

Spectral Diffusion for Sampling on $SU(N)$

Surtej Kanwar^a and Octavio Vega^{b,*}

^a*Higgs Centre for Theoretical Physics, University of Edinburgh, Edinburgh EH9 3FD, UK*

^b*Illinois Center for Advanced Studies of the Universe and Department of Physics,
University of Illinois Urbana-Champaign, Urbana, IL 61801, USA*

E-mail: octavio5@illinois.edu

Although ensemble generation remains a central challenge in lattice field theory simulations, recent advances in generative modeling may offer a path to accelerated sampling in these contexts. In this work, we implement a framework for efficiently training diffusion models acting on $SU(N)$ degrees of freedom, adapting the traditional score matching technique to the group manifold. We demonstrate that our models can effectively reproduce several target densities, resulting in precise unbiased expectation values. These results mark a step for diffusion models towards modeling full $SU(N)$ lattice field theories, including lattice Quantum Chromodynamics.

*The 42nd International Symposium on Lattice Field Theory (LATTICE2025)
2-8 November 2025
Tata Institute of Fundamental Research, Mumbai, India*

*Speaker

1. Introduction

The first step of any lattice field theory simulation is ensemble generation, which requires accurate sampling of physical probability densities in high dimensions. Traditional approaches, such as Hybrid Monte Carlo [1], are afflicted by topological freezing [2, 3] and critical slowing down [4], which become more pronounced with decreasing lattice spacing and increasing dimensionality. Recently, several advancements in machine learning have offered methods for using deep generative models to efficiently produce samples from arbitrarily complicated probability densities. Among these, score-based diffusion models [5–8] have shown promising impacts in image generation, drug discovery, and, recently, lattice field theory [9–11]. This work extends the formulation of score-based diffusion models to $SU(N)$ degrees of freedom, laying the groundwork for their application to lattice gauge theories, such as lattice Quantum Chromodynamics. While previous works have demonstrated a general construction of Riemannian manifold diffusion models [12, 13], here we resolve several practical questions required to implement these models for $SU(N)$. A software package containing this implementation is available at github.com/ovega14/sun_diffusion.

The central idea of score-based diffusion models is connecting the target probability distribution p_0 to a trivial distribution by incrementally applying Brownian noise. This *diffusion process* induces a time-dependent sequence of distributions p_t according to the associated Fokker-Planck equation; in practice a finite diffusion time T is sufficient for p_T to very closely approximate the equilibrium Gaussian (for non-compact variables) or uniform (for compact variables) distribution. The target distribution can be efficiently sampled using the reverse transport process, given by an ordinary differential equation written in terms of the *score function* $\nabla \log p_t$. Because the score function is related to the unknown density p_t , score-based diffusion models instead optimize a machine learning approximation of the true score, $s_t \approx \nabla \log p_t$. Regardless, exact unbiased estimates can be obtained by constructing appropriate reweighting factors.¹

2. Diffusion and the Heat Equation

A fundamental step in training score-matching diffusion models is drawing samples over the range of diffusion times t to be studied. Efficient training is possible when these samples can be drawn directly, without simulating the diffusion process.

2.1 Euclidean Diffusion

Mathematically, the diffusion and denoising processes are formulated through stochastic differential equations (SDE). A typical choice for the diffusion process on Euclidean data is the *variance-expanding* SDE,²

$$dx_t = g_t dW_t. \quad (1)$$

¹The learned reverse transport process is a continuous normalizing flow, and the unbiased estimators used in that context directly apply here [14].

²Here, dW_t is an infinitesimal Wiener process increment defined by $dW_t = W_{t+dt} - W_t \sim \mathcal{N}(0, dt)$.

The time-dependent probability density $p_t(x)$ of the samples evolves according to the Fokker-Planck Equation, which is in this case the heat equation

$$\partial_t p_t(x) = \frac{g_t^2}{2} \Delta p_t(x), \quad \Delta = \partial_x^2. \quad (2)$$

The fundamental solution to Eq. (2) is the *heat kernel*,

$$K_t(x) = \frac{1}{\sqrt{2\pi\sigma_t^2}} \exp\left[-\frac{1}{2\sigma_t^2} \|x\|^2\right], \quad (3)$$

where the diffusivity $\sigma_t \geq 0$ is defined by $\sigma_t = \sqrt{\int_0^t g_{t'}^2 dt'}$. The forward evolution of any target density $p_0(x)$ has an exact solution obtained by convolving against K :

$$p_0(x) \mapsto p_t(x) = \int_{\mathbb{R}^d} K_t(x-y) p_0(y) dy. \quad (4)$$

Thus, simulating the forward diffusion process from x_0 to x_t is equivalent to solving the heat equation to arbitrary time t . This requires sampling noise $\Delta x_t \sim K_t(\cdot)$, which is trivially accomplished in Euclidean space by sampling $\eta \sim \mathcal{N}(0, \mathbb{I})$ and setting $\Delta x_t = \sigma_t \eta$, leading to the forward map

$$x_0 \mapsto x_t = x_0 + \sigma_t \eta. \quad (5)$$

2.2 Diffusion Processes on $SU(N)$

An analogous diffusion process can be defined on the $SU(N)$ group manifold as

$$dU_t = ig_t \left[\sum_{a=1}^{N^2-1} T^a dW_t^a \right] U_t. \quad (6)$$

Due to the non-Abelian nature of $SU(N)$, the stochastic evolution of U_t is given by a path-ordered exponential $U_t = \mathcal{P} \exp \left[i \int_0^t g_{t'} \sum_{a=1}^{N^2-1} T^a dW_{t'}^a \right] U_0$. Sampling U_t using simulation would thus require a computationally challenging numerical approximation to this integral. However, we demonstrate in the following section that the analytically known heat kernel of $SU(N)$ can be sampled directly as an efficient alternative.

3. The Heat Kernel over $SU(N)$

The Fokker-Planck equation corresponding to (6) is obtained by replacing the Euclidean Laplace operator with the Laplace-Beltrami operator in (2), giving

$$\partial_t p_t(U) = \frac{g_t^2}{2} \Delta_U p_t(U), \quad \Delta_U f(U) = \sum_{a=1}^{N^2-1} \frac{\partial^2}{\partial \omega_a^2} f \left(e^{i\omega_a T^a} U \right) \Big|_{\omega=0}. \quad (7)$$

The $SU(N)$ heat kernel $K_t(U)$ is the Green's function solution of (7) sourced by a delta distribution at the identity, so that forward evolution under the heat equation is solved by convolving against K ,

$$p_0(U) \mapsto p_t(U) = \int dV K_t(UV^\dagger) p_0(V). \quad (8)$$

Combining left- and right-invariance of Δ_U with the invariance of the identity matrix I under unitary conjugation, the heat kernel satisfies $K_t(U) = K_t(P^\dagger U P)$. The heat kernel is therefore completely specified by its value on the subgroup of diagonal matrices.³ Matrices belonging to this subgroup can be identified with a list of eigenangles,

$$\Lambda = \text{diag}(\lambda_1, \dots, \lambda_N) = \text{diag}(e^{i\theta_1}, \dots, e^{i\theta_N}) \quad \leftrightarrow \quad \boldsymbol{\theta} = (\theta_1, \dots, \theta_N), \quad (9)$$

where $\sum_j \theta_j \equiv 0 \pmod{2\pi}$. By convention, we chose $\theta_j \in [-\pi, \pi]$ for all explicit representations of the eigenangles below. With a slight abuse of notation, we write $K_t(\boldsymbol{\theta})$ for the heat kernel on the diagonal subgroup.

3.1 Wrapped Normal Representation

The heat kernel on the group is closely related to the heat kernel on the Cartan subalgebra of $\mathfrak{su}(N)$, which is isomorphic to the space $\{\mathbf{x} \mid \sum_{i=1}^N x_i = 0\}$. In particular, applying an appropriate measure factor and wrapping using the exponential map, $\theta_j = \arg(e^{ix_j})$, leads to the representation [15]

$$K_t(\boldsymbol{\theta}) \propto \sum_{\mathbf{m} \in \mathbb{Z}^N} \tilde{K}_t(\boldsymbol{\theta} + 2\pi\mathbf{m}) \delta\left(\sum_i m_i\right), \quad \tilde{K}_t(\mathbf{x}) = \prod_{i < j} \frac{x_i - x_j}{2 \sin\left[\frac{x_i - x_j}{2}\right]} \exp\left[-\frac{1}{2\sigma_t^2} \sum_i x_i^2\right], \quad (10)$$

where the sum over $\mathbf{m} \in \mathbb{Z}^N$ represents the sum over all equivalent pre-images of $\boldsymbol{\theta}$.⁴

We can also analytically compute the score function for (10). First, the score for the heat kernel over the space of unwrapped eigenangles is given by

$$\partial_i \log \tilde{K}_t(\mathbf{x}) = -\frac{x_i}{\sigma_t^2} + \sum_{j \neq i} \left[\frac{1}{x_i - x_j} - \frac{1}{2} \cot \frac{x_i - x_j}{2} \right], \quad (11)$$

where the first term is the Euclidean score function for the Gaussian distribution and the second term corresponds to the gradient of the measure term. The wrapped score is then

$$\nabla \log K_t(\boldsymbol{\theta}) = \frac{1}{K_t(\boldsymbol{\theta})} \sum_{\mathbf{m} \in \mathbb{Z}^N} \tilde{K}_t(\boldsymbol{\theta} + 2\pi\mathbf{m}) \nabla \log \tilde{K}_t(\boldsymbol{\theta} + 2\pi\mathbf{m}) \delta\left(\sum_i m_i\right). \quad (12)$$

3.2 Dual Representation

The spectral heat kernel may also be written as a character sum or eigenfunction expansion which, for general N , takes the form [15, 16]

$$K_t(U) = \sum_{\boldsymbol{\mu}} \dim(\boldsymbol{\mu}) \chi_{\boldsymbol{\mu}}(U) \exp\left[-\sigma_t^2 c_2(\boldsymbol{\mu})\right] \quad (13)$$

where the sum runs over partitions $\boldsymbol{\mu} = \{\mu_1 \geq \dots \geq \mu_N = 0\}$ labeling the $SU(N)$ irreps. The value of the quadratic Casimir is the eigenvalue of the $SU(N)$ Laplace-Beltrami operator:

$$c_2(\boldsymbol{\mu}) = \sum_{j=1}^N \mu_j (\mu_j - 2j + N + 1) - \frac{1}{N} \left(\sum_{j=1}^N \mu_j \right)^2. \quad (14)$$

³This subgroup is known as the maximal torus $\mathbb{T} \leq SU(N)$.

⁴The sums in Eq. (10) and Eq. (12) can be calculated more stably using the **LogSumExp** function to avoid singularities.

The Weyl character and Weyl dimension formulas [17] give

$$\chi_{\mu}(U) = \frac{\det A(U)}{\det B(U)}, \quad \dim(\mu) = \prod_{j < k} \frac{\mu_j - \mu_k + k - j}{k - j} \quad (15)$$

where the matrix elements of A and B are given in terms of the eigenvalues λ_i of U as

$$[A(U)]_{ij} = \lambda_i^{\mu_j + N - j}, \quad [B(U)]_{ij} = \lambda_i^{N - j}. \quad (16)$$

The resulting score function is then

$$\partial_i \log K_t(\theta) = \frac{1}{K_t(U)} \sum_{\mu} \dim(\mu) \chi_{\mu}(U) \text{Tr} [A^{-1} \partial_{\theta_i} A - B^{-1} \partial_{\theta_i} B] \exp [-\sigma_t^2 c_2(\mu)]. \quad (17)$$

In contrast to (10), here σ_t^2 appears in the numerator of the exponent. These two representations are complementary: if σ_t is small, the sum over m rapidly converges in (10), while if σ_t is large, the sum over μ rapidly converges in (13). For the limited range of σ_t used in the one-variable targets below, we find that the wrapped normal representation remains stable for all times, but we expect that larger σ_t may be required for higher-dimensional examples to fully diffuse the data to a uniform prior distribution, making the dual representation critical for stable training.

3.3 Sampling the Heat Kernel

Given the explicit representations above, the full heat kernel over the group can be efficiently sampled in a two-step process. First, we sample the *marginal* distribution over eigenvalues,

$$\theta \sim \prod_{i < j} 4 \sin^2 \left[\frac{\theta_i - \theta_j}{2} \right] K_t(\theta), \quad (18)$$

where marginalizing over the eigenvectors introduces the extra Haar measure term. Second, we sample eigenvectors $P \in U(N)$ uniformly with respect to the Haar measure of $U(N)$, and construct

$$V = P \text{diag} \left(e^{i\theta} \right) P^{\dagger} \sim K_t(V). \quad (19)$$

In this work, we sample the marginal distribution using a Metropolis sampling scheme with a proposal distribution tuned separately for $\sigma_t < 0.5$ (wrapped normal) and $\sigma_t > 0.5$ (uniform). A similar sampling approach was successfully applied in Ref. [13].

4. Spectral Diffusion

The forward diffusion process on the group is simulated by

$$U_0 \mapsto U_t = V U_0, \quad V = P \text{diag} \left(e^{i\theta} \right) P^{\dagger} \sim K_t(V). \quad (20)$$

We show an example of the forward spectral diffusion process on $SU(2)$ in Figure 1.

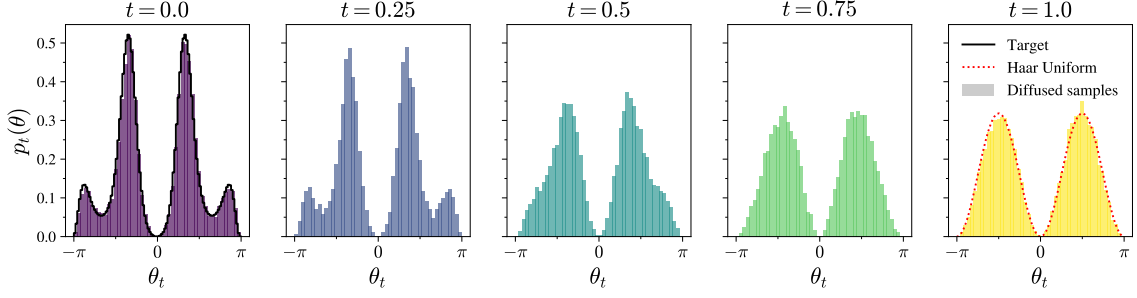


Figure 1: Evolution from left to right of the $SU(2)$ spectral density on the first eigenangle under the forward diffusion process. The starting (target) density is progressively smoothed out by the heat kernel. After 1 unit of diffusion time, the diffused density closely matches the $SU(2)$ Haar uniform density.

4.1 Group-Valued Score Matching

Score-based diffusion models are conventionally trained via *score matching*, which minimizes the distance between the model score function and an unbiased estimator of the true score function across diffusion time. The $SU(N)$ score-matching loss function, adapted to the $SU(N)$ group manifold, is

$$\mathcal{L} = \int_0^1 \mathbb{E}_{U_0 \sim p_0} \mathbb{E}_{V \sim K_t} \left[\sigma_t^2 \|s_t(U) - P^\dagger \text{diag}[\nabla \log K_t(\theta)] P\|^2 \right] dt. \quad (21)$$

Here, θ and P are the sampled eigenangles and eigenvectors used to construct V in (19), $s_t(V)$ takes values in the algebra $\mathfrak{su}(N)$, and the norm is over algebra components. As usual for score matching, computing (21) requires samples U_0 from the target distribution. To verify the correctness of our approach, we have applied the training routine detailed here to successfully learn the analytically known heat kernel score function (12) by training with a fixed initial condition $U_0 = I$.

4.2 Reverse Transport

Given a trained score function, the ODE for the reverse transport process is

$$\partial_t U_t = \left[-\frac{i}{2} g_t^2 s_t(U_t) \right] U_t, \quad (22)$$

to be integrated in reverse from $t = 1 \rightarrow 0$. We can compute the model probability density at each time t as

$$\log q_t(U_t) = \int_t^1 \frac{g_{t'}^2}{2} \nabla \cdot s_{t'}(U_{t'}) dt'. \quad (23)$$

This allows reweighting factors to be introduced for unbiased evaluation at $t = 0$: for each sample U drawn, we evaluate the model density $\log q(U) \equiv \log q_0(U)$ and the target density $\log p(U_i) = -S(U_i) + \text{const.}$, giving the unnormalized weights

$$w(U) = e^{-S(U)} / q(U). \quad (24)$$

	$c^{(1)}$			$c^{(2)}$			$c^{(3)}$		
β	1.0	5.0	9.0	1.0	5.0	9.0	1.0	5.0	9.0
SU(2) ESS (%)	99.6	96.6	97.1	96.7	97.8	95.4	99.5	98.4	98.7
SU(3) ESS (%)	98.8	93.7	97.7	97.6	95.1	93.8	99.2	97.1	96.8

Table 1: Final ESS values for different β values and coefficient sets in the $SU(N)$ toy model.

5. Toy Example

To illustrate the ability of our method to sample from non-trivial target distributions, we define a family of simple densities $p^{(i)}$ over a single $SU(N)$ matrix U , where the action for the toy theory is polynomial in the matrix:

$$S^{(i)}(U) = -\frac{\beta}{N} \text{Re Tr} \left[\sum_{n=1}^3 c_n^{(i)} U^n \right], \quad p^{(i)}(U) = \frac{1}{Z} e^{-S^{(i)}(U)}. \quad (25)$$

The action is parameterized by a set of coefficients $c_n^{(i)}$, $i \in \{1, 2, 3\}$ which determine the shape of the density. For comparison to the spectral flow-based approach, we choose the same sets of coefficients as in Table I of Ref. [18]. We investigate both $SU(2)$ and $SU(3)$ targets.

The diffusion process uses a diffusion schedule $g_t = \kappa t^\alpha$ with $\kappa = 3.0$, $\alpha = 1.0$, and we stop diffusion at $t = 1$, which is sufficient to reach a uniform distribution. For training, 32,768 configurations were generated using the Metropolis algorithm with 1000 thermalization steps, keeping every tenth configuration. Unless otherwise stated, we train for 100 epochs using the Adam optimizer [19] with a batch size of 1024 and a base learning rate of 10^{-3} ; for $SU(3)$ experiments we instead use a batch size of 3072 and a learning rate of 10^{-2} . Using the fact that the target distribution is a function of only the eigenvalues, we define a specialized score network $s_t(U_t) = s_t(\theta_t)$ acting directly on eigenvalues and returning non-zero values only in the Cartan subalgebra. The score network is parameterized by a multilayer perceptron with three hidden layers and **SiLU** activations, using sinusoidal positional encodings to embed the input times.

For inference, we sample the uniform distribution and apply reverse transport using a simple group-valued Euler integrator with 200 steps. Fig. 2 displays the measure target and model log-likelihoods, $-S$ and $\log q$, the unnormalized reverse Kullback-Leibler divergence $D_{\text{KL}}(q||p) \equiv \mathbb{E}_{U \sim q}[\log q + S]$, the effective sample size (ESS),

$$\text{ESS} := \frac{\mathbb{E}[w(U)]^2}{\mathbb{E}[w(U)^2]}, \quad 0 \leq \text{ESS} \leq 1, \quad (26)$$

as well as an unbiased estimator of the partition function, $Z = \int dU e^{-S(U)} = \mathbb{E}[w]$. We summarize the final effective sample sizes we obtain for the toy models in Table 1 and compare the learned distributions with the target densities in Figs. 3 and 4.

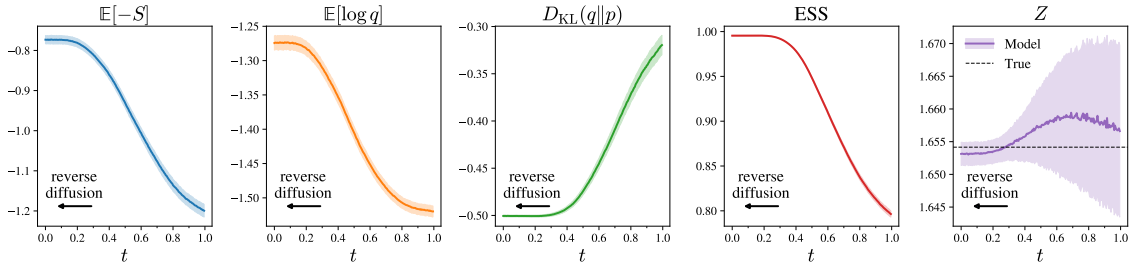


Figure 2: Reverse transport process for the toy $SU(2)$ model with $\beta = 1.0$ and coefficient set $c^{(3)}$. The horizontal axis on each plot represents forward diffusion time, so the evolution should be interpreted as right to left. The KL divergence is seen to decrease while the ESS increases over time, signifying that the target density is being approximately reproduced by the model during sampling.

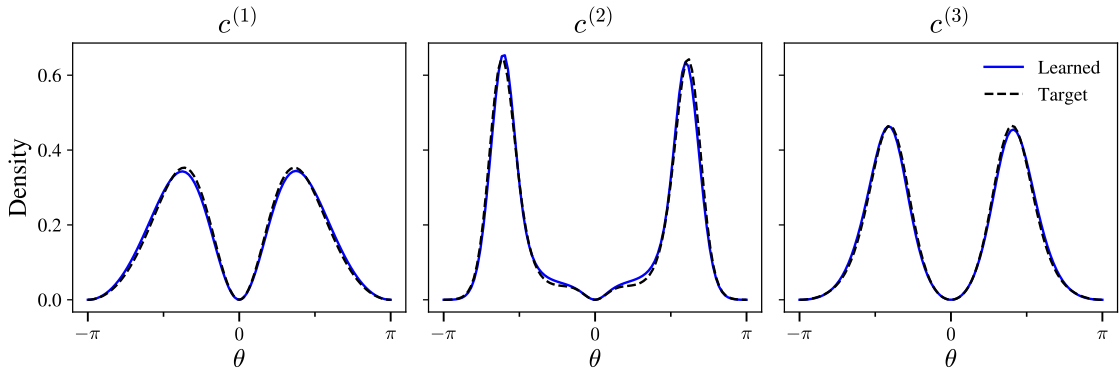


Figure 3: Learned and target $SU(2)$ densities at $\beta = 1.0$. Models were trained for 200 epochs. Unbiased results can be obtained by reweighting estimates from the model to the target density with (24).

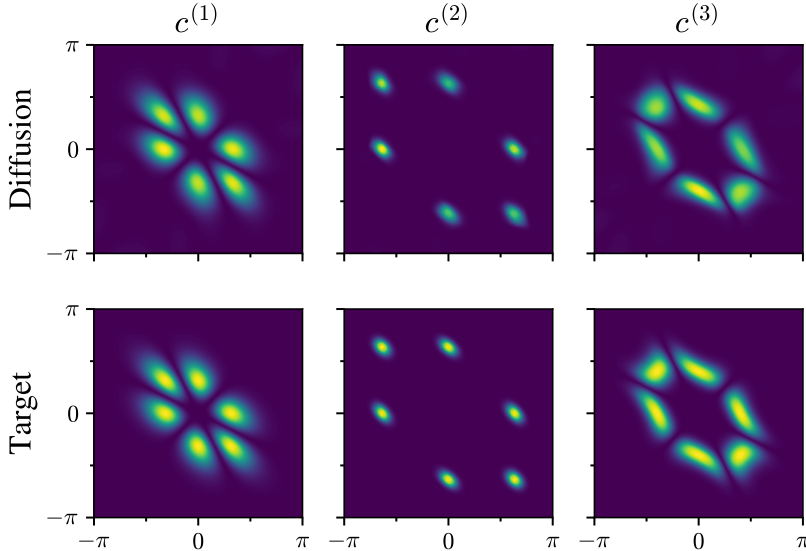


Figure 4: Comparison between the learned and target $SU(3)$ densities at $\beta = 9.0$, shown over the first two eigenangles θ_1 and θ_2 . Unbiased results can be obtained from the diffusion model with (24).

6. Outlook

We have demonstrated an implementation of the $SU(N)$ heat kernel, its score function, and training using $SU(N)$ score matching. Analytic results for the score function of the heat kernel and a numerically stable implementation were key steps to making this method possible. This lays the groundwork for future studies with the potential to address the challenges of ensemble generation for $SU(N)$ lattice gauge theories.

Acknowledgments

We thank Diaa Habibi for bringing our attention to related results [13] during the preparation of this work. We also wish to thank Javad Komijani and Aida El-Khadra for guidance and helpful discussions. We are grateful to the MITP Summer School 2025 for hosting an engaging exchange leading to the ideas presented here. OV acknowledges support from the U.S. Department of Energy, Office of Science under grant Contract Number DE-SC001565, as well as 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** [20] and **NumPy** [21]. Figures were created with **Matplotlib** [22]. Some of the experiments in this research used the DeltaAI advanced computing and data resource, which is supported by the National Science Foundation (award OAC 2320345) and the State of Illinois.

References

- [1] S. Duane, A. D. Kennedy, B. J. Pendleton, and D. Roweth, *Hybrid Monte Carlo*, *Phys. Lett. B* **195** (1987) 216.
- [2] P. Woit, *Topological Charge in Lattice Gauge Theory*, *Phys. Rev. Lett.* **51** (1983) 638.
- [3] 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].
- [4] U. Wolff, *Critical Slowing Down*, *Nucl. Phys. B Proc. Suppl.* **17** (1990) 93.
- [5] J. Sohl-Dickstein, E. Weiss, N. Maheswaranathan, S. Ganguli, *Deep Unsupervised Learning using Nonequilibrium Thermodynamics*, *PMLR* **37** (2015) 2256 [1503.03585].
- [6] J. Ho, A. Jain, and P. Abbeel, *Denosing diffusion probabilistic models*, *NIPS ’20: Proceedings of the 34th International Conference on Neural Information Processing Systems* **574** (2020) 6840 [2006.11239].
- [7] Y. Song and S. Ermon, *Generative modeling by estimating gradients of the data distribution*, *Proceedings of the 33rd International Conference on Neural Information Processing Systems* **1067** (2019) 11918 [1907.05600].
- [8] Y. Song, J. Sohl-Dickstein, D. P. Kingma, A. Kumar, S. Ermon and B. Poole, *Score-Based Generative Modeling through Stochastic Differential Equations*, 2011.13456.

- [9] L. Wang, G. Aarts and K. Zhou, *Diffusion models as stochastic quantization in lattice field theory*, *JHEP* **05** (2024) 060 [2309.17082].
- [10] Q. Zhu, G. Aarts, W. Wang, K. Zhou, and L. Wang, *Physics-Conditioned Diffusion Models for Lattice Gauge Theory*, 2502.05504.
- [11] O. Vega, J. Komijani, A. X. El-Khadra, and M. Marinkovic, *Group-Equivariant Diffusion Models for Lattice Field Theory*, 2510.26081.
- [12] C. Huang, M. Aghajohari, A. J. Bose, P. Panangaden, and A. Courville, *Riemannian Diffusion Models*, *NIPS'22: Proceedings of the 36th International Conference on Neural Information Processing Systems* **199** (2022) 2750 [2208.07949].
- [13] A. Lou, M. Xu, A. Farris, and S. Ermon, *Scaling Riemannian Diffusion Models*, *NIPS '23: Proceedings of the 37th International Conference on Neural Information Processing Systems* **3519** (2023) 80291 [2310.20030].
- [14] M. Gerdes, P. de Haan, C. Rainone, R. Bondesan and M. C. N. Cheng, *Learning lattice quantum field theories with equivariant continuous flows*, *SciPost Phys.* **15** (2023) 238 [2207.00283].
- [15] P. Menotti and E. Onofri, *The Action of $SU(N)$ Lattice Gauge Theory in Terms of the Heat Kernel on the Group Manifold*, *Nucl. Phys. B* **190** (1981) 288.
- [16] B. E. Ehsan, *Character functions of $SU(3)$* , *J. Phys. A* **21** (1988) 2651.
- [17] J. M. Drouffe and J. B. Zuber, *Strong Coupling and Mean Field Methods in Lattice Gauge Theories*, *Phys. Rept.* **102** (1983) 1.
- [18] D. Boyda, G. Kanwar, S. Racanière, D. J. Rezende, M. S. Albergo, K. Cranmer, D. C. Hackett and P. E. Shanahan, *Sampling using $SU(N)$ gauge equivariant flows*, *Phys. Rev. D* **103** (2021) 074504 [2008.05456].
- [19] D. P. Kingma and J. Ba, *Adam: A Method for Stochastic Optimization*, 1412.6980.
- [20] A. Paszke, S. Gross, F. Massa, A. Lerer, J. Bradbury, G. Chanan et al., *PyTorch: An Imperative Style, High-Performance Deep Learning Library*, *Proceedings of the 33rd International Conference on Neural Information Processing Systems* **721** (2019) 8024 [1912.01703].
- [21] 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].
- [22] J. D. Hunter, *Matplotlib: A 2D Graphics Environment*, *Computing in Science & Engineering* **9** (2007) 90.