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, New York 10003, USA*
[2]*Argonne Leadership Computing Facility, Argonne National Laboratory, Lemont, Illinois 60439, USA*
[3]*Center for Theoretical Physics, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA*
[4]*The NSF AI Institute for Artificial Intelligence and Fundamental Interactions, Cambridge, Massachusetts 02139, USA*
[5]*Albert Einstein Center, Institute for Theoretical Physics, University of Bern, 3012 Bern, Switzerland*
[6]*DeepMind, S2, 8 Handyside Street, London, N1C 4DJ, United Kingdom*
[7]*Institut für Theoretische Physik, Universität Heidelberg, Philosophenweg 16, 69120 Heidelberg, Germany*

(Received 9 March 2022; accepted 12 July 2022; published 29 July 2022)

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.

DOI: 10.1103/PhysRevD.106.014514

I. INTRODUCTION

Many important physical systems across particle and condensed matter physics can be described in the language of quantum field theory (QFT). Lattice field theory (LFT) is the only known systematically improvable approach to *ab initio* calculations of QFTs in nonperturbative regimes, such as quantum chromodynamics (QCD) at low energies. LFT is based on the path-integral formulation of QFT, discretized on a Euclidean spacetime lattice. Monte Carlo techniques render the high-dimensional discretized path integral tractable by recasting the problem as statistical sampling: the expectation value of some observable $\mathcal{O}$ can be computed as

$$\langle \mathcal{O} \rangle = \frac{1}{\mathcal{Z}} \int dU e^{-S_E(U)} \mathcal{O}(U) \simeq \frac{1}{N} \sum_{i=1}^{N} \mathcal{O}(U_i), \quad (1)$$

where $\mathcal{Z}$ is the partition function, S_E is the Euclidean action, and $\{U_i\}$ is a set of N samples of the lattice-field degrees of freedom distributed as $p(U) = \exp[-S_E(U)]/\mathcal{Z}$. Statistical uncertainties decrease as $1/\sqrt{N}$ as the estimate converges to the true value.

In theories such as QCD, for which exact sampling algorithms are not known, Markov chain Monte Carlo (MCMC) techniques are typically used. However, field samples or "configurations" from MCMC are correlated, i.e., subsequently generated configurations are not statistically independent. Depending on the MCMC approach, these "autocorrelations" may grow as the system is tuned towards criticality [1], e.g., to describe universal properties of condensed matter theories or access the continuum or large-N_c limits of gauge theories [2,3]. Autocorrelations may become especially severe if MCMC updates are unlikely to generate transitions between modes that are separated in configuration space. This effect, known as "freezing," can prevent effective exploration of the distribution for any practical sample size, amounting to an in-practice violation of ergodicity—a necessary condition for the validity of MCMC.

Importantly, this affects hybrid Monte Carlo (HMC), the state-of-the-art algorithm for sampling QCD field configurations, which generates samples by continuously evolving the fields through configuration space via Hamiltonian dynamics [4]. These dynamics make the algorithm susceptible to freezing due to the topological properties of gauge fields, which divide the distribution into different modes or "topological sectors." As the system is tuned towards criticality, increasingly large potential barriers

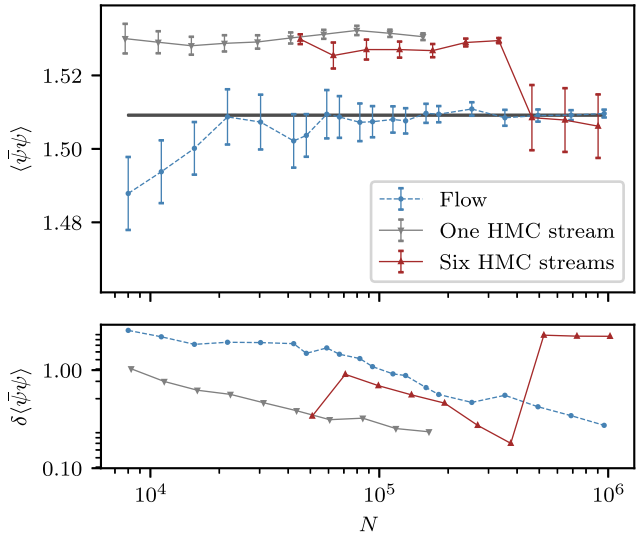

FIG. 1. Demonstration of underestimated uncertainties when using HMC, in contrast to flow-based sampling which provides consistent results converging to a baseline value (black). Top panel: estimates of the chiral condensate $\langle\bar{\psi}\psi\rangle$ in the Schwinger model at critical parameters $\beta = 2.0$, $\kappa = 0.276$, and $L = 16$, as a function of the sample size N. Bottom panel: scaling of the statistical uncertainty $\delta\langle\bar{\psi}\psi\rangle$. Flow-based sampling (blue) converges to the baseline value with uncertainties scaling as $1/\sqrt{N}$. Meanwhile, HMC (gray, red) exhibits seemingly convergent uncertainties that are in fact severely underestimated, as indicated by the discrepancy with the baseline and sudden jumps when tunneling events occur (red).

suppress tunneling between sectors, presenting a well-known obstacle to extending the reach of state-of-the-art lattice QCD calculations [1,5,6] (formalisms for lattice QCD at fixed topological charge are also being explored [7–9]). In contrast, emerging flow-based sampling algorithms [10–16] propose random hops throughout configuration space, unaffected by density barriers. Promising results of flow-based samplers and related approaches in theories without fermions [11,12,17,18], without gauge fields [14], or away from criticality [19] suggest that these methods may provide a path towards mitigating freezing in this context.

In this paper, we show that flow-based sampling can circumvent topological freezing in a fermionic gauge theory at criticality. Specifically, we provide a numerical demonstration in the Schwinger model (two-dimensional quantum electrodynamics) at the critical value of the fermion mass, illustrating that the flow-based approach is robust at sample sizes where HMC fails. A stark example is shown in Fig. 1, where HMC estimates of a key observable in the theory appear to be converging to a biased value, while in fact the uncertainty is underestimated due to insufficient sampling of the different topological sectors. In contrast, the flow-based sampling estimate is accurate with reliable uncertainties.

II. FLOW-BASED SAMPLING FOR THE SCHWINGER MODEL

Normalizing flow models [20–22] are based on applying a diffeomorphic "flow" transformation f to (possibly high-dimensional) samples z drawn from a base distribution, $r(z)$. This procedure yields samples $U = f(z)$ distributed according to the model density $q(U) = r(z)|\det \partial f/\partial z|^{-1}$. Flow-based sampling uses the model q to approximate a target distribution p. Neural networks can be used to construct an expressive and trainable flow, which can be optimized by minimizing the distance between p and q. Provably exact sampling that corrects for deviations between p and q can be obtained with independence Metropolis [23,24] or reweighting; we use the former in this work. These may be applied *a posteriori*, enabling embarrassingly parallel sampling that can provide practical advantages over HMC and sequential algorithms incorporating flows [17–19].

Here, we apply flow-based sampling to the $N_f = 2$ Schwinger model, a strongly interacting $(1+1)$d $U(1)$ gauge theory coupled to two fermions that exhibits similar features to QCD: confinement, spontaneous chiral symmetry breaking due to a chiral condensate, and nontrivial topology [25,26]. It commonly serves as a toy model for QCD, and is often used for testing new approaches to LFT [19,27–33], including methods using quantum technologies [34–36]. It has also been used to study properties of QFTs [26,37–46].

Wick rotating, discretizing, and integrating out the fermionic degrees of freedom yields a Euclidean lattice Schwinger model action amenable to Monte Carlo sampling [47–49],

$$S_E(U) = -\beta\sum_x \mathrm{Re}P(x) - \log\det D[U]^\dagger D[U], \quad (2)$$

given in terms of gauge links $U_\mu(x)$ at position x in direction μ. The first term is the gauge action, where β is the inverse of the squared gauge coupling, and the plaquette $P(x)$ is the smallest possible Wilson loop—a gauge-invariant product of links around a 1×1 square. It is defined as $P(x) = U_0(x)U_1(x+\hat{0})U_0^\dagger(x+\hat{1})U_1^\dagger(x)$, where $\hat{\mu}$ is the unit vector in the μ direction. The second term, given in terms of the Wilson Dirac operator D [50,51], encodes the effect of fermions and gauge-fermion interactions. The bare fermion mass m_0 is controlled by the hopping parameter $\kappa = 1/(4+2m_0)$ that parametrizes D.

To achieve efficient sampling via a flow-based approach, it is critical to incorporate the physical properties of the target distribution. For the Schwinger model specifically, gauge invariance imposes strong constraints on the target distribution, which we build into our models using the framework of gauge-equivariant flows on compact manifolds developed in Refs. [11,12,52]. Another challenge is sampling of theories with fermionic degrees of freedom.

Out of the four treatments in Ref. [14], here we consider a "marginal sampler" using exact evaluation of the fermion determinant. The model thus approximates a distribution of gauge links that encodes the effect of fermions as a nonlocal effective action for the gauge degrees of freedom.

Following Ref. [11], gauge-equivariant flows are constructed by composing a sequence of equivariant coupling layers. Each coupling layer updates an "active" subset of the links conditioned on a disjoint "frozen" subset. Different partitionings are used in each layer so that all variables are updated. In each layer, gauge-invariant closed Wilson loops are computed from the frozen links and fed into a "context function" constructed from neural networks. The outputs are used to parametrize the transformation of the active links, which is constrained to commute with gauge transformations. Combined with a gauge-invariant base distribution, this yields a gauge-invariant model.

Unlike in the $\kappa = 0$ limit of pure-gauge theory, the Schwinger model exhibits long-range correlations, with the correlation length defined by the inverse of the mass of the lightest particle. At criticality, the correlation length diverges; this demands new architectural features over those developed previously in Ref. [11] to model ultralocal U(1) dynamics. First, we use a subset of active links that is locally more sparse, with each active link completely surrounded by frozen ones, to allow for better propagation of information. Second, we provide larger 2×1 Wilson loops along with 1×1 plaquettes as inputs for context functions. Third, our architecture includes dilated convolutions, which have translational equivariance and better context aggregation, i.e., an exponential expansion of the receptive field without loss of resolution or coverage [53]. Fourth, we parametrize our transformations using highly expressive neural splines [54]. Finally, we decay the learning rate over the course of training.

We train this flow model for the Schwinger model at criticality and compare the performance of flow-based MCMC using this model against that of HMC. At finite lattice spacing, a diverging correlation length is realized by tuning κ to its critical value, resulting in a vanishing renormalized fermion mass. To achieve this, we take $\beta = 2.0$ and $\kappa = 0.276$, which correspond to the near-critical parameters found in Ref. [47]. We use a square lattice of extent $L = 16$. Details of the architecture, training scheme, and HMC parameters are in the Supplemental Material [55].

III. ADVANTAGES OF FLOW-BASED SAMPLING

A clear illustration of the advantages of flow-based sampling for the Schwinger model at criticality is given in Fig. 1, which compares estimates of the chiral condensate from HMC with those from flow-based MCMC. This quantity

$$\langle \bar{\psi}\psi \rangle = \frac{1}{V} \mathrm{Tr} D^{-1}[U] \tag{3}$$

is a simple fermionic observable whose value is correlated with the topological sectors and is therefore sensitive to freezing. We quantify uncertainties using the integrated autocorrelation time with the "gamma method" [56] and compute the baseline result using an augmented version of HMC that efficiently samples topological sectors. Clearly, the single frozen HMC stream yields estimates that are manifestly inconsistent with the baseline result, indicating severely underestimated uncertainties even at very large sample sizes, $N \approx 10^5$. The dataset of samples from six independent HMC streams can incorporate information from multiple topological sectors even in the presence of freezing. However, as the figure shows, this estimate is still biased for $N \approx 10^5$ samples with incorrect uncertainties deceptively scaling as $1/\sqrt{N}$. The estimate becomes consistent with the ground truth only when $N \gtrsim 10^6$. The uncertainty, however, catastrophically increases—a clear indication of an ergodicity problem. This analysis suggests that affordable HMC stream lengths may not be sufficient to diagnose bias. By contrast, flow-based results converge smoothly to the baseline value, with errors scaling as $1/\sqrt{N}$.

Figure 2 provides a more direct illustration of freezing in the Monte Carlo histories of topological quantities. The topological sectors of the Schwinger model are distinguished by the integer-valued topological charge. A common discretization is [42]

$$Q = \frac{1}{2\pi} \sum_x \mathrm{Im} \log P(x), \tag{4}$$

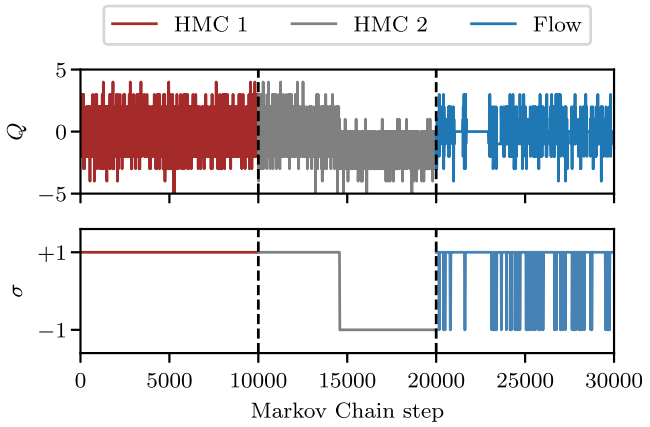

FIG. 2. Monte Carlo history of the topological charge, Q (top), and the sign of the real part of the determinant of the Dirac operator, σ (bottom). Results shown are based on three different streams of configurations: two HMC streams (red and gray) and a stream from our flow model (blue). In HMC, Q exhibits ultraviolet fluctuations associated with discretization effects, but σ rarely changes—an indication that true tunneling events between sectors are infrequent. Thus the HMC streams show clear evidence of topological freezing, while flow-based MCMC mixes rapidly.

where the imaginary part of the complex logarithm is restricted to $(-\pi, \pi]$. Due to lattice artifacts, this observable fluctuates even when the topological sector is fixed. A better-suited observable to identify true tunneling events was proposed in Ref. [47]:

$$\sigma = \text{sign}(\text{Re det } D). \tag{5}$$

The value of σ is positive (negative) for even (odd) topological sectors, and changes in its value are correlated with tunneling events.

Results for these observables based on two Markov chains generated with HMC (with different random number seeds) and one with flow-based sampling are illustrated in Fig. 2. In the first HMC stream, Q appears to fluctuate without any evidence of freezing. However, σ is completely frozen for all samples shown, implying that these fluctuations arise from discretization effects and do not correspond to tunneling events between topological sectors. In the second HMC stream, we see an abrupt change in the behavior of Q. This coincides with a change in σ, confirming that a true tunneling event has occurred. By contrast, flow-based sampling exhibits rapid fluctuation in both Q and σ, demonstrating sampling which rapidly mixes topological sectors.

A fair and comprehensive comparison of the costs of HMC and flow-based MCMC requires quantifying three factors for each: setup costs, the raw computational cost of a sampling step, and the sampling efficiency (i.e., the degree of autocorrelation). Setup costs—predominantly, equilibration for HMC and training for flows—are particularly difficult to compare in this case. Full equilibration of HMC requires observing and discarding many tunneling events, which occur stochastically, while training costs for the flow-based approach may vary over orders of magnitude depending on the training scheme. Raw computational costs may be measured directly, but depend strongly on implementation details. On the same GPU hardware, we find that flow-based MCMC steps are $\sim$10 times less expensive than HMC trajectories, due to the frequent inversions of the Dirac operator in HMC. However, there is room for optimization in both cases.

Nevertheless, disregarding setup and raw computational costs, an approximate comparison of sampling efficiency is sufficient to show the advantage of flow-based sampling over HMC. Each algorithm exhibits some characteristic time between tunneling events; a chain with many times that number of steps will be required to incorporate information from all topological sectors with appropriate weights. For these HMC parameters, we find tunneling events are separated by $\sim$20 k trajectories on average. Meanwhile, sampling with our flow model, the topological sector changes every $\sim$6 steps on average. Thus, for this model at the parameters investigated, we estimate that the advantage in sampling efficiency of flow-based MCMC over HMC is more than three orders of magnitude.

IV. CONCLUSION AND OUTLOOK

In this paper, we demonstrate for the first time that flow-based sampling can be applied to lattice gauge theories with fermion content at criticality. Specifically, we have developed an architecture that can successfully model long-range correlations in the Schwinger model at vanishing renormalized fermion mass. The resulting flow-based sampler does not suffer from topological freezing at these parameters and thus outperforms HMC by orders of magnitude. These results represent an important milestone in first-principles calculations in gauge field theories with fermions, such as QCD, using provably exact machine learning.

Importantly, the flow models developed in this work—and in particular, their features which allow long-range correlations to be modeled—may serve as an "engine" to improve a much broader class of sampling algorithms. For example, flow-based MCMC updates may be interleaved with steps of HMC [16] or other MCMC algorithms [57]. Such composite algorithms may provide improved sampling over either method alone. Furthermore, our technical advances can be adapted for more general MCMC schemes, e.g., generalizations of the HMC algorithm [17,18], stochastic normalizing flows [58–60], or hierarchical multilevel MCMC schemes [19].

Challenges remain on the road to large-scale applications, such as state-of-the-art QCD calculations. The sampling approach here employs exact evaluation of the fermion determinant, but more scalable approaches will be needed for larger volumes and theories in higher dimensions; extending the machine-learned stochastic determinant estimators of Ref. [14] to lattice gauge theories presents a promising opportunity. If the success demonstrated here for the Schwinger model can be extended to other theories, and in particular at scale, it will have widespread impact across nuclear and particle physics, as well as in condensed matter applications.

ACKNOWLEDGMENTS

We thank Ryan Abbott for useful comments on this manuscript. G. K., D. C. H., F. R.-L., and P. E. S. are supported in part by the U.S. Department of Energy, Office of Science, Office of Nuclear Physics, under Grant Contract No. DE-SC0011090. P. E. S. is additionally supported by the National Science Foundation under EAGER Grant No. 2035015, by the U.S. DOE Early Career Award No. DE-SC0021006, by an NEC Research Award, and by the Carl G. and Shirley Sontheimer Research Fund. G. K. is additionally supported by the Schweizerischer Nationalfonds. K. C. and M. S. A. are supported by the National Science Foundation under Award No. PHY-2141336. M. S. A. thanks the Flatiron Institute for their hospitality. D. B. is supported by the Argonne Leadership Computing Facility, which is a U.S.

Department of Energy Office of Science User Facility operated under Contract No. DE-AC02-06CH11357. This work is funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under Germany's Excellence Strategy Grant No. EXC 2181/1–390900948 (the Heidelberg STRUCTURES Excellence Cluster), the Collaborative Research Centre Grant No. SFB 1225 (ISOQUANT), and the U.S. National Science Foundation under Cooperative Agreement No. PHY-2019786 (The NSF AI Institute for Artificial Intelligence and Fundamental Interactions, [61]). This work is associated with an ALCF Aurora Early Science Program project, and used resources of the Argonne Leadership Computing Facility, which is a DOE Office of Science User Facility supported under Contract No. DEAC02-06CH11357. The authors acknowledge the MIT SuperCloud and Lincoln Laboratory Supercomputing Center [62] for providing HPC resources that have contributed to the research results reported within this paper. Numerical experiments and data analysis used PyTorch [63], HOROVOD [64], NumPy [65], and SciPy [66]. Figures were produced using MATPLOTLIB [67].

[1] S. Schaefer, R. Sommer, and F. Virotta (ALPHA Collaboration), Nucl. Phys. **B845**, 93 (2011).
[2] L. Del Debbio, H. Panagopoulos, and E. Vicari, J. High Energy Phys. 08 (2002) 044.
[3] M. Cè, M. García Vera, L. Giusti, and S. Schaefer, Phys. Lett. B **762**, 232 (2016).
[4] S. Duane, A. Kennedy, B. J. Pendleton, and D. Roweth, Phys. Lett. B **195**, 216 (1987).
[5] B. Alles, G. Boyd, M. D'Elia, A. Di Giacomo, and E. Vicari, Phys. Lett. B **389**, 107 (1996).
[6] L. Del Debbio, G. M. Manca, and E. Vicari, Phys. Lett. B **594**, 315 (2004).
[7] R. Brower, S. Chandrasekharan, J. W. Negele, and U. J. Wiese, Phys. Lett. B **560**, 64 (2003).
[8] S. Aoki, H. Fukaya, S. Hashimoto, and T. Onogi, Phys. Rev. D **76**, 054508 (2007).
[9] Y. Aoki *et al.* arXiv:2111.09849.
[10] M. S. Albergo, G. Kanwar, and P. E. Shanahan, Phys. Rev. D **100**, 034515 (2019).
[11] G. Kanwar, M. S. Albergo, D. Boyda, K. Cranmer, D. C. Hackett, S. Racanière, D. J. Rezende, and P. E. Shanahan, Phys. Rev. Lett. **125**, 121601 (2020).
[12] D. Boyda, G. Kanwar, S. Racanière, D. J. Rezende, M. S. Albergo, K. Cranmer, D. C. Hackett, and P. E. Shanahan, Phys. Rev. D **103**, 074504 (2021).
[13] K. A. Nicoli, C. J. Anders, L. Funcke, T. Hartung, K. Jansen, P. Kessel, S. Nakajima, and P. Stornati, Phys. Rev. Lett. **126**, 032001 (2021).
[14] M. S. Albergo, G. Kanwar, S. Racanière, D. J. Rezende, J. M. Urban, D. Boyda, K. Cranmer, D. C. Hackett, and P. E. Shanahan, Phys. Rev. D **104**, 114507 (2021).
[15] L. Del Debbio, J. M. Rossney, and M. Wilson, Phys. Rev. D **104**, 094507 (2021).
[16] D. C. Hackett, C.-C. Hsieh, M. S. Albergo, D. Boyda, J.-W. Chen, K.-F. Chen, K. Cranmer, G. Kanwar, and P. E. Shanahan, arXiv:2107.00734.
[17] S. Foreman, X.-Y. Jin, and J. C. Osborn, in *Proceedings of the 9th International Conference on Learning Representations* (2021), arXiv:2105.03418.
[18] S. Foreman, T. Izubuchi, L. Jin, X.-Y. Jin, J. C. Osborn, and A. Tomiya, Proc. Sci., LATTICE2021 (**2022**) 073 arXiv:2112.01586].
[19] J. Finkenrath, arXiv:2201.02216.
[20] D. J. Rezende and S. Mohamed, arXiv:1505.05770.
[21] L. Dinh, J. Sohl-Dickstein, and S. Bengio, arXiv:1605.08803.
[22] G. Papamakarios, E. Nalisnick, D. J. Rezende, S. Mohamed, and B. Lakshminarayanan, J. Mach. Learn. Res. **22**, 1 (2021).
[23] N. Metropolis, A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller, and E. Teller, J. Chem. Phys. **21**, 1087 (1953).
[24] L. Tierney, Ann. Stat. **22**, 1701 (1994).
[25] J. S. Schwinger, Phys. Rev. **128**, 2425 (1962).
[26] S. R. Coleman, R. Jackiw, and L. Susskind, Ann. Phys. (N.Y.) **93**, 267 (1975).
[27] J. Smit and J. C. Vink, Nucl. Phys. **B303**, 36 (1988).
[28] H. Dilger, Phys. Lett. B **294**, 263 (1992).
[29] H. Dilger, Int. J. Mod. Phys. C **06**, 123 (1995).
[30] S. Durr, Phys. Rev. D **85**, 114503 (2012).
[31] D. Albandea, P. Hernández, A. Ramos, and F. Romero-López, Eur. Phys. J. C **81**, 873 (2021).
[32] T. Hartung, K. Jansen, F. Y. Kuo, H. Leövey, D. Nuyens, and I. H. Sloan, Proc. Sci., LATTICE2021 (**2022**) 010 [arXiv:2112.05069].
[33] T. Eichhorn and C. Hoelbling, Proc. Sci., LATTICE2021 (**2022**) 573 [arXiv:2112.05188].
[34] L. Funcke, K. Jansen, and S. Kühn, Phys. Rev. D **101**, 054507 (2020).
[35] N. Butt, S. Catterall, Y. Meurice, R. Sakai, and J. Unmuth-Yockey, Phys. Rev. D **101**, 094509 (2020).
[36] M. C. Bañuls *et al.*, Eur. Phys. J. D **74**, 165 (2020).
[37] S. R. Coleman, Ann. Phys. (N.Y.) **101**, 239 (1976).
[38] A. V. Smilga, Phys. Lett. B **278**, 371 (1992).
[39] L. Giusti, G. C. Rossi, M. Testa, and G. Veneziano, Nucl. Phys. **B628**, 234 (2002).
[40] G. D'Amico, R. Gobbetti, M. Kleban, and M. Schillo, J. Cosmol. Astropart. Phys. 03 (**2013**) 004.
[41] Y. Shimizu and Y. Kuramashi, Phys. Rev. D **90**, 014508 (2014).
[42] Y. Shimizu and Y. Kuramashi, Phys. Rev. D **90**, 074503 (2014).
[43] C. Nagele, J. E. Cejudo, T. Byrnes, and M. Kleban, Phys. Rev. D **99**, 094501 (2019).
[44] C. Nagele, O. Janssen, and M. Kleban, arXiv:2010.04803.

[45] H. Georgi, Phys. Rev. Lett. **125**, 181601 (2020).
[46] X.-Y. Hu, M. Kleban, and C. Yu, J. High Energy Phys. 03 (2022) 197.
[47] C. R. Gattringer, I. Hip, and C. B. Lang, Nucl. Phys. **B508**, 329 (1997).
[48] I. Hip, C. B. Lang, and R. Teppner, Nucl. Phys. B, Proc. Suppl. **63**, 682 (1998).
[49] C. Czaban and M. Wagner, *Proc. Sci.*, LATTICE2013 (2013) 465 [arXiv:1310.5258].
[50] K. G. Wilson, Phys. Rev. D **10**, 2445 (1974).
[51] K. G. Wilson, in *Proceedings of the 13th International School of Subnuclear Physics: New Phenomena in Subnuclear Physics* (Springer New York, New York, NY, 1975).
[52] D. J. Rezende, G. Papamakarios, S. Racanière, M. S. Albergo, G. Kanwar, P. E. Shanahan, and K. Cranmer, arXiv:2002.02428.
[53] F. Yu and V. Koltun, arXiv:1511.07122.
[54] C. Durkan, A. Bekasov, I. Murray, and G. Papamakarios, in Advances in Neural Information Processing Systems (2019), pp. 7511–7522, arXiv:1906.04032.
[55] See Supplemental Material at http://link.aps.org/supplemental/10.1103/PhysRevD.106.014514 for details of architecture, training scheme, and HMC parameters.
[56] U. Wolff (ALPHA Collaboration), Comput. Phys. Commun. **156**, 143 (2004); **176**, 383(E) (2007).
[57] M. Gabrié, G. M. Rotskoff, and E. Vanden-Eijnden, Proc. Natl. Acad. Sci. U.S.A. **119**, e2109420119 (2022).
[58] A. G. D. G. Matthews, M. Arbel, D. J. Rezende, and A. Doucet, arXiv:2201.13117.
[59] H. Wu, J. Köhler, and F. Noé, Adv. Neural Inf. Process. Syst. **33**, 5933 (2020).
[60] D. Nielsen, P. Jaini, E. Hoogeboom, O. Winther, and M. Welling, Adv. Neural Inf. Process. Syst. **33**, 12685 (2020).
[61] http://iaifi.org/.
[62] A. Reuther, J. Kepner, C. Byun, S. Samsi, W. Arcand, D. Bestor, B. Bergeron, V. Gadepally, M. Houle, M. Hubbell *et al.*, *2018 IEEE High Performance extreme Computing Conference (HPEC)* (Curran Associates, Inc., Red Hook, NY, 2018), pp. 1–6.
[63] A. Paszke *et al.*, in *Advances in Neural Information Processing Systems 32*, edited by H. Wallach, H. Larochelle, A. Beygelzimer, F. d′Alché-Buc, E. Fox, and R. Garnett (Curran Associates, Inc., 2019), pp. 8024–8035.
[64] A. Sergeev and M. Del Balso, arXiv:1802.05799.
[65] C. R. Harris, K. J. Millman, S. J. Van Der Walt, R. Gommers, P. Virtanen, D. Cournapeau, E. Wieser, J. Taylor, S. Berg, N. J. Smith *et al.*, Nature (London) **585**, 357 (2020).
[66] P. Virtanen, R. Gommers, T. E. Oliphant, M. Haberland, T. Reddy, D. Cournapeau, E. Burovski, P. Peterson, W. Weckesser, J. Bright *et al.*, Nat. Methods **17**, 261 (2020).
[67] J. D. Hunter, Comput. Sci. Eng. **9**, 90 (2007).