

Equivariant Flow-Based Sampling for Lattice Gauge Theory

Gurtej Kanwar¹, Michael S. Albergo², Denis Boyda¹, Kyle Cranmer², Daniel C. Hackett¹,
Sébastien Racanière³, Danilo Jimenez Rezende³, and Phiala E. Shanahan¹

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](https://doi.org/10.1103/PhysRevLett.125.121601)

Many important physical theories are described by Lagrangians that are invariant under local symmetry transformations that form Lie groups; such theories are named gauge theories. For example, the standard model of particle physics, which is our most accurate description of nature at the shortest length scales, is a quantum field theory centered around the action of three gauge groups [1–4], and several important condensed matter systems can be described by effective gauge theories [5–8]. In the strong-coupling limit, these theories are nonperturbative, and numerical formulations on discrete spacetime lattices offer the only known way to compute properties of interest from first principles.

Calculations within lattice frameworks typically proceed by estimating expectation values of observables using Markov chain Monte Carlo (MCMC) to sample from thermodynamic distributions or Euclidean-time path integrals. In both cases, samples U (typically high dimensional) are drawn from an exponentially weighted distribution $p(U) = e^{-S(U)}/Z$, where the physics is encoded in an energy or action functional $S(U)$, and the normalizing constant Z is unknown. When MCMC sampling from the distribution $p(U)$ is efficient, precise physical predictions can be made from the theory. However, as the model parameters are tuned towards criticality, e.g., to describe universal properties of condensed matter theories or to access the continuum or large- N limits of quantum field theories, critical slowing-down (CSD) can cause the computational cost of sampling to diverge [9,10].

Specialized approaches have been developed to avoid CSD for specific theories [11–20]. For several theories of interest, however, CSD obstructs calculations. This is true, in particular, for the lattice formulation of quantum chromodynamics (QCD) [21–23], which enables calculations of nonperturbative phenomena arising from the standard model of particle physics. Recently, there has been progress in the development of flow-based generative models which can be trained to directly produce samples from a given probability distribution; early success has been demonstrated in theories of bosonic matter, spin systems, molecular systems, and for Brownian motion [24–34]. This progress builds on the great success of flow-based approaches for image, text, and structured object generation [35–42], as well as non-flow-based machine learning techniques applied to sampling for physics [43–48]. If flow-based algorithms can be designed and implemented at the scale of state-of-the-art calculations, they would enable efficient sampling in lattice theories that are currently hindered by CSD.

In this Letter, we develop a provably correct flow-based sampling algorithm designed for lattice gauge theories with continuous gauge groups. Our approach is immediately applicable to U(1) gauge theory and can be applied to non-Abelian gauge theories, such as lattice QCD, upon development of bijective maps acting on single elements of the gauge group. Here, we demonstrate the application of this approach to U(1) gauge theory in two spacetime dimensions. In this theory, the field strength tensor fluctuates independently on each lattice site, and therefore, there is no critical limit corresponding to infinite correlation length; however, in the limit of small bare coupling, topological freezing results in severe difficulties when general sampling methods are applied. Hybrid Monte Carlo (HMC) [49] and heat bath (HB) [50–52] are two widely used MCMC approaches that are applicable to pure gauge

Published by the American Physical Society under the terms of the [Creative Commons Attribution 4.0 International license](https://creativecommons.org/licenses/by/4.0/). Further distribution of this work must maintain attribution to the author(s) and the published article's title, journal citation, and DOI. Funded by SCOAP³.

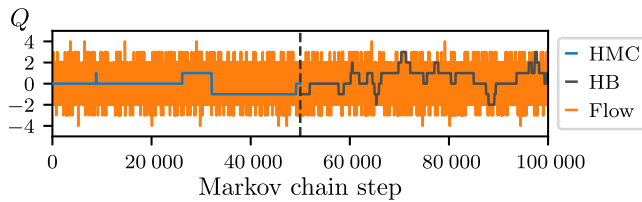


FIG. 1. Standard approaches (HMC and HB) to MCMC sampling for U(1) gauge theory explore the distribution of topological charge Q very slowly compared with the flow-based approach introduced here. Results are shown for coupling $\beta = 7$ on a 16×16 lattice, see Eq. (6). The first (second) half of the Markov chain history is displayed for HMC (HB).

theories with continuous gauge group (HMC is also applicable to a much broader class of theories, including gauge theories with fermion content). Both methods explore topological sectors very slowly, as seen in the Markov chain histories of topological charge depicted in Fig. 1. For 2D U(1) gauge theory, in particular, there are also a number of specialized calculation techniques that do not suffer from topological freezing, including a cluster algorithm [53], complex Langevin [54], dual variable approaches [55–57], and direct integration of the path integral [57]. We use the latter to check the accuracy of the numerical methods under study. In contrast, our flow-based algorithm is generally applicable to pure-gauge theories with continuous gauge groups; yet it produces independent samples of field configurations with appropriate frequency from each topological sector, enabling far more accurate estimation of topological quantities at a given computational cost when compared with similarly general methods. Critical to the success of this approach is enforcing exact gauge symmetry in the flow-based distribution: when the symmetry is enforced, we can successfully train flow-based models at a range of parameters approaching topologically frozen, while without it, models of similar scale fail to learn the distributions under the same training procedure.

Flow-based sampling.—Flow-based generative models allow sampling from an expressive family of distributions by applying an invertible function f to samples U from a fixed, simple prior distribution defined by a density $r(U)$ [58]. The resulting samples $U' = f(U)$ are distributed according to a model density $q(U')$. The invertible function is constructed specifically to allow efficient evaluation of the Jacobian factor for any given sample, so that the associated normalized probability density,

$$q(U') = q[f(U)] = r(U) \left| \det \frac{\partial f(U)}{\partial U} \right|^{-1}, \quad (1)$$

is returned with each sample drawn. This feature enables training the flow model, i.e., optimizing the function f , by minimizing the distance between the model probability density $q(U')$ and the desired density $p(U')$ using a chosen

metric. Any deviation from the true distribution due to an imperfect model can be corrected by a number of techniques; in this Letter, we apply independence Metropolis sampling [26]. (Reweighted observables can also be used [59,60]. This is efficient when measurements of the action are more costly than measurements of observables.)

A powerful approach to defining a flexible invertible function f is through composition of several coupling layers, $f := g_m \circ \dots \circ g_1$. Coupling layers act on samples U by applying an analytically invertible transformation (such as a scaling) to a subset of the components $U^A := \{U^i : i \in A\}$, where the superscript i indexes components of the multi-dimensional sample U and the set A indicates the components that are transformed. The remaining (unmodified) components U^B , defined by $U^B = U \setminus U^A$, are given as input to a feed-forward neural network that parametrizes the transformation. This variable splitting guarantees invertibility despite the use of noninvertible feed-forward networks.

Gauge-invariant flows.—Lattice gauge theories can be defined in terms of one gauge variable $U_\mu(x)$ per nearest-neighbor link $(x, x + \hat{\mu})$ of the lattice. Thus, samples live in the compact manifold $G^{N_d V}$, where G is the manifold of the gauge group, N_d is the spacetime dimension, and V is the lattice volume. The physical distribution $p(U)$ is exactly invariant under a discrete translational symmetry group with V elements and a continuous V -dimensional gauge symmetry group, meaning that the density associated with any transformed field configuration \tilde{U} is identical to that of the untransformed configuration, $p(\tilde{U}) = p(U)$. Under a gauge transformation, links $U_\mu(x)$ are transformed by a group-valued field $\Omega(x)$ as

$$U_\mu(x) \rightarrow \tilde{U}_\mu(x) = \Omega(x) U_\mu(x) \Omega^\dagger(x + \hat{\mu}). \quad (2)$$

In the flow-based approach, symmetries correspond to flat directions of the probability density that must be reproduced by the model. Exactly encoding symmetries in machine learning models can improve training and model quality compared with learning the symmetries over the course of training [25,28,61–65]. The incorporation of translational symmetries into models is possible using convolutional architectures, as studied for example, in Ref. [61]. To address gauge symmetry, one could attempt to use a gauge-fixing procedure to select a single configuration from each gauge-equivalent class, leaving only physical degrees of freedom; however, the only known gauge fixing procedures that preserve translational invariance are based on implicit differential equation constraints [66], which do not have a straightforward implementation in flows. Here, instead, we introduce a method to preserve exact gauge invariance in flow-based models.

When a flow-based model is defined in terms of coupling layers, its output distribution will be invariant under a symmetry group if two conditions are met: (1) The prior

distribution is symmetric; (2) Each coupling layer is equivariant under the symmetry, i.e., all transformations commute through application of the coupling layer [61,63,67–69].

Choosing a prior distribution that is symmetric is typically straightforward, for example, a uniform distribution with respect to the Haar measure over gauge links is both translationally and gauge invariant. Using gauge-equivariant coupling layers with such a prior distribution then defines a gauge-invariant flow-based model.

Gauge-equivariant coupling layers.—We construct an explicitly gauge-equivariant and invertible coupling layer $g: G^{N_d V} \rightarrow G^{N_d V}$ by splitting the input variables into subsets U^A and U^B . In terms of these subsets, we define the action of the coupling layer to be $g(U^A, U^B) = (U^A, U^B)$, where link $U^i \in U^A$ is mapped to

$$U^i \rightarrow h(U^i S^i | I^i) S^{i\dagger}, \quad (3)$$

in which $h: G \rightarrow G$ is an invertible kernel which is explicitly parametrized by a set of gauge-invariant quantities I^i constructed from the elements of U^B . Here, S^i is a product of links such that $U^i S^i$ forms a loop that starts and ends at a common point x , and, therefore, transforms under the gauge symmetry to $\Omega(x) U^i \Omega^\dagger(x + \hat{\mu}) \Omega(x + \hat{\mu}) S^i \Omega^\dagger(x) = \Omega(x) U^i S^i \Omega^\dagger(x)$. With this definition, the coupling layer is gauge equivariant if the kernel satisfies

$$h(XWX^\dagger) = Xh(W)X^\dagger, \quad \forall X, W \in G, \quad (4)$$

which implies that $U^i \rightarrow \tilde{U}^i$ transforms according to Eq. (2),

$$\begin{aligned} \tilde{U}^i &= h[\Omega(x) U^i S^i \Omega^\dagger(x) | I^i] \Omega(x) S^{i\dagger} \Omega^\dagger(x + \hat{\mu}) \\ &= \Omega(x) U^i \Omega^\dagger(x + \hat{\mu}). \end{aligned} \quad (5)$$

To ensure invertibility, the product of links S^i must not contain any links in U^A .

For an Abelian group, the transformation property in Eq. (4) is trivially satisfied by any kernel. In the $U(1)$ gauge theory considered below, therefore, we define the kernel using invertible flows parametrized by neural networks. For non-Abelian theories, it has been shown that it is possible to construct invertible functions on spheres [70] and surjective functions on general Lie groups [71]. If these approaches can be generalized to produce invertible functions with convergent power expansions, they will satisfy the necessary kernel transformation property, since $h(XWX^\dagger) = \sum_n \alpha_n (XWX^\dagger)^n = Xh(W)X^\dagger$.

An example of a variable splitting suitable for both Abelian and non-Abelian gauge theories is given by the pattern depicted in Fig. 2. In this example, the set of updated links U^A consists of vertical links spaced by four sites, and the products $U^i S^i$ are 1×1 loops adjacent to each U^i . This is sufficiently sparse such that every S^i is

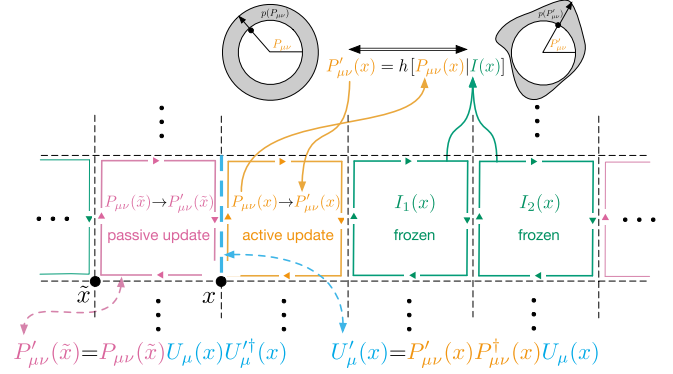


FIG. 2. An example of a variable splitting based on a tiled 4×1 pattern with an actively updated link $U^i \equiv U_\mu(x)$ and 1×1 loop $U^i S^i \equiv P_{\mu\nu}(x)$ located at x , a passively updated 1×1 loop at $\tilde{x} = x - \hat{\nu}$, and two frozen traced 1×1 loops at $x + \hat{\nu}$ and $x + 2\hat{\nu}$ included in the set I^i .

independent of all updated links in U^A , and a nontrivial set of invariants I^i (e.g., all traced 1×1 loops that are not adjacent to updated links) can be constructed to parametrize the transformation. Composing coupling layers using rotations and offsets of the pattern allows all links to be updated. (For example, composing eight such layers is sufficient to update all links in 2D.)

Using gauge-equivariant coupling layers constructed in terms of kernels generalizes the “trivializing map” proposed in Ref. [72]. There, repeatedly applying a specific kernel based on gradients of the action theoretically trivializes a gauge theory, i.e., maps the Euclidean time distribution to a uniform one. The family of gauge equivariant flows defined here includes the trivializing map in the limit of a large number of coupling layers and arbitrarily expressive kernel, indicating that, in this limiting case, exact sampling as described in Ref. [72] is possible. However, the approach presented here allows for more general and inexpensive parametrizations of h . These can be optimized to produce flows that similarly trivialize the theory and which may have a lower cost of evaluation than implementations of the analytical trivializing map [73].

Application to $U(1)$ gauge theory.—Gauge theory with a $U(1)$ gauge group defined in two spacetime dimensions is the quenched limit of $1 + 1D$ electrodynamics, i.e., the Schwinger model [74]. The full Schwinger model reproduces many features of quantum chromodynamics (confinement, an axial anomaly, topology, and chiral symmetry breaking) while being analytically tractable. Even in the quenched limit, the well-defined gauge field topology results in severe slowing-down of MCMC methods for sampling lattice discretizations of the model as the coupling is taken to criticality. We consider the lattice discretization given by the Wilson gauge action [75]

$$S(U) := -\beta \sum_x \text{Re} P(x), \quad (6)$$

where $P(x)$ is the plaquette at x defined in terms of link variables $U_\mu(x) \in U(1)$

$$P(x) := U_0(x)U_1(x + \hat{0})U_0^\dagger(x + \hat{1})U_1^\dagger(x), \quad (7)$$

and $x = (x_0, x_1)$ runs over coordinates in an $L \times L$ square lattice with periodic boundary conditions. Physical information may be extracted from the model by considering expectation values of observables \mathcal{O} under the Euclidean time path integral

$$\langle \mathcal{O} \rangle := \frac{1}{Z} \int \mathcal{D}U \mathcal{O}(U) e^{-S(U)}, \quad (8)$$

where $\int \mathcal{D}U$ denotes integration over the product of Haar measures for each link, and $Z = \int \mathcal{D}U e^{-S(U)}$. In this study, three key observables were considered: (1) Expectation values of powers of plaquettes; (2) Expectation values of $\ell \times \ell$ Wilson loops $W_{\ell \times \ell} = \prod_{x \in \ell \times \ell} P(x)$; (3) Topological susceptibility $\chi_Q = \langle Q^2/V \rangle$, where topological charge $Q := (1/2\pi) \sum_x \arg[P(x)]$ is defined in terms of plaquette phase in the principal interval, $\arg[P(x)] \in [-\pi, \pi]$.

To investigate slowing-down due to topological freezing, we studied the theory at a fixed lattice size, $L = 16$, using seven choices of the parameter $\beta = \{1, 2, 3, 4, 5, 6, 7\}$; the topology is increasingly frozen as $\beta \rightarrow \infty$. For each parameter choice, we trained gauge invariant flow-based models using a uniform prior distribution and a composition of 24 gauge-equivariant coupling layers. The kernels h were chosen to be mixtures of noncompact projections [70], which are suitable for $U(1)$ group elements; in particular, we used six components for each mixture and parametrized each transformation with a convolutional neural network. The model architecture was held fixed across all choices of β , ensuring identical cost to draw samples for each parameter choice. To train the models, we minimized the Kullback-Leibler divergence between the model density $q(U)$ and the target density $e^{-S(U)}/Z$. Training was halted when the loss function reached a plateau. For this proof-of-principle study, we did not perform extensive optimization over the variable splitting pattern, neural network architecture, or training hyperparameters, and it is likely that better models can be trained, for example, using automatic hyperparameter and architecture searches [76].

After training, the flow-based models were used to generate proposals for an independence Metropolis Markov chain [26], producing ensembles of 100000 samples each. For comparison, ensembles of identical size were produced using the HMC and heat bath algorithms. For all choices of β , we fixed the HMC trajectory length to achieve $> 80\%$ acceptance rate when using a leapfrog integrator with five steps. Each HB step was defined as one sweep, i.e., a single update of every link. To within 10%, the computational cost per HMC trajectory was equal to the cost per proposal from the flow-based model in a

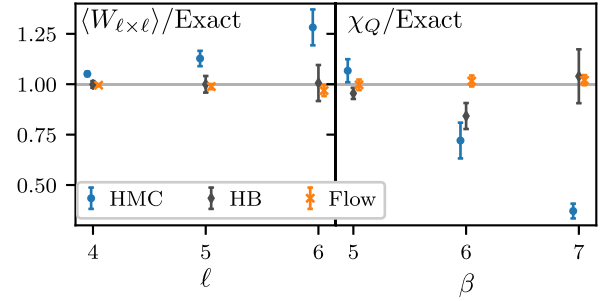


FIG. 3. Left: estimates of average Wilson loops $\langle W_{\ell \times \ell} \rangle$ measured on the most frozen ensemble studied here ($\beta = 7$). Right: estimates of topological susceptibility measured on the three most frozen ensembles studied here ($\beta = 5, 6, 7$). All values are plotted as ratios to the exact results. The flow-based estimates are consistent with the exact values, while the HMC and heat bath estimates have larger and underestimated uncertainties due to long autocorrelation times on the order of the Markov chain length or longer.

single-threaded CPU environment, while the cost per heat bath step was half that of HMC or flow.

Using samples from a flow-based model as proposals within a Markov chain ensures unbiased estimates after thermalization; at the finite ensemble size used here, all observables were found to agree with analytical results within statistical uncertainties. Of the observables we studied, local quantities like powers of plaquettes and expectation values of small Wilson loops were estimated up to two times as precisely by HMC and HB than by the flow-based algorithm. However, Fig. 3 shows that, for observables with larger extent such as $W_{\ell \times \ell}$ with $\ell \geq 4$, and particularly for χ_Q , large autocorrelations in the HMC and HB samples, in some cases on the order of the Markov chain length or longer, result in estimates that have lower precision than the flow-based estimates and have underestimated uncertainties, despite accounting for measured autocorrelations.

For Markov chain methods, the characteristic length of autocorrelations for an observable \mathcal{O} can be defined by the integrated autocorrelation time $\tau_{\mathcal{O}}^{\text{int}}$ [77]. Figure 4 compares $\tau_{\mathcal{O}}^{\text{int}}$ for HMC and HB to that in the flow-based algorithm as an indicator of how well the three methods explore the distribution of topological charge. For all three methods, $\tau_{\mathcal{O}}^{\text{int}}$ grows as β is increased. However, this problem is far less severe for the flow-based algorithm than for HMC or HB. For example, the autocorrelation time in the flow-based algorithm is approximately ten at the largest value of β , whereas $\tau_{\mathcal{O}}^{\text{int}} \approx 4000$ for HB and $\tau_{\mathcal{O}}^{\text{int}} \approx 15000$ for HMC. Accounting for the relative cost per step of each Markov chain, the flow-based Metropolis sampler is, therefore, roughly 1500 times more efficient than HMC and 200 times more efficient than heat bath in determining topological quantities. A promising possibility for further development is mixing flow-based Markov chain steps with HMC

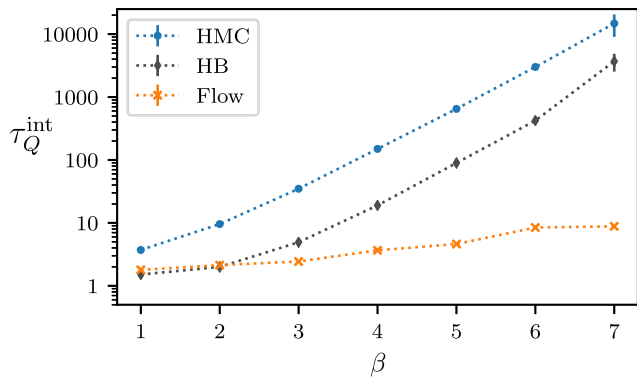


FIG. 4. Integrated autocorrelation time for the topological charge, τ_Q^{int} , measured on ensembles of 16×16 lattices generated using HMC, heat bath, and the flow-based algorithm. Ten replicas of each ensemble were used to estimate errors, which are smaller than the plot markers for most points.

trajectories or heat bath sweeps to gain the benefits of standard Markov chain steps for local observables and of the flow-based algorithm for extended and topological observables.

Summary.—Critical slowing-down of sampling in lattice gauge theories is an obstacle to precisely estimating quantities of physical interest as critical limits of the theories are approached. Topological freezing is a particularly severe slowing-down preventing estimation of quantities coupled to gauge field topology in many lattice gauge theories. Flow-based models enable direct sampling from an approximation to the distribution of interest, from which estimates of physical observables can be derived that are exact in the infinite-statistics limit. Here, we introduce flow-based models constructed to satisfy exact gauge invariance and find that applying this approach to a two-dimensional Abelian gauge theory enables more efficient estimation of topological quantities than the two other algorithms that are applicable to general lattice gauge theories, HMC and heat bath.

While the numerical investigations in this Letter demonstrate the efficiency of the proposed sampling approach in a region of parameter space where HMC and heat bath suffer significantly from topological freezing, it remains to be investigated how the algorithm scales in the thermodynamic and continuum limits. In the thermodynamic limit, one naively expects an exponential degradation of the independence Metropolis acceptance rate with volume (given a fixed model); the experimental question will be how the model complexity must be scaled to maintain the acceptance rate. In the continuum limit, it will be important to efficiently construct correlations over large length scales, which may be enabled by architectures such as hierarchical models [78,79] and dilated convolutions [80].

The approach presented here is generally applicable to gauge theories defined by Lie groups, including non-Abelian theories such as QCD. To extend this method to

such theories, expressive invertible functions must be defined as the kernels of gauge equivariant coupling layers. There are several possible avenues forward; for example, Ref. [70] defines flows on spherical manifolds and Ref. [71] defines surjective (though not bijective) maps on Lie groups, both using neural network parametrizations. Future work will explore constructing kernels based on generalizations of these methods and, thus, producing gauge invariant flows for non-Abelian theories like QCD.

The authors thank Jiunn-Wei Chen, Will Detmold, and Andrew Pochinsky for helpful discussions. G. K., D. B., D. C. H., 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 CAREER Grant No. 1841699. K. C. is supported by the National Science Foundation under Grants No. ACI-1450310, No. OAC-1836650, and No. OAC-1841471 and by the Moore-Sloan data science environment at NYU. M. S. A. is supported by the Carl Feinberg Fellowship in Theoretical Physics. 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. DE-AC02-06CH11357.

-
- [1] S. Weinberg, *Phys. Rev. Lett.* **19**, 1264 (1967).
 - [2] A. Salam, *Conf. Proc.* **C680519**, 367 (1968).
 - [3] O. W. Greenberg, *Phys. Rev. Lett.* **13**, 598 (1964).
 - [4] M. Y. Han and Y. Nambu, *Phys. Rev.* **139**, B1006 (1965).
 - [5] J. B. Kogut, *Rev. Mod. Phys.* **51**, 659 (1979).
 - [6] N. Seiberg, T. Senthil, C. Wang, and E. Witten, *Ann. Phys. (Amsterdam)* **374**, 395 (2016).
 - [7] S. Sachdev, *Rep. Prog. Phys.* **82**, 014001 (2019).
 - [8] H. Guo, R. Samajdar, M. S. Scheurer, and S. Sachdev, *Phys. Rev. B* **101**, 195126 (2020).
 - [9] U. Wolff, *Nucl. Phys. B, Proc. Suppl.* **17**, 93 (1990).
 - [10] C. Bonanno, C. Bonati, and M. D’Elia, *J. High Energy Phys.* **01** (2019) 003.
 - [11] D. Kandel, E. Domany, D. Ron, A. Brandt, and E. Loh, *Phys. Rev. Lett.* **60**, 1591 (1988).
 - [12] C. Bonati and M. D’Elia, *Phys. Rev. E* **98**, 013308 (2018).
 - [13] U. Wolff, *Phys. Rev. Lett.* **62**, 361 (1989).
 - [14] M. Hasenbusch, *Phys. Rev. D* **96**, 054504 (2017).
 - [15] M. Hasenbusch and S. Schaefer, *Phys. Rev. D* **98**, 054502 (2018).
 - [16] R. H. Swendsen and J.-S. Wang, *Phys. Rev. Lett.* **58**, 86 (1987).
 - [17] N. Prokof’ev and B. Svistunov, *Phys. Rev. Lett.* **87**, 160601 (2001).
 - [18] N. Kawashima and K. Harada, *J. Phys. Soc. Jpn.* **73**, 1379 (2004).
 - [19] W. Bietenholz, A. Pochinsky, and U.-J. Wiese, *Phys. Rev. Lett.* **75**, 4524 (1995).
 - [20] M. Lüscher and S. Schaefer, *J. High Energy Phys.* **07** (2011) 036.

- [21] Y. Aoki, T. Blum, N. H. Christ, C. Dawson, T. Izubuchi, R. D. Mawhinney, J. Noaki, S. Ohta, K. Orginos, A. Soni, and N. Yamada, *Phys. Rev. D* **73**, 094507 (2006).
- [22] S. Schaefer, R. Sommer, and F. Virotta, *Proc. Sci. LAT2009* (2009) 032.
- [23] G. McGlynn and R. D. Mawhinney, *Phys. Rev. D* **90**, 074502 (2014).
- [24] S.-H. Li and L. Wang, *Phys. Rev. Lett.* **121**, 260601 (2018).
- [25] L. Zhang, E. Weinan, and L. Wang, [arXiv:1809.10188](https://arxiv.org/abs/1809.10188).
- [26] M. Albergo, G. Kanwar, and P. Shanahan, *Phys. Rev. D* **100**, 034515 (2019).
- [27] G. S. Hartnett and M. Mohseni, [arXiv:2001.00585](https://arxiv.org/abs/2001.00585).
- [28] J. Köhler, L. Klein, and F. Noé, [arXiv:1910.00753](https://arxiv.org/abs/1910.00753).
- [29] F. Noé, S. Olsson, J. Köhler, and H. Wu, *Science* **365**, eaaw1147 (2019).
- [30] S.-H. Li, C.-X. Dong, L. Zhang, and L. Wang, *Phys. Rev. X* **10**, 021020 (2020).
- [31] K. Madhawa, K. Ishiguro, K. Nakago, and M. Abe, [arXiv:1905.11600](https://arxiv.org/abs/1905.11600).
- [32] J. Liu, A. Kumar, J. Ba, J. Kiros, and K. Swersky, [arXiv:1905.13177](https://arxiv.org/abs/1905.13177).
- [33] C. Shi, M. Xu, Z. Zhu, W. Zhang, M. Zhang, and J. Tang, [arXiv:2001.09382](https://arxiv.org/abs/2001.09382).
- [34] G.-J. Both and R. Kusters, [arXiv:1912.09092](https://arxiv.org/abs/1912.09092).
- [35] L. Dinh, D. Krueger, and Y. Bengio, [arXiv:1410.8516](https://arxiv.org/abs/1410.8516).
- [36] L. Dinh, J. Sohl-Dickstein, and S. Bengio, [arXiv:1605.08803](https://arxiv.org/abs/1605.08803).
- [37] D. P. Kingma and P. Dhariwal, [arXiv:1807.03039](https://arxiv.org/abs/1807.03039).
- [38] J. Ho, X. Chen, A. Srinivas, Y. Duan, and P. Abbeel, [arXiv:1902.00275](https://arxiv.org/abs/1902.00275).
- [39] D. Tran, K. Vafa, K. K. Agrawal, L. Dinh, and B. Poole, [arXiv:1905.10347](https://arxiv.org/abs/1905.10347).
- [40] Z. M. Ziegler and A. M. Rush, [arXiv:1901.10548](https://arxiv.org/abs/1901.10548).
- [41] M. Kumar, M. Babaeizadeh, D. Erhan, C. Finn, S. Levine, L. Dinh, and D. Kingma, [arXiv:1903.01434](https://arxiv.org/abs/1903.01434).
- [42] G. Yang, X. Huang, Z. Hao, M.-Y. Liu, S. Belongie, and B. Hariharan, [arXiv:1906.12320](https://arxiv.org/abs/1906.12320).
- [43] L. Huang and L. Wang, *Phys. Rev. B* **95**, 035105 (2017).
- [44] L. Wang, *Phys. Rev. E* **96**, 051301(R) (2017).
- [45] K. Zhou, G. Endrődi, L.-G. Pang, and H. Stöcker, *Phys. Rev. D* **100**, 011501(R) (2019).
- [46] D. Wu, L. Wang, and P. Zhang, *Phys. Rev. Lett.* **122**, 080602 (2019).
- [47] J. Zhang, Y. I. Yang, and F. Noé, *J. Phys. Chem. Lett.* **10**, 5791 (2019).
- [48] J. Carrasquilla, G. Torlai, R. G. Melko, and L. Aolita, *Nat. Mach. Intell.* **1**, 155 (2019).
- [49] S. Duane, A. Kennedy, B. J. Pendleton, and D. Roweth, *Phys. Lett. B* **195**, 216 (1987).
- [50] M. Creutz, *Phys. Rev. Lett.* **43**, 553 (1979); **43**, 890(E) (1979).
- [51] N. Cabibbo and E. Marinari, *Phys. Lett.* **119B**, 387 (1982).
- [52] A. D. Kennedy and B. J. Pendleton, *Phys. Lett.* **156B**, 393 (1985).
- [53] R. Sinclair, *Phys. Rev. D* **45**, 2098 (1992).
- [54] J. Ambjorn, M. Flensburg, and C. Peterson, *Phys. Lett.* **159B**, 335 (1985).
- [55] R. Savit, *Rev. Mod. Phys.* **52**, 453 (1980).
- [56] V. Azcoiti, E. Follana, A. Vaquero, and G. Di Carlo, *J. High Energy Phys.* **08** (2009) 008.
- [57] C. Gattringer, T. Kloiber, and M. Müller-Preussker, *Phys. Rev. D* **92**, 114508 (2015).
- [58] G. Papamakarios, E. Nalisnick, D. Jimenez Rezende, S. Mohamed, and B. Lakshminarayanan, [arXiv:1912.02762](https://arxiv.org/abs/1912.02762).
- [59] A. M. Ferrenberg and R. H. Swendsen, *Phys. Rev. Lett.* **61**, 2635 (1988).
- [60] T. Müller, B. McWilliams, F. Rousselle, M. Gross, and J. Novák, [arXiv:1808.03856](https://arxiv.org/abs/1808.03856).
- [61] T. Cohen and M. Welling, *Proc. Mach. Learn. Res.* **48**, 2990 (2016).
- [62] T. Cohen, M. Weiler, B. Kicanaoglu, and M. Welling, *Proc. Mach. Learn. Res.* **97**, 1321 (2019).
- [63] D. Jimenez Rezende, S. Racanière, I. Higgins, and P. Toth, [arXiv:1909.13739](https://arxiv.org/abs/1909.13739).
- [64] P. Wirsberger, A. J. Ballard, G. Papamakarios, S. Abercrombie, S. Racanière, A. Pritzel, D. Jimenez Rezende, and C. Blundell, [arXiv:2002.04913](https://arxiv.org/abs/2002.04913).
- [65] M. Finzi, S. Stanton, P. Izmailov, and A. G. Wilson, [arXiv:2002.12880](https://arxiv.org/abs/2002.12880).
- [66] T. DeGrand and C. E. Detar, *Lattice Methods for Quantum Chromodynamics* (World Scientific, Singapore, 2006).
- [67] G. E. Hinton, A. Krizhevsky, and S. D. Wang, in *Artificial Neural Networks and Machine Learning—ICANN 2011*, edited by T. Honkela, W. Duch, M. Girolami, and S. Kaski (Springer, Berlin, Heidelberg, 2011), pp. 44–51.
- [68] J. J. Kivinen and C. K. Williams, in *International Conference on Artificial Neural Networks* (Springer, New York, 2011), pp. 1–9.
- [69] U. Schmidt and S. Roth, in *2012 IEEE Conference on Computer Vision and Pattern Recognition* (IEEE, New York, 2012), pp. 2050–2057.
- [70] D. J. Rezende, G. Papamakarios, S. Racanière, M. S. Albergo, G. Kanwar, P. E. Shanahan, and K. Cranmer, [arXiv:2002.02428](https://arxiv.org/abs/2002.02428).
- [71] L. Falorsi, P. de Haan, T. R. Davidson, and P. Forré, *Proc. Mach. Learn. Res.* **89**, 3244 (2019).
- [72] M. Lüscher, *Commun. Math. Phys.* **293**, 899 (2009).
- [73] G. P. Engel and S. Schaefer, *Comput. Phys. Commun.* **182**, 2107 (2011).
- [74] J. S. Schwinger, *Phys. Rev.* **128**, 2425 (1962).
- [75] K. G. Wilson, *Phys. Rev. D* **10**, 2445 (1974).
- [76] P. Balaprakash, M. Salim, T. D. Uram, V. Vishwanath, and S. M. Wild, in *2018 IEEE 25th International Conference on High Performance Computing (HiPC)* (Conference Publishing Services, New Jersey, 2018), pp. 42–51.
- [77] N. Madras and A. D. Sokal, *J. Stat. Phys.* **50**, 109 (1988).
- [78] L. Dinh, J. Sohl-Dickstein, and S. Bengio, [arXiv:1605.08803](https://arxiv.org/abs/1605.08803).
- [79] A. Bhattacharyya, S. Mahajan, M. Fritz, B. Schiele, and S. Roth, in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (IEEE Xplore, New Jersey, 2020), pp. 8415–8424.
- [80] F. Yu and V. Koltun, [arXiv:1511.07122](https://arxiv.org/abs/1511.07122).