Sampling the Schwinger Model with Gauge-Equivariant Diffusion

Octavio Vega **Aida X. El-Khadra**
Department of Physics
University of Illinois Urbana-Champaign
Urbana, IL 61801
octavio5@illinois.edu

Abstract

We present a first study of a diffusion-based approach to accelerated sampling of the $N_f = 2$ lattice Schwinger model. Our work is inspired by recent and growing successes in developing such generative models for ensemble generation in LFT to overcome the well-known critical slowing down problem [16–22]. We train a U(1)-equivariant score-based generative model to sample gauge link configurations from the marginal Schwinger model. By computing model likelihoods, we obtain unbiased estimates for observables that closely match those produced by MCMC simulations. We also demonstrate improvement over HMC as measured qualitatively by a reduction in topological freezing near critical parameters.

1 Introduction

Across many domains in physics, quantum field theory (QFT) is a useful tool for studying interacting systems. Lattice field theory (LFT) offers a non-perturbative avenue for simulating these systems by discretizing continuum theories onto a Euclidean spacetime grid. LFT is based on the path integral formalism of QFT, where physical observables are computed as weighted expectations:

$$\langle \mathcal{O} \rangle = \frac{1}{\mathcal{Z}} \int [\mathcal{D}U] \; \mathcal{O}(U) e^{-S[U]} \approx \frac{1}{N} \sum_{i=1}^{N} \mathcal{O}(U_i). \tag{1}$$

The functional measure $[\mathcal{D}U]$ represents a sum over all possible configurations of the fields U. In LFT, numerical simulations are used to approximate the expectation values of Eq. (1) via empirical means as in the right-most expression of (1), which amounts to sampling large ensembles of field configurations $\{U_i\}_{i=1}^N$ from the physical distribution $p(U) = {}^{e^{-S[U]}}/_{\mathcal{Z}}$.

As LFT simulations approach the continuum limit of the underlying models, Markov chain Monte Carlo (MCMC) sampling algorithms become severely hindered, a phenomenon known as *critical slowing down* [2]. Autocorrelation times between samples grow exponentially large due to low acceptance rates between new proposals. At the same time, for topologically non-trivial theories, samples can become “frozen" in particular modes and thereby fail to explore the various topological sectors of configuration space [3, 4].

One particular system of interest is the Schwinger model [5], a fermionic QFT representing quantum electrodynamics in two dimensions. It is an effective toy model for quantum chromodynamics (QCD), as it exhibits confinement, a chiral anomaly, and non-trivial topology. Previous works using normalizing flows [6–8] have demonstrated an accelerated approach to sampling in the lattice Schwinger model [10, 11]. In this work, we present another method rooted in diffusion models. To our knowledge, this is the first application of diffusion-based sampling to a lattice field theory with fermions.

2 Diffusion Models

Diffusion models [12–14] are a type of generative model formulated through stochastic differential equations (SDE). The *forward process*, relevant for training, is responsible for corrupting initial configurations by repeatedly injecting increasing levels of random noise over diffusion time. The *reverse process*, used to generate new samples, employs a trained score network to interpolate from noise to data by integrating an SDE along reverse diffusion time.

The forward process is captured by the variance-expanding SDE, written in Itô form as

$$d\phi_t = g_t dW_t, \tag{2}$$

and describes the progressive corruption of initial configurations ϕ_0 towards noise. The diffusion coefficient $g_t : [0,1] \to \mathbb{R}^+$ is a thermal diffusivity controlling the level of noise added to samples over time. When adapted to the Lie group U(1), the analytical solution to Eq. (2) yields

$$\phi_0 \mapsto \phi_t = \texttt{wrap}\left(\phi_0 + \sigma_t \eta\right), \quad \sigma_t = \sqrt{\int_0^t g_s^2 ds} \tag{3}$$

where $\eta \sim \mathcal{N}(0, \mathbb{I})$ and the `wrap` operation folds into the unit circle modulo 2π.

The SDE which reverses the forward process in Eq. (2) is itself a diffusion process:

$$d\phi_t = -\frac{g_t^2 + \tilde{g}_t^2}{2} \nabla \log p_t(\phi_t) + \tilde{g}_t dW_t. \tag{4}$$

The gradient field $\nabla \log p_t$ is known as the *score function*, and can be estimated by a time-dependent neural network $\hat{\boldsymbol{s}}_t$ trained through score matching [15]. The parameters of $\hat{\boldsymbol{s}}_t$ are optimized by minimizing the conditional score matching objective:

$$\mathcal{L}\left[\hat{\boldsymbol{s}}\right] = \int_0^1 \mathbb{E}_{\phi_0 \sim p_0} \mathbb{E}_{\phi_t \sim p_{t|0}} \left[\sigma_t^2 \left\| \hat{\boldsymbol{s}}_t(\phi_t) - \nabla \log p_{t|0}(\phi_t|\phi_0) \right\|^2\right] dt. \tag{5}$$

Then, the reverse process (4) can be simulated using a numerical ODE solver to obtain new samples. In order to calculate the likelihoods of samples under the model, we set $\tilde{g}_t \equiv 0$ and integrate the deterministic ODE in reverse as in continuous normalizing flows [9].

3 The Schwinger Model on the Lattice

3.1 Definitions

The action for the lattice Schwinger model is defined as

$$S[U, \psi, \bar{\psi}] = -\beta \sum_x \operatorname{Re} P(x) + \sum_{x,y} \bar{\psi}_\alpha(x) D[U](x,y)^{\alpha\beta} \psi_\beta(y), \tag{6}$$

where $P(x) = U_0(x) U_1(x+\hat{0}) U_0^*(x+\hat{1}) U_1^*(x)$ is the U(1) plaquette. Distinct lattice sites are denoted by x, y and $\alpha, \beta \in \{1, 2\}$ are spin indices that are summed. The differential operator $D[U]$ is the Wilson Dirac operator, given by

$$\begin{aligned} D[U](x,y)^{\alpha\beta} = \delta(x,y)\delta^{\alpha\beta} - \kappa \sum_\mu \Big\{ & U_\mu^*(y)\delta(x, y+\hat{\mu}) \left[\mathbb{I} + \sigma_\mu\right]^{\alpha\beta} \\ & + U_\mu(y-\hat{\mu})\delta(x, y-\hat{\mu}) \left[\mathbb{I} - \sigma_\mu\right]^{\alpha\beta} \Big\}. \end{aligned} \tag{7}$$

The matrices σ_μ for $\mu = 1, 2$ are the Hermitian 2×2 Pauli matrices and $\mathbb{I}$ is the identity matrix in spin space. The strength of the coupling between the bosonic variables U and the fermions ψ is controlled by the *Hopping parameter* κ, related to the bare fermion mass m_0 via $\kappa = \frac{1}{4+2m_0}$.

The full action (6) defines a joint probability $p\left(U, \psi, \bar{\psi}\right)$ over the gauge links and fermions which can be marginalized by integrating out the fermions [23, 24], giving a marginal density

$$p(U) \propto \det\left(D[U]^\dagger D[U]\right) e^{-S[U]}. \tag{8}$$

In turn, the marginal density defines an effective action $S[U] = -\log p(U)$ via

$$S[U] = -\beta \sum_x \operatorname{Re} P(x) - \log\det\left(D[U]^\dagger D[U]\right). \tag{9}$$

As the gauge links $U_\mu \in \mathbb{C}$ are complex numbers, we use the parameterization $U_\mu = \exp(iA_\mu)$ and work directly with the phases $A_\mu \in [0, 2\pi)$. Our goal is therefore to sample gauge links A_μ from the marginal density defined by the exact-determinant action in Eq. (9).

3.2 Computational Setup

For our simulations, we investigate two lattice sizes: 8×8 and 16×16. We consider $N_f = 2$ flavors of degenerate fermions, and the action parameters we use are $\beta = 2.0$ and $\kappa = 0.276$. Our training datasets consist of 8192 gauge link configurations generated using Hybrid Monte Carlo (HMC) [25] with 200 trajectories each consisting of 10 leapfrog steps at step size 0.1.

The prior distribution is taken to be uniform with respect to the Haar measure, which for U(1) is $1/2\pi$ on $[0, 2\pi)$ and 0 elsewhere. The diffusion schedule is chosen to be linear in time: $g_t = 3t$, yielding a marginal width $\sigma_t = \sqrt{3t^3}$.

Models are trained for 1000 epochs in batches of 512 with a learning rate of 0.004 using the Adam optimizer [28]. The score network uses a U-Net parameterization [26] incorporating time encoding via Gaussian Fourier embeddings [27]. For inference, we integrate the reverse ODE using an Euler integrator with 500 steps and step size 0.002. Likelihoods are obtained using the Skilling-Hutchinson trace estimator [29, 30] with 10 Rademacher random vectors.

The high degree of non-locality introduced by the fermion determinant demands architectural features with greater expressive potential. To that end, we have thus far experimented with dilated convolutions [31] as well as axial attention layers [32] to learn long-range correlations and expand the receptive field of our network. Moreover, in Ref. [18] we showed how to construct a U(1) gauge-equivariant score network by feeding it gauge-invariant plaquettes. Here, we use rectangular 2×1 Wilson loops as additional gauge-invariant input channels to provide greater coverage over the lattice, as was done in Ref. [10].

4 Results

4.1 Sampling Dynamics

The integer-valued topological charge is used to label topological sectors and is defined as

$$Q = \frac{1}{2\pi} \sum_x \arg P(x) \in \mathbb{Z}. \tag{10}$$

Another observable was suggested in Ref. [23] to identify tunneling events between even and odd topological sectors, given by

$$\sigma = \operatorname{sign}\left(\operatorname{Re}\det D\right). \tag{11}$$

For comparison, we resample [33] the diffusion-generated configurations using 10,000 Metropolis accept/reject steps [34] and generate an HMC dataset with the same number of steps. We track both Q and σ during MCMC time for both HMC and diffusion, as depicted in Fig. 1. Though the evolution of Q is comparable, we observe significantly more even-odd transitions with the diffusion model when compared to HMC, suggesting that diffusion-based sampling can enable much wider and more efficient phase space exploration.

4.2 Observables

We validate the diffusion-generated configurations by computing physical observables on them and comparing them to those computed over a reference dataset generated with HMC.

Two relevant observables carry over from pure gauge theory, which are the average plaquette $P = \operatorname{Re}\sum_x P(x)$ and the topological susceptibility $\chi_Q = \langle Q^2 \rangle$. A fermionic observable we also investigate is the chiral condensate $\langle \bar{\psi}\psi \rangle$, given by the volume-averaged trace of $D^{-1}[U]$.

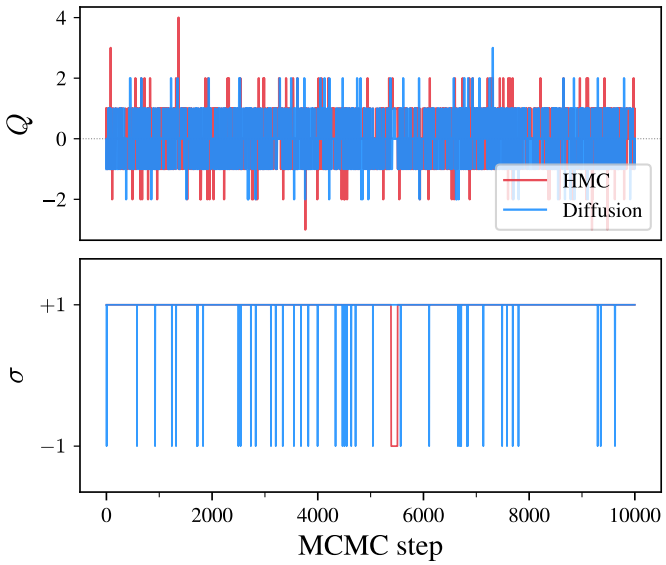

Figure 1: Evolution of the topological charge Q and Dirac sign σ during sampling on an 8×8 lattice.

In Table 1, we show the effective sample size (ESS), MCMC resampling acceptance rate (AR), and the estimated unnormalized Kullback-Leibler divergence (KL) along with the three observables across the two lattice sizes for both HMC and diffusion. We emphasize that these observable measurements are preliminary, and more work is needed to understand their uncertainties and scaling behavior.

		Metrics				**Observables**		
L	**Method**	ESS ↑	AR ↑	KL ↓	$\tau_{\text{int}}(Q)$ ↓	$\langle P \rangle$	χ_Q	$\langle \bar{\psi}\psi \rangle$
8	HMC	–	–	–	1.22	0.7597(14)	0.0038(3)	1.502(4)
	Diffusion	0.29	35%	−295.0	2.92	0.7607(32)	0.0037(5)	1.504(14)
16	HMC	–	–	–	12.95	0.7585(8)	0.0040(2)	1.507(4)
	Diffusion	0.08	20%	−1180.2	4.91	0.7585(46)	0.0035(14)	1.501(25)

Table 1: Quality metrics and observables computed on 1024 samples generated from HMC and diffusion. Diffusion model observables are reweighted.

5 Outlook

Further experimentation across model architectures is needed to identify a scalable approach to sampling with fermions. Future work will investigate adaptations of diffusion-based sampling to different schemes for handling fermions [35], including with pseudofermions both in the Schwinger model and in $\mathrm{SU}(N)$ lattice gauge theory [36].

Acknowledgments and Disclosure of Funding

We thank Gurtej Kanwar, Javad Komikani, and Marina Marinkovic for helpful discussions. AXK and OV acknowledge support from the U.S. Department of Energy, Office of Science under grant Contract Number DE-SC001565. OV is also grateful for support received from a UIUC Graduate College Fellowship and from a Sloan Scholarship through the Alfred P. Sloan Foundation. Numerical experiments and data analysis were executed using **PyTorch** [37] and **NumPy** [38]. Figures were created with **Matplotlib** [39]. The simulations presented in this paper utilized the DeltaAI advanced computing and data resource, which is supported by the National Science Foundation (award OAC2320345) and the State of Illinois.

References

[1] K. G. Wilson, *Confinement of quarks*, *Phys. Rev. D* **10** (1974) 2445.

[2] U. Wolff, *Critical Slowing Down*, *Nucl. Phys. B Proc. Suppl.* **17** (1990) 93.

[3] P. Woit, *Topological Charge in Lattice Gauge Theory*, *Phys. Rev. Lett.* **51** (1983) 638.

[4] B. Allés, G. Boyd, M. D'Elia, A. Di Giacomo, and E. Vicari, *Hybrid Monte Carlo and topological modes of full QCD*, *Phys. Lett. B* **389** (1996) 107 [hep-lat/9607049].

[5] J. Schwinger, *Gauge Invariance and Mass. II*, *Phys. Rev.* **128** (1962) 2425.

[6] D. Rezende and S. Mohamed, *Variational Inference with Normalizing Flows*, *PMLR* **37** (2015) 1530 [1505.05770].

[7] L. Dinh, J. Sohl-Dickstein and S. Bengio, *Density Estimation Using Real NVP*, 1605.08803.

[8] G. Papamakarios, E. Nalisnick, D. J. Rezende, S. Mohamed, and B. Lakshminarayanan, *Normalizing Flows for Probabilistic Modeling and Inference*, *JMLR* **22** (2021) 2617 [1912.02762].

[9] R. T. Q. Chen, Y. Rubanova, J. Bettencourt, and D. Duvenaud, *Neural ordinary differential equations*, *Adv. Neural Inf. Process. Syst.* **31** (2018) 6572 [1806.07366].

[10] 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 [2202.11712].

[11] J. Finkenrath, *Tackling critical slowing down using global correction steps with equivariant flows: the case of the Schwinger model*, 2201.02216.

[12] J. Sohl-Dickstein, E. Weiss, N. Maheswaranathan, S. Ganguli, *Deep Unsupervised Learning using Nonequilibrium Thermodynamics*, *PMLR* **37** (2015) 2256 [1503.03585].

[13] J. Ho, A. Jain, and P. Abbeel, *Denoising diffusion probabilistic models*, *Adv. Neural Inf. Process. Syst.* **33** (2020) 6840 [2006.11239].

[14] Y. Song, J. Sohl-Dickstein, D. P. Kingma, A. Kumar, S. Ermon, and B. Poole, *Score-Based Generative Modeling through Stochastic Differential Equations*, *ICLR* (2021) [2011.13456].

[15] Y. Song and S. Ermon, *Generative modeling by estimating gradients of the data distribution*, *Adv. Neural Inf. Process. Syst.* **32** (2019) 11918 [1907.05600].

[16] L. Wang, G. Aarts and K. Zhou, *Diffusion models as stochastic quantization in lattice field theory*, *JHEP* **05** (2024) 060 [2309.17082].

[17] Q. Zhu, G. Aarts, W. Wang, K. Zhou, and L. Wang, *Physics-Conditioned Diffusion Models for Lattice Gauge Theory*, *JHEP* **03** (2026) 111 [2502.05504].

[18] O. Vega, J. Komijani, A. El-Khadra, and M. Marinkovic, *Group-Equivariant Diffusion Models for Lattice Field Theory*, 2510.26081.

[19] G. Kanwar and O. Vega, *Spectral Diffusion for Sampling on* $\mathrm{SU}(N)$, *PoS* **LATTICE2025** (2026) 040 [2512.19877].

[20] G. Aarts, D. E. Habibi, A. Ipp, D. I. Müller, T. R. Ranner, L. Wang, W. Wang and Q. Zhu, *Generalizable Equivariant Diffusion Models for Non-Abelian Lattice Gauge Theory*, 2601.19552.

[21] H. Alharazin, J. Y. Panteleeva and B. D. Sun, *Diffusion Models for* $\mathrm{SU}(2)$ *Lattice Gauge Theory in Two Dimensions*, 2602.09045.

[22] J. Komijani, M. K. Marinkovic and L. Turgut, *Diffusion model for* $\mathrm{SU}(N)$ *gauge theories*, 2605.06134.

[23] C. R. Gattringer, I. Hip, and C. B. Lang, *Topological charge and the spectrum of the fermion matrix in lattice QED*$_2$, *Nucl. Phys. B* **508** (1997) 329 [hep-lat/9707011].

[24] C. Gattringer and C. Lang, *Quantum Chromodynamics on the Lattice: An Introductory Presentation*, *Lecture Notes in Physics* **788**, Springer Berlin, Heidelberg (2010) [DOI:10.1007/978-3-642-01850-3].

[25] S. Duane, A. D. Kennedy, B. J. Pendleton, and D. Roweth, *Hybrid Monte Carlo*, *Phys. Lett. B* **195** (1987) 216.

[26] O. Ronneberger, P. Fischer, and T. Brox, *U-Net: Convolutional Networks for Biomedical Image Segmentation*, *Proc. MICCAI* (2015) [1505.04597].

[27] M. Tancik, P. P. Srinivasan, B. Mildenhall et al., *Fourier Features Let Networks Learn High Frequency Functions in Low Dimensional Domains*, *Adv. Neural Inf. Process. Syst.* **33** (2020) 7537 [2006.10739].

[28] D. P. Kingma and J. Ba, *Adam: A Method for Stochastic Optimization*, 1412.6980.

[29] J. Skilling, *Maximum Entropy and Bayesian Methods*, *Fundamental Theories of Physics* **36**, Springer Netherlands, Dordrecht (1989) [DOI:10.1007/978-94-015-7860-8_48].

[30] M. F. Hutchinson, *A Stochastic Estimator of the Trace of the Influence Matrix for Laplacian Smoothing Splines*, *Commun. Stat. Simul. Comput.* **18** (1989) 1059.

[31] F. Yu and V. Koltun, *Multi-Scale Context Aggregation by Dilated Convolutions*, *ICLR* (2016) [1511.07122].

[32] J. Ho, N. Kalchbrenner, D. Weissenborn, and T. Salimans, *Axial Attention in Multidimensional Transformers*, 1912.12180.

[33] A. Doucet, N. Freitas, and N. Gordon, *Sequential Monte Carlo Methods in Practice*, Springer New York (2001) [DOI:10.1007/978-1-4757-3437-9].

[34] 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.

[35] M. S. Albergo, G. Kanwar, S. Racanière, D. J. Rezende, J. M. Urban, D. Boyda, K. Cranmer, D. C. Hackett, and P. E. Shanahan, *Flow-based sampling for fermionic lattice field theories*, *Phys. Rev. D* **104** (2021) 114507 [2106.05934].

[36] R. Abbott, M. S. Albergo, D. Boyda, K. Cranmer, D. C. Hackett, G. Kanwar, S. Racanière, D. J. Rezende, F. Romero-López, P. E. Shanahan, B. Tian, and J. M. Urban, *Gauge-Equivariant Flow Models for Sampling in Lattice Field Theories with Pseudofermions*, *Phys. Rev. D* **106** (2022) 074506 [2207.08945].

[37] A. Paszke, S. Gross, F. Massa, A. Lerer et al., *PyTorch: An Imperative Style, High-Performance Deep Learning Library*, *Adv. Neural Inf. Process. Syst.* **32** (2019) 8026 [1912.01703].

[38] C. R. Harris, K. J. Millman, S. J. van der Walt, R. Gommers, P. Virtanen, D. Cournapeau et al., *Array programming with NumPy*, *Nature* **585** (2020) 357 [2006.10256].

[39] J. D. Hunter, *Matplotlib: A 2D Graphics Environment*, *Comp. Sci. Eng.* **9** (2007) 90.