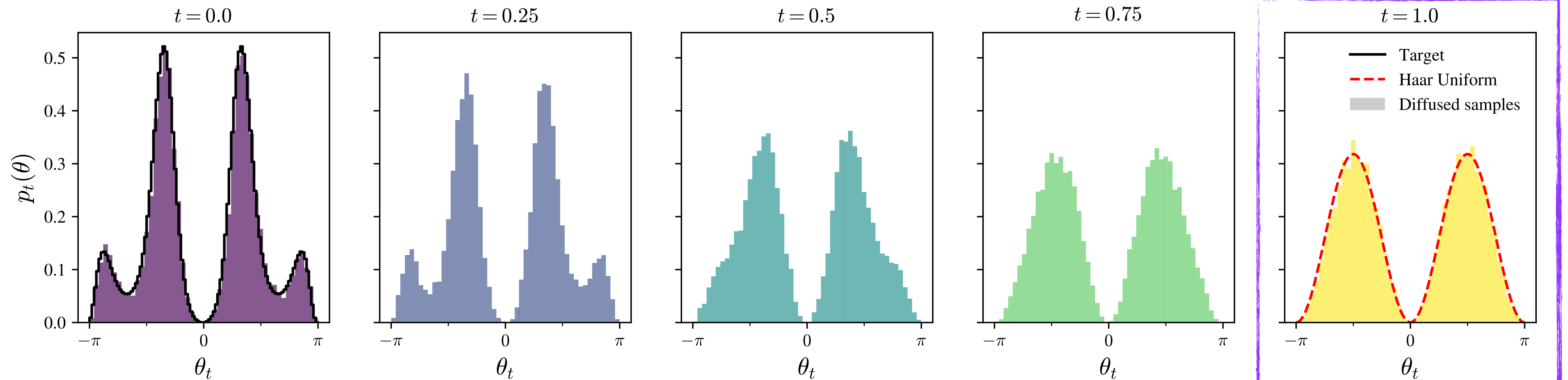
$$\begin{array}{ccc} \mathfrak{su}(N) & \xrightarrow{\text{diag}} & \mathfrak{t} \\ \exp \downarrow & & \downarrow \exp \\ SU(N) & \xrightarrow{\text{diag}} & \mathbb{T} \end{array}$$

Spectral Diffusion

Just need samples $\theta \sim K_t, P \sim \text{Haar} [U(N)]$

$$U_0 \rightarrow U_t = P^\dagger \text{diag} (e^{i\theta}) P U_0$$

Fully diffused to Haar uniform



Learning Toy Densities

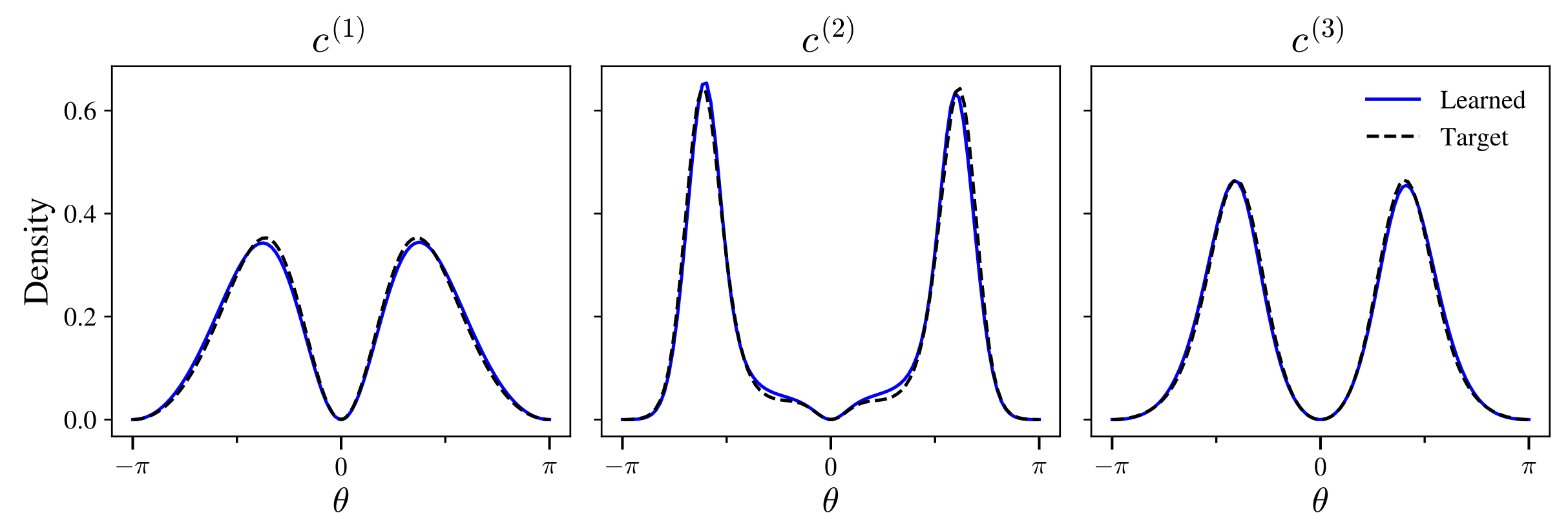
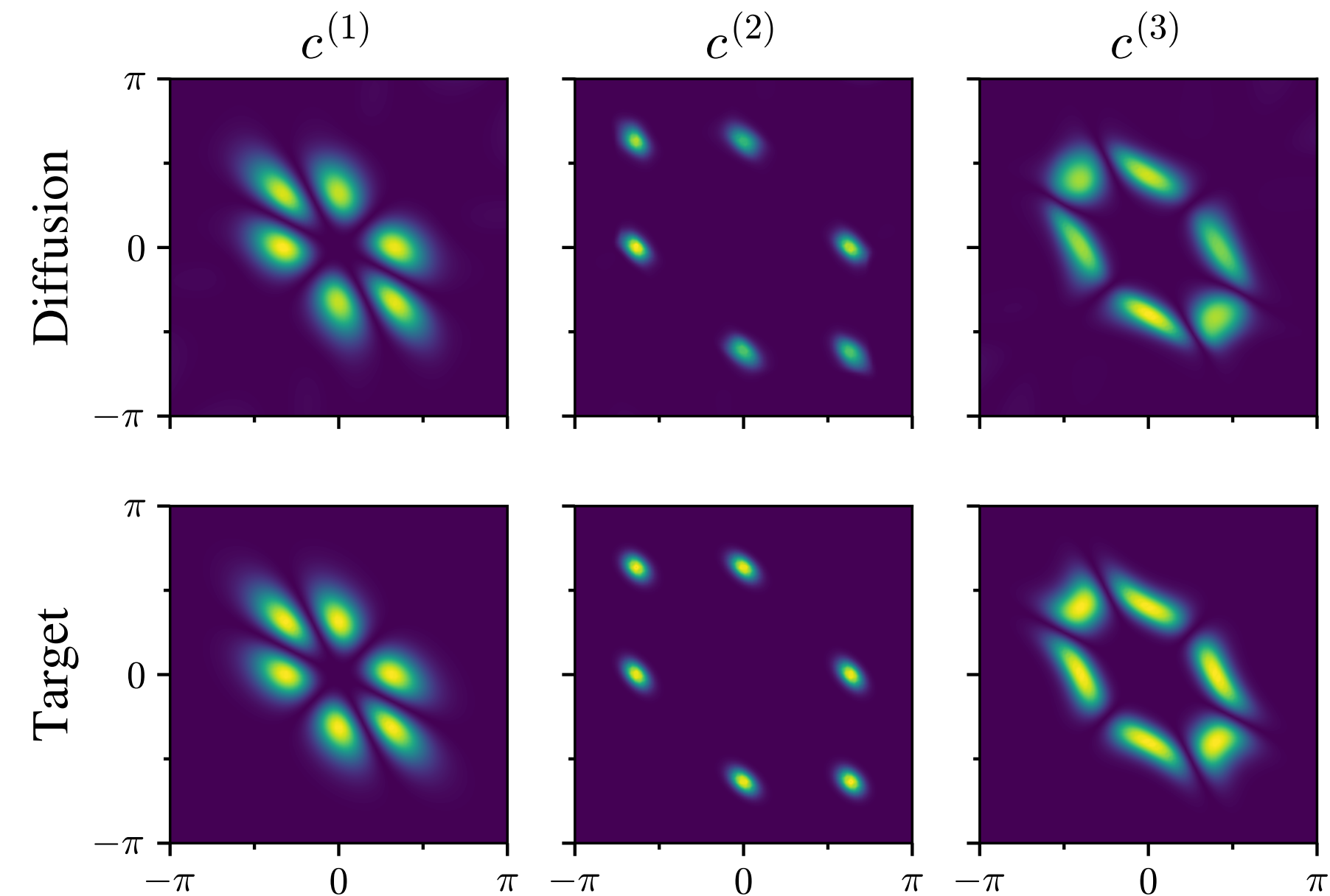
Sample from a family of toy densities

$$S_i(U) = -\frac{\beta}{N} \text{ReTr} \left[\sum_k c_k^{(i)} U^k \right]$$

Boyda, et al. *PRD* **103** (2021)

Multiple couplings β , coefficient sets

- obtain **ESS = 93% - 100%**



Ongoing Work

The Schwinger Model

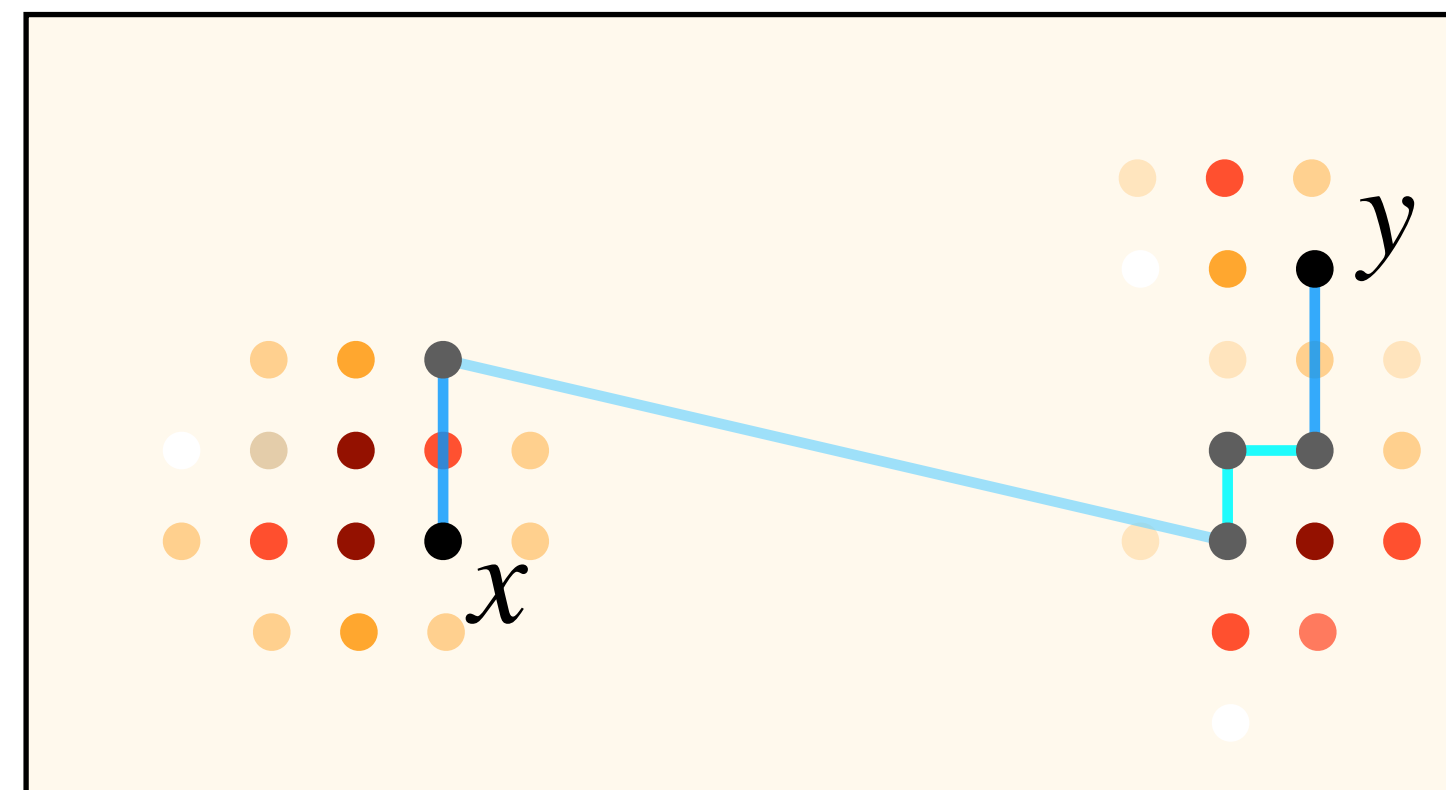
J. Schwinger, *Phys. Rev.* **128** (1962)

Strongly interacting model of QED in 2D

- Confinement ✓, chiral anomaly ✓, non-trivial topology ✓

Action defines a **joint** probability over bosons and fermions: $p(U, \bar{\psi}, \psi) \propto e^{-S[U, \bar{\psi}, \psi]}$

$$S[U, \bar{\psi}, \psi] = -\beta \sum_x \text{Re } P(x) + \sum_{x,y} \bar{\psi}_\alpha(x) D[U](x, y)^{\alpha\beta} \psi_\beta(y)$$



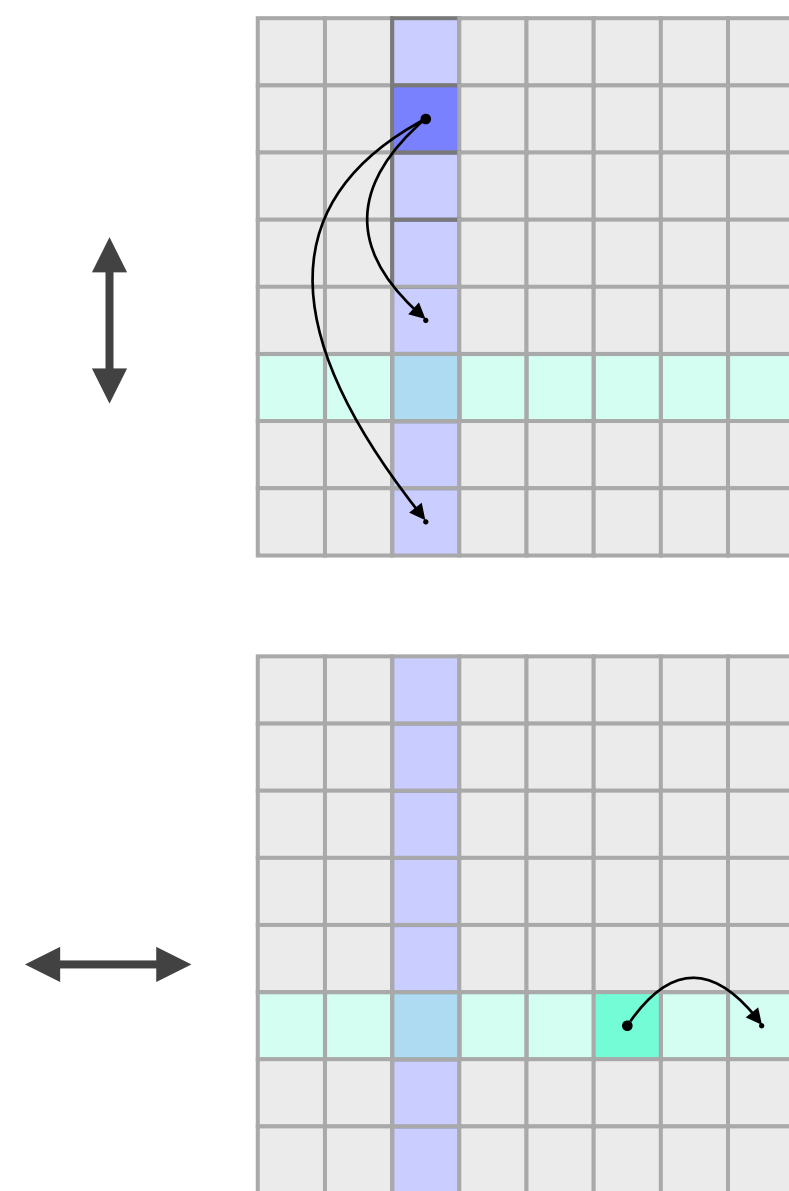
large sparse matrix

Fermions Demand Better Architectures

How to deal with **long-range correlations**?

Axial Attention

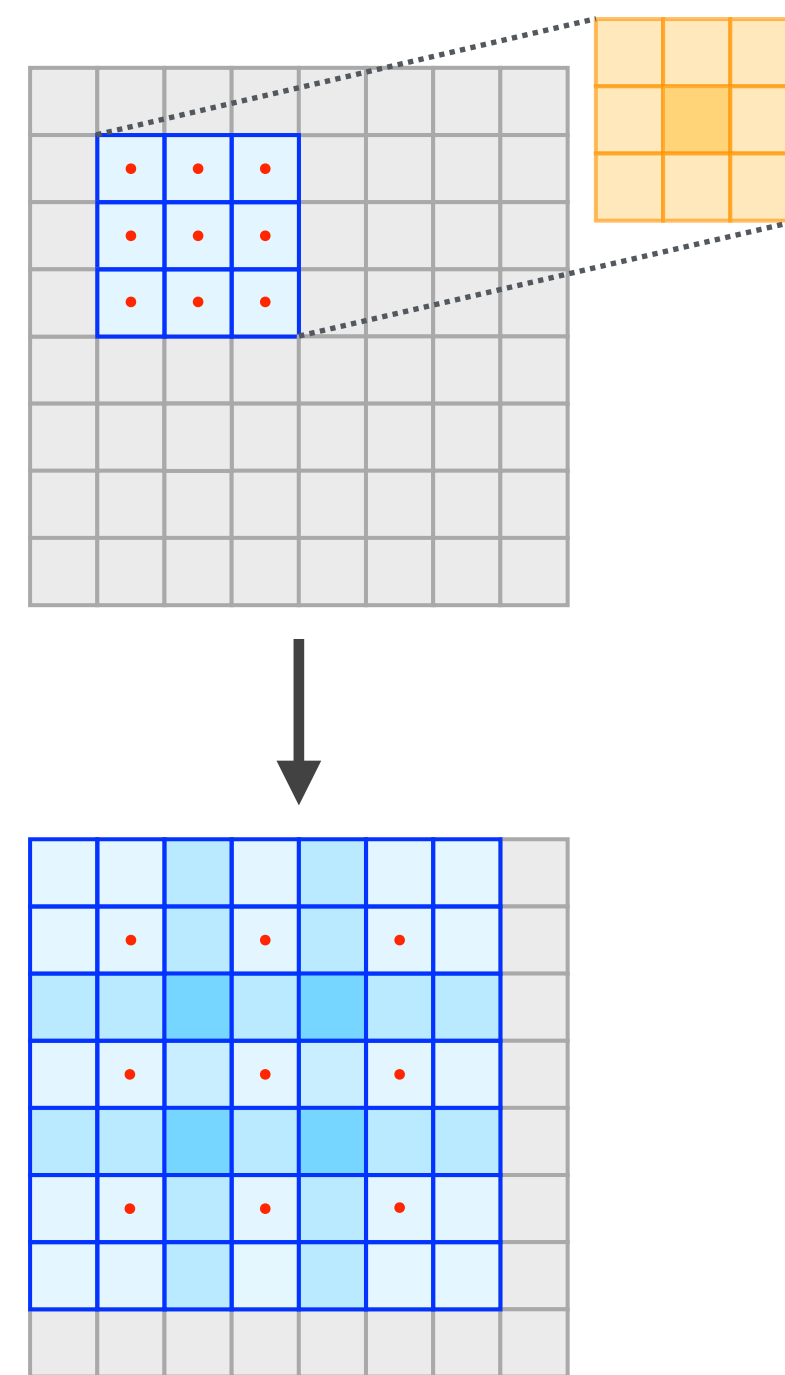
Ho et al., 1912.12180



Dilated Convolutions

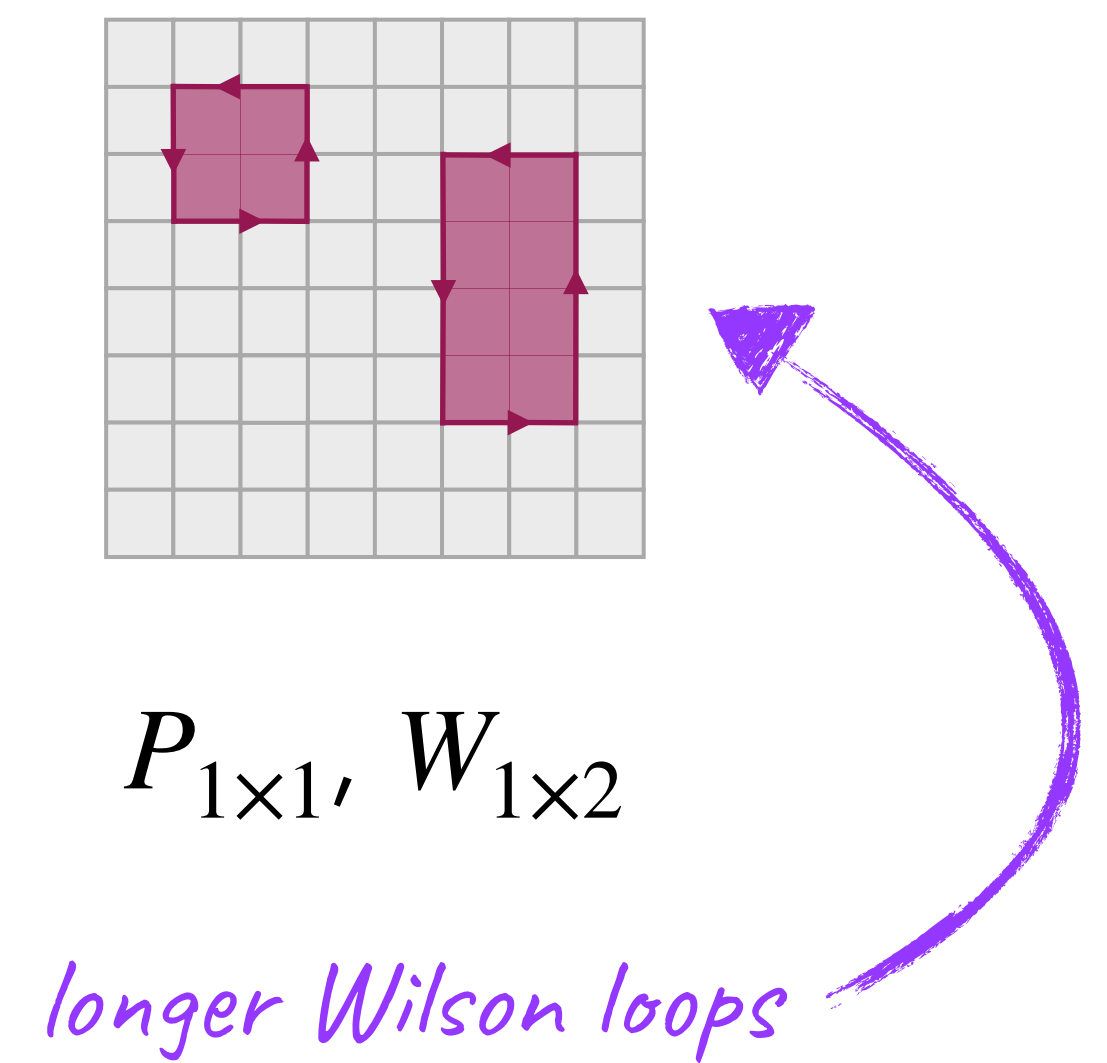
Yu and Koltun, 1511.07122

*expanded
receptive field*



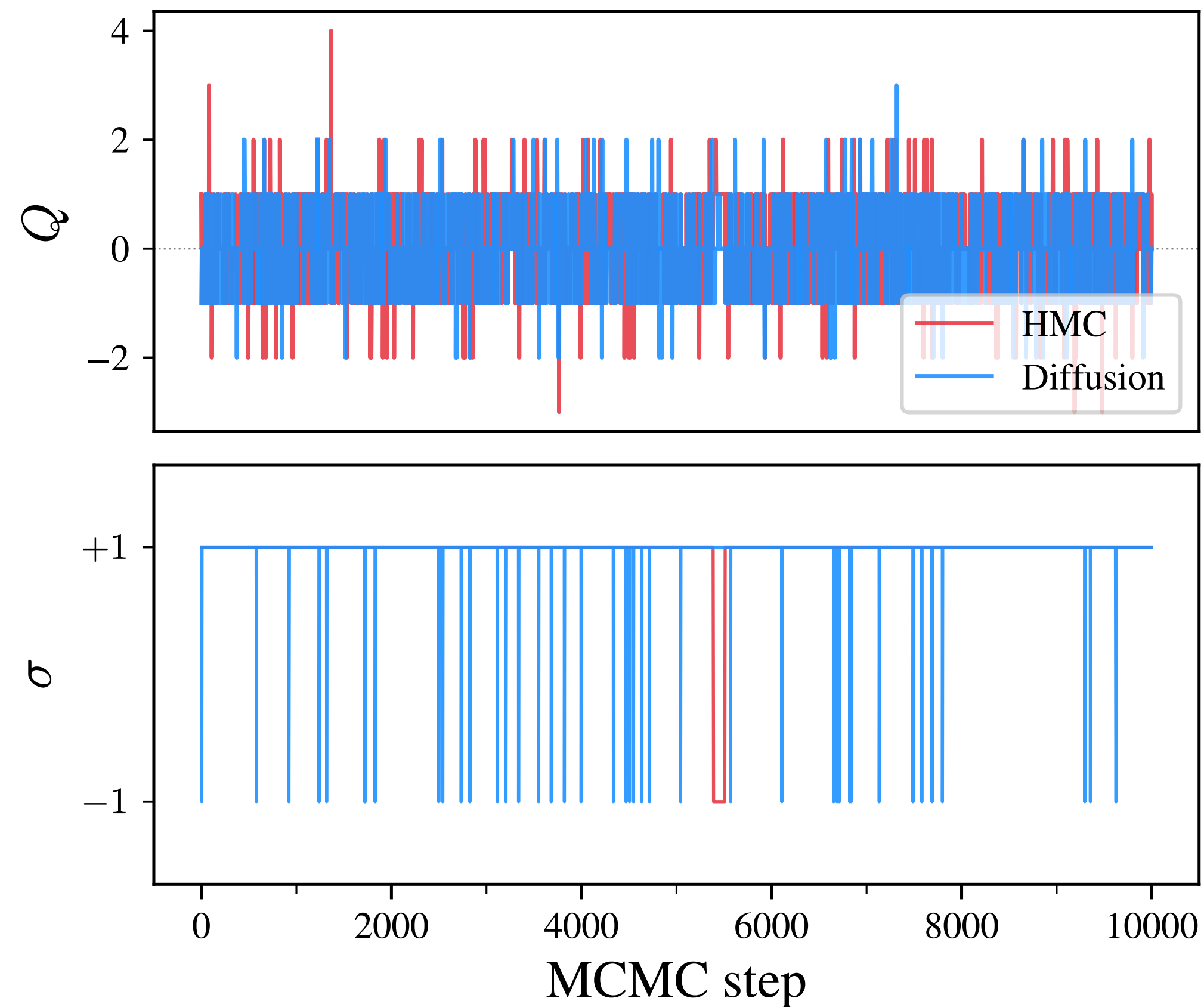
Larger Inputs

Albergo et al., PRD **106** (2022)

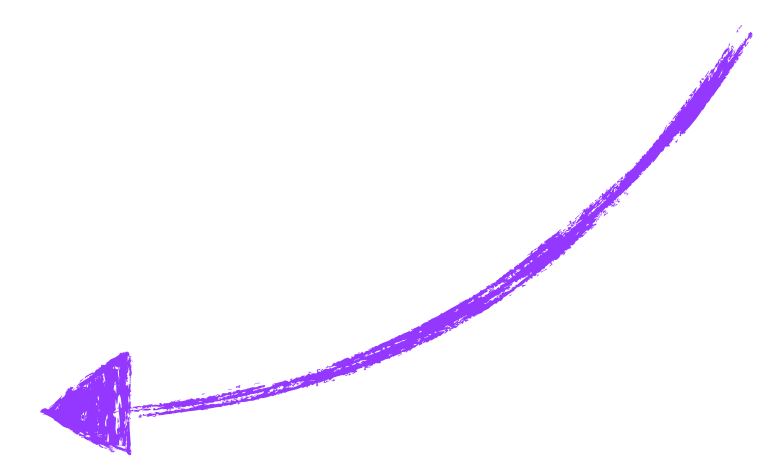


Improved Topological Sampling

Label even (+1) and odd (-1) sectors with $\sigma = \text{sign}(\text{Re det } D)$



many more transitions!



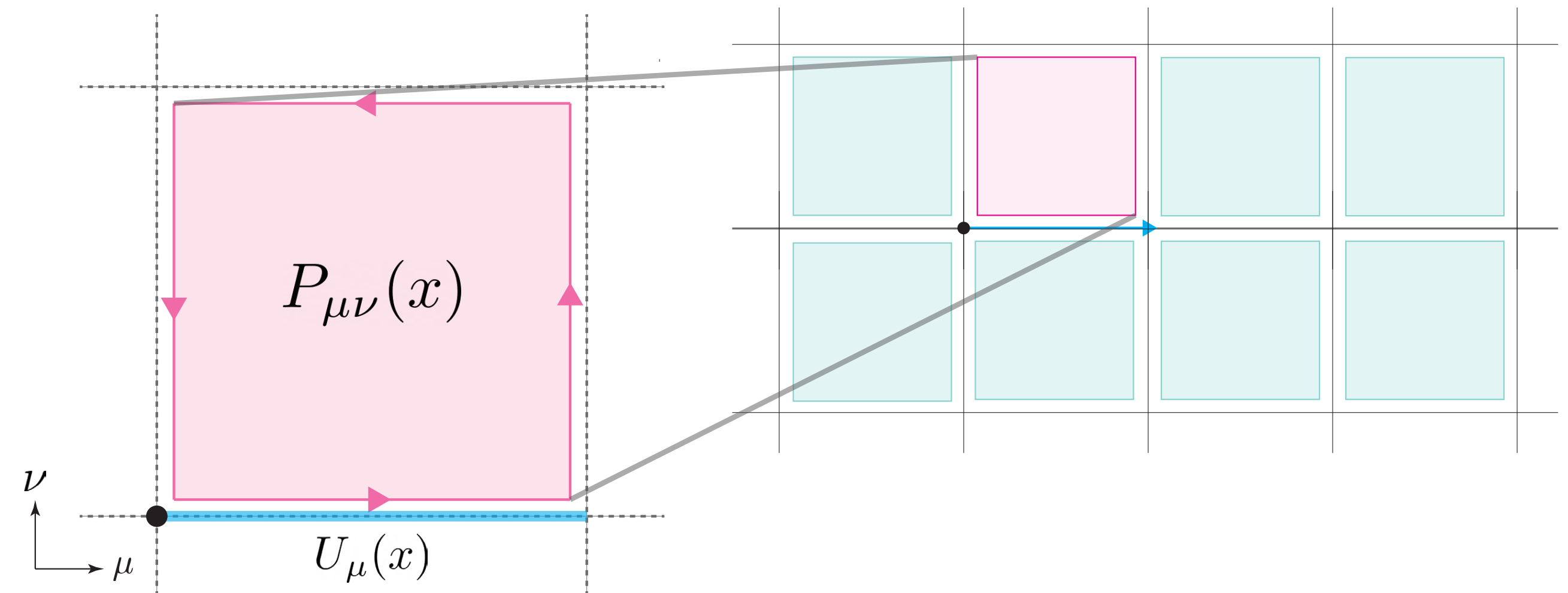
Sampling $SU(N)$ Lattice Gauge Theory

Non-Abelian link variables $U_\mu(x) \in SU(N)$

$$S[U] = -\frac{\beta}{N} \sum_x \sum_{\mu < \nu} \text{ReTr} P_{\mu\nu}(x)$$

Idea: *Energy-based model*

Du and Mordatch, *NeurIPS* **32** (2019)



Learn gauge-invariant effective action $\hat{S}_t^\theta[U_t] \approx S_t[U_t]$

- Autograd to get **effective force**: $\hat{b}_t^\theta(U_t) = -\nabla_U \hat{S}_t^\theta[U_t] \in \mathfrak{su}(N)$

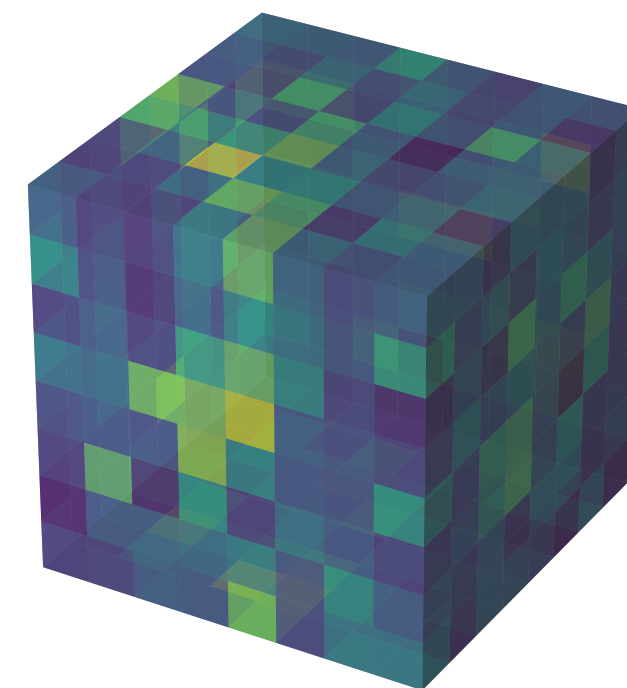
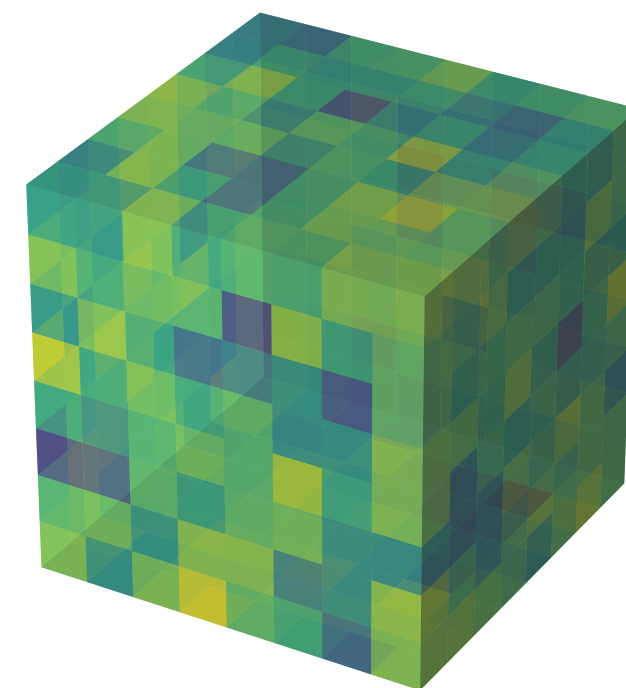
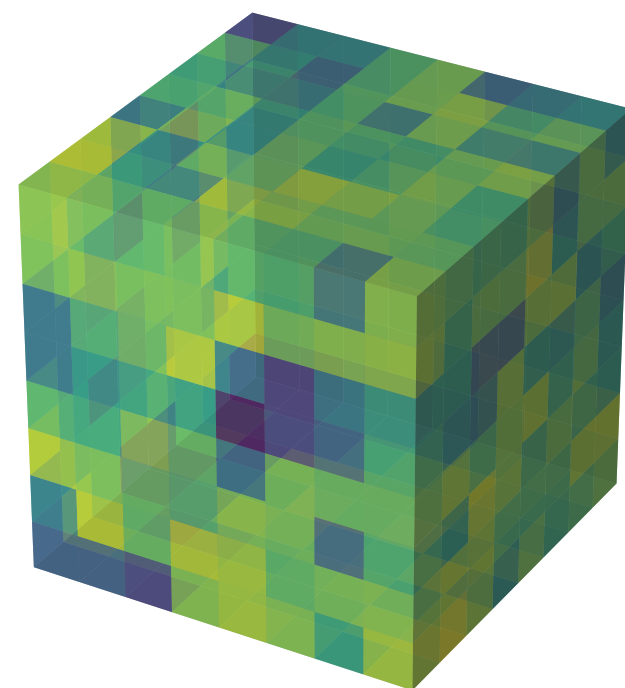
Outlook

High-Dimensional Generation

Spacetime is **4D**, but ESS **scales exponentially poorly** with lattice volume

What we need to combat this:

- More expressive model architectures
- New sampling methods?



Enhanced Sampling

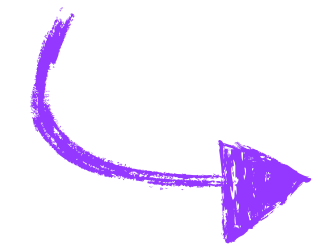
Albergo and Vanden-Eijnden, *PMLR* **267** (2024)

Looking into *Non-Equilibrium Transport Sampling (NETS)*

Solve coupled SDE/ODE system for samples ϕ_t and log-weights A_t :

re-weighting
factors

$$d\phi_t = b_t(\phi_t)dt - \epsilon_t \nabla S_t[\phi_t]dt + \sqrt{2\epsilon_t}dW_t, \quad \phi_0 \sim p_0$$



$$dA_t = \nabla \cdot b_t(\phi_t)dt - \nabla S_t[\phi_t] \cdot b_t(\phi_t)dt - \partial_t S_t[\phi_t]dt, \quad A_0 = 0$$

Enables injecting noise into reverse process, but **preserves likelihood calculation**

- Gives true **unbiased** diffusion-based sampling
- Tunable sampling parameters **after training** to improve ESS

The Road to QCD

SU(3) gauge theory coupling N_f flavors of **quarks** to **bosons** in **color space**

$$S_{\text{QCD}} [U, \bar{\psi}, \psi] = -\frac{\beta}{3} \sum_x \sum_{\mu < \nu} \text{ReTr} P_{\mu\nu}(x) + \sum_{x,y} \sum_{f=1}^{N_f} \bar{\psi}_\alpha^{(f)}(x)^i D[U](x,y)_{ij}^{\alpha\beta} \psi_\beta^{(f)}(y)^j$$

Modular approach: combine all of our machinery

- SU(N)-equivariant sampling of **gauge fields** U_μ
- Architectures suited to **pseudofermions**
- Scalable sampling for **large lattice volumes** in 4D

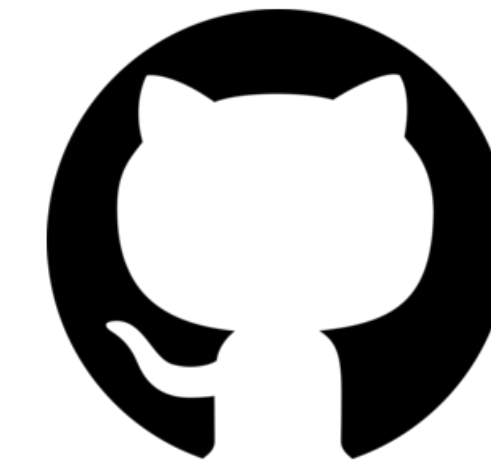
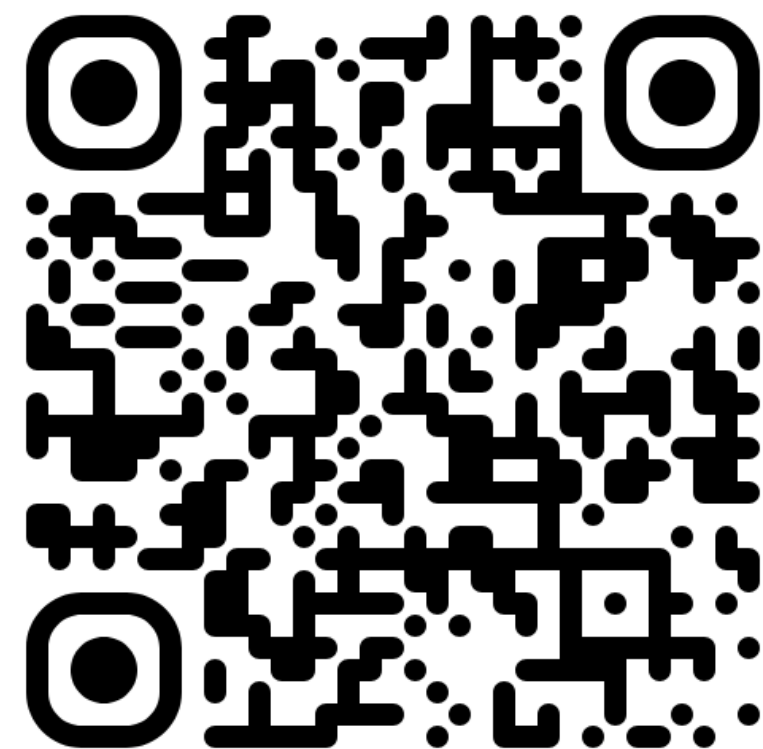
Computing and Software Tools

Development of internal codebase

- based on PyTorch
- distributed training and inference

Sharing public **software toolkits**

- see [sun_diffusion](https://github.com/ovega14/sun_diffusion) repo!



https://github.com/ovega14/sun_diffusion

```
1 import torch
2 import sun_diffusion
3 from sun_diffusion.diffusion import PowerDiffusionSUN
4
5 batch_size = 512
6 U_0 = random_sun_element(batch_size, Nc=2, scale=0.1)
7
8 # Diffuse U_0 -> U_1 on SU(2)
9 diffuser = PowerDiffusionSUN(kappa=3.0, alpha=1.0)
10 U_1, xs, V = diffuser.diffuse(U_0, torch.ones(batch_size))
```

G. Kanwar and OV, Zenodo (2026)

References I

- K. G. Wilson, *Confinement of quarks*, *Phys. Rev. D* **10** (1974) 2445.
- N. Metropolis, A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller, and E. Teller, *Equation of State Calculations by Fast Computing Machines*, *J. Chem. Phys.* **21** (1953) 1087.
- S. Duane, A. D. Kennedy, B. J. Pendleton, and D. Roweth, *Hybrid Monte Carlo*, *Phys. Lett. B* **195** (1987) 216.
- M. Lüscher, *Trivializing maps, the Wilson flow and the HMC algorithm*, *Commun. Math. Phys.* **293** (2010) 899 [[arXiv:0907.5491](https://arxiv.org/abs/0907.5491)].
- Yang Song, Jascha Sohl-Dickstein, Diederik P. Kingma, Abhishek Kumar, Stefano Ermon, and Ben Poole, *Score-Based Generative Modeling through Stochastic Differential Equations*, *ICLR (2021)* [[arXiv:2011.13456](https://arxiv.org/abs/2011.13456)].
- C. Saharia et al, *Photorealistic text-to-image diffusion models with deep language understanding*, *NeurIPS* **36** (2022) 36479 [[arXiv:2205.11487](https://arxiv.org/abs/2205.11487)].
- O. Vega, J. Komijani, A. El-Khadra, and M. Marinkovic, *Group Equivariant Diffusion Models for Lattice Field Theory*, [arXiv:2510.26081](https://arxiv.org/abs/2510.26081).
- O. Ronneberger, P. Fischer, and T. Brox, *U-Net: Convolutional networks for biomedical image segmentation*, *Lecture Notes in Computer Science* **9351** (2015) 234 [[arXiv:1505.04597](https://arxiv.org/abs/1505.04597)].

References II

- G. Kanwar and O. Vega, *Spectral Diffusion for Sampling on $SU(N)$* , *PoS LATTICE2025* (2026) 040 [[arXiv:2512.19877](https://arxiv.org/abs/2512.19877)].
- D. Boyda, G. Kanwar, S. Racanière, D. J. Rezende, M. S. Albergo, K. Cranmer, D. C. Hackett, and P. E. Shanahan, *Sampling Using $SU(N)$ Equivariant Flows*, *Phys. Rev. D* **103** (2021) 074504 [[arXiv:2008.05456](https://arxiv.org/abs/2008.05456)].
- J. Schwinger, *Gauge Invariance and Mass. II*, *Phys. Rev.* **128** (1962) 2425.
- F. Yu and V. Koltun, *Multi-Scale Context Aggregation by Dilated Convolutions*, [arXiv:1511.07122](https://arxiv.org/abs/1511.07122).
- J. Ho, N. Kalchbrenner, D. Weissenborn, and T. Salimans, *Axial Attention in Multidimensional Transformers*, [arXiv:1912.12180](https://arxiv.org/abs/1912.12180).
- M. S. Albergo, D. Boyda, K. Cranmer, D. C. Hackett, G. Kanwar, S. Racanière, D. J. Rezende, F. Romero-López, P. E. Shanahan, and J. M. Urban, *Flow-Based Sampling in the Lattice Schwinger Model at Criticality*, *Phys. Rev. D* **106** (2022) 014514 [[arXiv:2202.11712](https://arxiv.org/abs/2202.11712)].
- Du and I. Mordatch, *Implicit Generation and Generalization in Energy-Based Models*, *NeurIPS* **32** (2019) 3608 [[arXiv:1903.08689](https://arxiv.org/abs/1903.08689)].
- M. Albergo and E. Vanden-Eijnden, *NETS: A Non-Equilibrium Transport Sampler*, *PMLR* **267** (2025) 1026 [[arXiv:2410.02711](https://arxiv.org/abs/2410.02711)].
- G. Kanwar and O. Vega, [sun_diffusion v1.0.0](https://zenodo.org/record/14888887), *Zenodo* (2026).

Backup

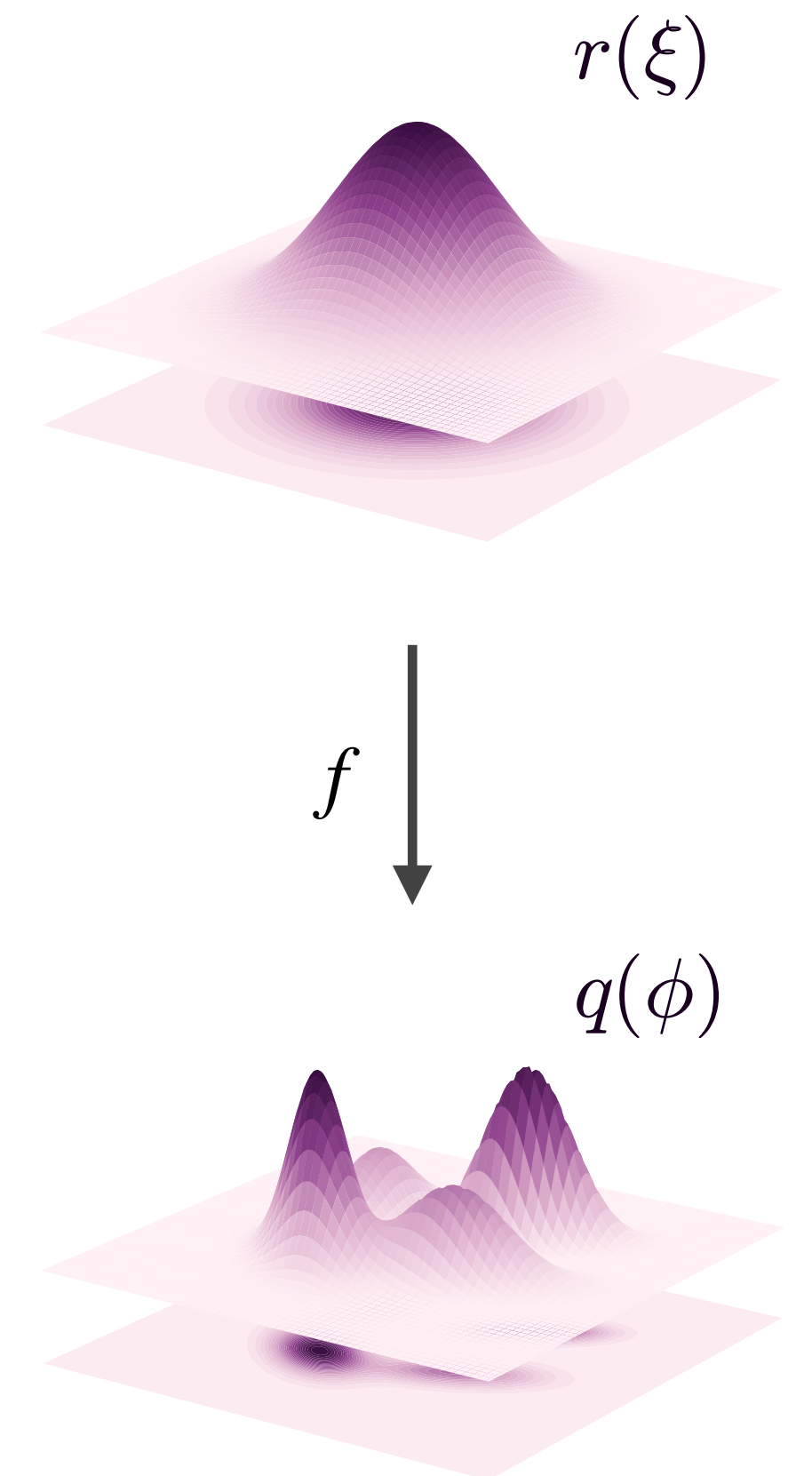
Computing Model Likelihoods

Change of variables: Differentiable bijection $\phi = f(\xi)$, $\xi \sim r(\xi)$

$$q(\phi) = r(\xi) \left| \det \frac{\partial \xi_i}{\partial \phi_j} \right|$$

Continuous time: Solution to the Fokker-Planck equation

$$q_t(\phi_t) = \underbrace{\exp \left[- \int_{t'}^t \nabla \cdot \hat{b}_\tau(\phi_\tau) d\tau \right]}_{\text{Jacobian}} q_{t'}(\phi_{t'})$$



Bias Correction and Model Quality

Given **importance weights** $w(\phi) = \frac{p(\phi)}{q(\phi)}$, define *effective sample size* (ESS)

$$\text{ESS} := \frac{1}{n} \frac{\left(\sum_i w(\phi_i)\right)^2}{\sum_i w(\phi_i)^2}$$

fraction of "good" samples

Likelihood ratios enable **exact** calculations by

Re-weighting

$$\langle \mathcal{O}(\phi) \rangle_p = \left\langle \mathcal{O}(\phi) \frac{p(\phi)}{q(\phi)} \right\rangle_q$$

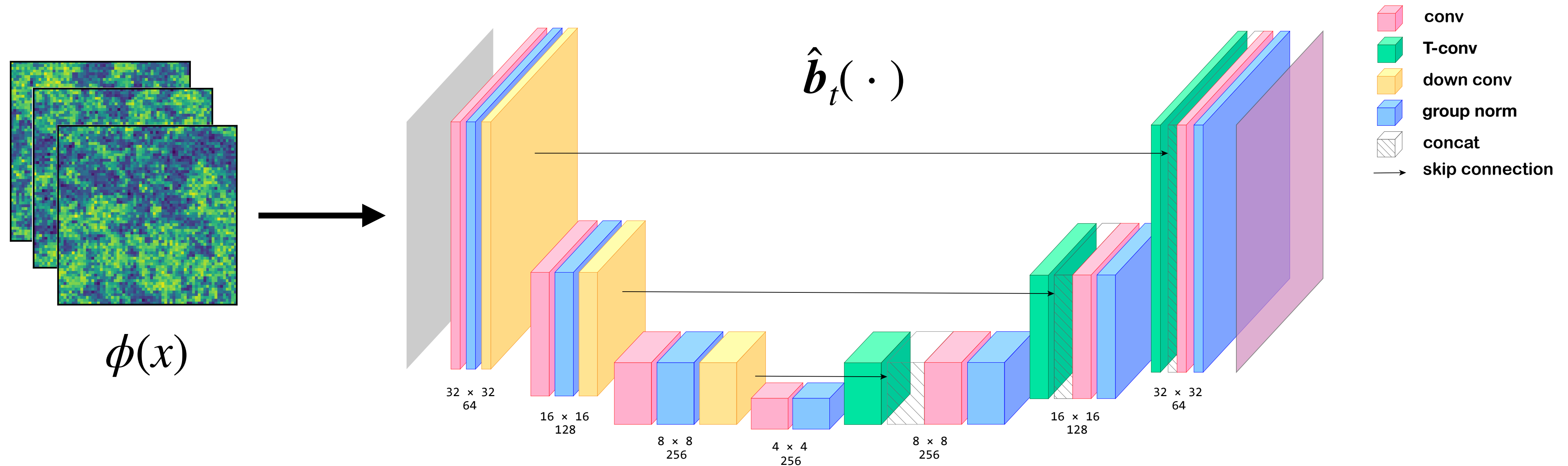
Re-sampling

$$\phi_1 \xrightarrow[\text{reject}]{\text{accept}} \phi_2 \xrightarrow[\times]{\checkmark} \cdots \xrightarrow[\times]{\checkmark} \phi_n$$

The U-Net Architecture

Ronneberger et al., *MICCAI* (2015)

2D field configurations $\phi(x)$ are like "images" \rightarrow use vision models



multiscale segmentation \leftrightarrow renormalization

The Spectral $SU(N)$ Heat Kernel

Two formulations for K_t over $SU(N)$ **eigenangles** θ

1. **Wrapped normal**: sum over equivalent pre-images under **exp**

$$K_t(\boldsymbol{\theta}) \propto \sum_{\mathbf{m} \in \mathbb{Z}^N} \delta\left(\sum_i m_i\right) \prod_{i < j} \frac{\theta_i - \theta_j + 2\pi(m_i - m_j)}{2 \sin\left[\frac{\theta_i - \theta_j}{2}\right]} \exp\left[-\frac{1}{2\sigma_t^2} \|\boldsymbol{\theta} + 2\pi\mathbf{m}\|^2\right]$$

2. **Dual character**: eigenfunction expansion

$$K_t(U) = \sum_{\mu} \dim(\mu) \chi_{\mu}(U) \exp\left[-c_2(\mu) \sigma_t^2\right]$$

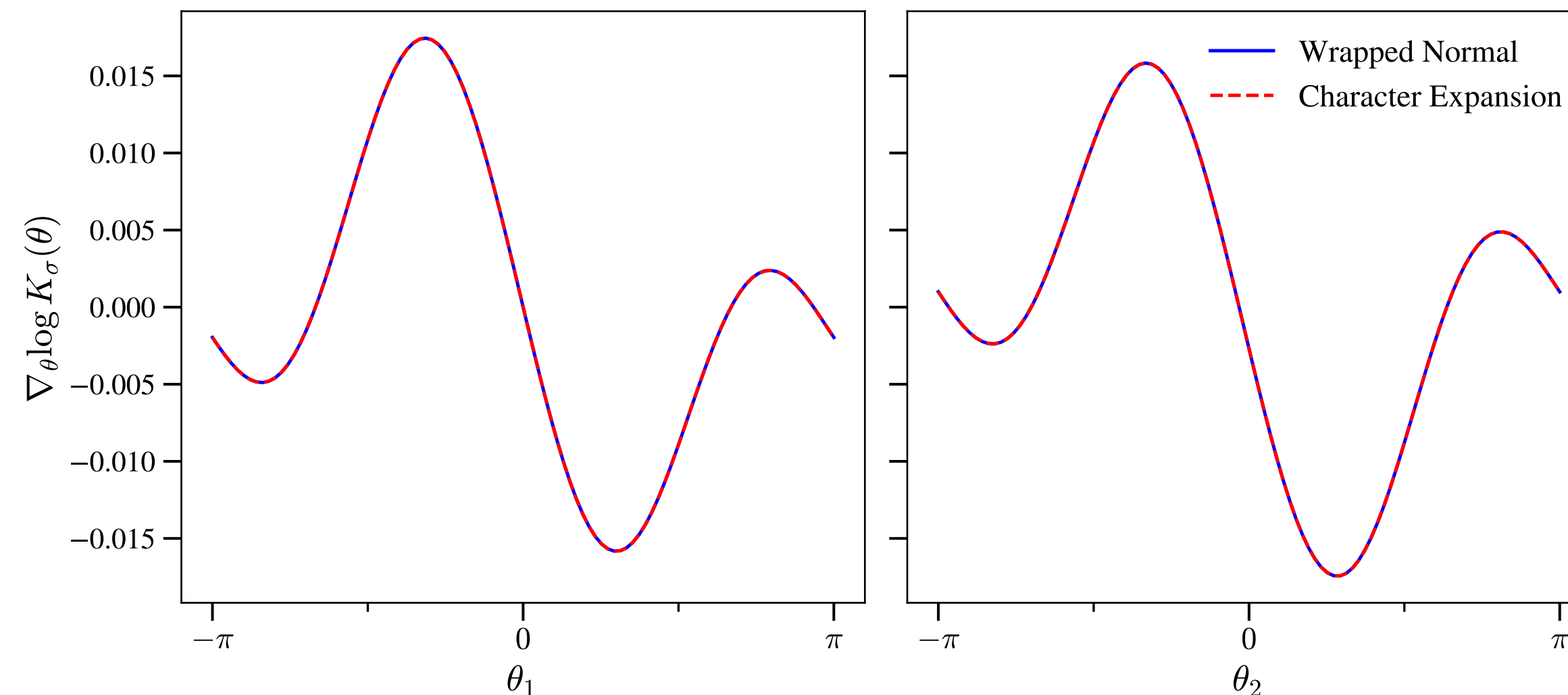
Training on the Group

Train according to modified objective function

$$\mathcal{L} \left[\hat{b} \right] = \int_0^1 \mathbb{E}_{U_0 \sim p_0} \mathbb{E}_{V \sim K_t} \left[\sigma_t^2 \left\| \hat{b}_t(VU_0) - P^\dagger \text{diag}[\nabla \log K_t(\theta)] P \right\|^2 \right] dt$$

Have access to **exact** heat kernel score functions $\nabla \log K_t(\theta)$

SU(3) HK score, $\sigma = 2.0$



Marginal Schwinger Action

Integrating out fermions \leftrightarrow marginal probability:

$$p(U) = \int \mathcal{D} [\bar{\psi}, \psi] p(U, \bar{\psi}, \psi) \propto e^{-S[U]}$$

Marginal density defines an **effective action**

$$S[U] = -\beta \sum_x \text{Re } P(x) - \log \det (D[U]^\dagger D[U])$$

⚠ fermion determinant $\det D^\dagger D$ highly **nonlocal**

encodes all fermion info