
Generative Diffusion Models for Lattice Field Theory

Lingxiao Wang

Frankfurt Institute for Advanced Studies
& Xidian-FIAS International Joint Research Center
Frankfurt am Main, D-60438, Germany
lwang@fias.uni-frankfurt.de

Gert Aarts

Department of Physics
Swansea University
SA2 8PP, Swansea, United Kingdom

European Centre for Theoretical Studies in Nuclear Physics and Related Areas (ECT*)
& Fondazione Bruno Kessler Strada delle Tabarelle
286, 38123 Villazzano (TN), Italy
g.aarts@swansea.ac.uk

Kai Zhou

Frankfurt Institute for Advanced Studies
Frankfurt am Main, D-60438, Germany
zhou@fias.uni-frankfurt.de

Abstract

This study delves into the connection between machine learning and lattice field theory by linking generative diffusion models (DMs) with stochastic quantization, from a stochastic differential equation perspective. We show that DMs can be conceptualized by reversing a stochastic process driven by the Langevin equation, which then produces samples from an initial distribution to approximate the target distribution. In a toy model, we highlight the capability of DMs to learn effective actions. Furthermore, we demonstrate its feasibility to act as a global sampler for generating configurations in the two-dimensional ϕ^4 quantum lattice field theory.

1 Introduction

In lattice field theory, physical observables are obtained by approximating path integrals via summing over field configurations, traditionally done through Monte Carlo methods. However, these methods can be computationally expensive. A promising alternative lies in employing generative models, a machine learning framework, to create new configurations following the target physical distribution, thus potentially enhancing the efficiency of Monte Carlo simulations [1, 2].

Generative models fall into two main categories based on likelihood estimation methods. Implicit Maximum Likelihood Estimation (MLE) employs, for instance, Generative Adversarial Networks (GANs) to generate new configurations through a min-max game, shown effective in lattice simulations [3, 4]. Explicit MLE uses explicit probability descriptions, e.g., autoregressive models [5] and flow-based models [6–11], improving efficiency in lattice simulations without needing prepared training data. Despite being plagued by model-collapse and scalability [12–14], flow-based models have developed rapidly and yielded achievements [2]. Recently, Diffusion Models (DMs), a new

implicit MLE class but with an explicit probability description, have shown promise in generating high-quality images via stochastic processes [15], hinting at potential applications also in high-energy physics [16, 17].

This work explores the potential of DMs in generating lattice field configurations, and its intrinsic connection with stochastic quantization (SQ) [18–20] from a stochastic differential equation (SDE) perspective. We demonstrate its efficiency and accuracy of learning effective actions in a toy model, and verify its feasibility of generating configurations in a two-dimensional ϕ^4 lattice field theory. Our findings suggest that DMs can serve as a significant tool to address computational challenges in lattice simulations, encouraging further explorations in this direction.

Related Work Langevin dynamics has been utilized in Bayesian learning as a stochastic gradient optimization [21], which introduces stochasticity into the parameter updates, thereby avoiding collapses into local minima. Recent related work introduces stochasticity into the hybrid Monte-Carlo algorithm [22] and explores the correspondence between the exact renormalizing group (ERG) and DMs based upon the heat equation [23]. From a flow-based model perspective, Ref. [24] designed a continuous-time normalizing flow with an inferred velocity field from the probability current of a time-dependent density that interpolates between the prior and target densities stochastically.

2 Stochastic Differential Equation

In DMs, a denoising model reconstructs original data from its noisy version obtained through a diffusion process. The predetermined process of adding noises, also called the forward process, smoothens the data’s probability distribution by introducing noise. The denoising model learns the inverse process—eliminating this noise. Once trained, it generates samples following the data distribution through a reverse diffusion process, starting with random noise and iteratively "cleaning" it until a convergent sample resembling the target data distribution is achieved.

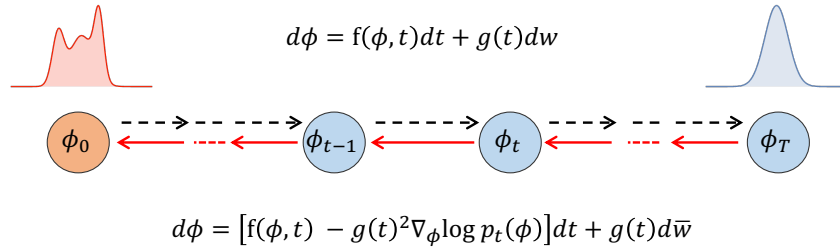


Figure 1: A sketch of the forward diffusion process (upper arrows) and the reverse denoising process (bottom arrows). The two stochastic processes are described by two stochastic differential equations. The target distribution is typically unknown but learnt from the initial data.

The forward process mentioned above can be introduced into the propagation of the field, ϕ_i , via the following Markov chain, $\phi_i = \phi_{i-1} + f_{i-1} + g_{i-1}z_{i-1}$, with a total of N steps ($i = 1, \dots, N$) and random noise $z_i \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$. Both the random noise z_i and drift force f_i have the same dimensionality as the field ϕ_i . Given an time interval $T \equiv Ndt$ and $N \rightarrow \infty$ ($dt \rightarrow 0$), the above forward process converges to its continuous-time limit, which follows an Itô SDE, $d\phi = \mathbf{f}(\phi, t)dt + g(t)d\mathbf{w}$, where $t \in [0, T]$, \mathbf{w} is the standard Wiener process, i.e, Brownian motion, $\mathbf{f}(\phi, t)$ is the *drift term*, and $g(t)$ is the scalar *diffusion coefficient*. The forward diffusion process $\phi(t)$ can be modeled as the solution of such a generic SDE.

As Figure 1 demonstrates, starting from a sample taken from the prior distribution p_T and reversing the above diffusion process enables obtaining a sample from the data distribution p_0 . Importantly, the reverse process represents a diffusion process evolving backward in time, which is expressed by the following reverse SDE [25], $d\phi = [\mathbf{f}(\phi, t) - g^2(t)\nabla_\phi \log p_t(\phi)]dt + g(t)d\bar{\mathbf{w}}$, where the reverse time $t \equiv T - t$, and $p_t(\phi)$ is the probabilistic distribution at time-step t , $\bar{\mathbf{w}}$ is a Wiener process in the reverse time direction, and dt represents an infinitesimal negative time step. This reverse SDE can be

solved once we know the drift term and diffusion coefficient of the forward SDE, and in particular $\nabla_\phi \log p_t(\phi)$ for each $t \in [0, T]$.

The reverse SDEs of DMs are mathematically related to Langevin dynamics. For a concise implementation, we choose the *variance expanding* picture of DMs, i.e. setting $f(\phi, t) \equiv 0, g(t) \equiv g_\tau$. Its Langevin equation (labeled by a new reverse time τ) now reads,

$$\frac{d\phi}{d\tau} = g_\tau^2 \nabla_\phi \log p_\tau(\phi) + g_\tau \bar{\eta}(\tau), \quad (1)$$

where the noise term $\langle \bar{\eta}(\tau) \rangle = 0, \langle \bar{\eta}(\tau) \bar{\eta}(\tau') \rangle = 2\bar{\alpha} \delta(\tau - \tau')$, with $\bar{\alpha}$ being the diffusion constant. Solving the reverse SDE (1) to depict denoising is difficult due to the intractable ‘‘time-dependent’’ drift term. A U-Net neural network is used to parameterize the score function, $\mathbf{s}_\theta(\phi, \tau)$, which estimates the drift term, $-\nabla_\phi \log p_\tau(\phi)$, in Eq. (1). The U-Net accepts time and a trajectory configuration as inputs and outputs the same size as the input. More details about the architecture of the U-Net can be found in Ref. [26].

2.1 Stochastic Quantization

In field theory, as an alternative quantization scheme, one can introduce SQ for real actions [18, 19], or complex Langevin dynamics for complex actions [27, 19]. Starting from a generic Euclidean path integral, $Z = \int D\phi e^{-S_E}$, SQ introduces a *fictitious* time τ for the field ϕ , whose evolution is described by Langevin dynamics,

$$\frac{\partial \phi(x, \tau)}{\partial \tau} = -\frac{\delta S_E[\phi]}{\delta \phi(x, \tau)} + \eta(x, \tau), \quad (2)$$

where the noise term satisfies $\langle \eta(x, \tau) \rangle = 0, \langle \eta(x, \tau) \eta(x', \tau') \rangle = 2\alpha \delta(x - x') \delta(\tau - \tau')$, with α being the diffusion constant. In the long-time limit, for real actions the system reaches an equilibrium state $P_{\text{eq}}(\phi) \propto \exp(-S_E(\phi)/\alpha)$, which follows from properties of the associated Fokker-Planck Hamiltonian [19]. For complex actions, there are additional criteria to consider [28].

Comparing Eqs. (1) and (2), one notices the presence of g_τ^2 , which rescales both the drift term and the variance of the noise, and is known as a kernel [19]. Its effect can be absorbed by rescaling time with $g^2(\tau)$, or equivalently absorbing it in the time step, $g_\tau^2 \Delta\tau$. One can then identify the drift term in Eq. (1) with the gradient of an effective DM action S_{DM} , using $\nabla_\phi S_{\text{DM}}(\phi, \tau) \equiv -\nabla_\phi \log p_\tau(\phi) \approx \mathbf{s}_\theta(\phi, \tau)$. In the $\tau \rightarrow T$ limit, the distribution $p_{\tau=T}(\phi) \rightarrow P[\phi, T] \propto \exp(-S_{\text{DM}}/\bar{\alpha})$. Upon identifying $\bar{\alpha}$ and α , this implies that the equilibrium state from a SQ perspective can be obtained by denoising a naive distribution using the DM prescription. Concurrently, sampling from a DM is equivalent to optimizing a stochastic trajectory to approach the equilibrium state in Euclidean quantum field theory, $P_{\text{eq}}[\phi] \propto \exp(-S_E/\alpha)$. This will be demonstrated in the following Section.

3 Numerical Results

Toy Model To demonstrate the capacity for learning effective DM actions, $S_{\text{DM}}(\phi, \tau)$, defined via the relation $\nabla_\phi S_{\text{DM}}(\phi, \tau) = -\nabla_\phi \log p_\tau(\phi) \approx \mathbf{s}_\theta(\phi, \tau)$, we introduce an oversimplified 0 + 0-dimensional field theory, i.e., a toy model with only one degree of freedom, and the action and drift term,

$$S(\phi) = \frac{\mu^2}{2} \phi^2 + \frac{g}{4!} \phi^4, \quad f(\phi) = -\frac{\partial S(\phi)}{\partial \phi} = -\mu^2 \phi - \frac{g}{3!} \phi^3, \quad (3)$$

with parameters μ^2 and g . We prepared 5120 configurations as training datasets in two setups: $\mu_1^2 = 1.0, g_1 = 0.4$ (single-well action), and $\mu_2^2 = -1.0, g_2 = 0.4$ (double-well action). A one-to-one neural network with time-embedding is implemented to represent the score function $\mathbf{s}_\theta(\phi, \tau)$.

After 500 epochs of training, the learned effective action $S_{\text{DM}}(\phi, \tau) = \int^\phi \hat{\mathbf{s}}_\theta(\phi', \tau) d\phi'$ is seen to approximate the action $S(\phi)$ in the upper panel of Fig. 2; it approaches the physical action as τ increases. We have added to S_{DM} a constant ΔS_0 , which is the difference between $\min[S(\phi)]$ and $\min[S_{\text{DM}}(\phi, \tau)]$. Generally, the learned effective actions are accurate approximations in both the single and double-well cases, around $\phi \sim 0$. In the bottom panel of Fig. 2, samples generated from the trained DM are compared with samples from the underlying theory. In this case, we utilized an Apple M2 Pro with 16GB of RAM and PyTorch for model training, achieving a total training time of 35 seconds over 500 epochs.

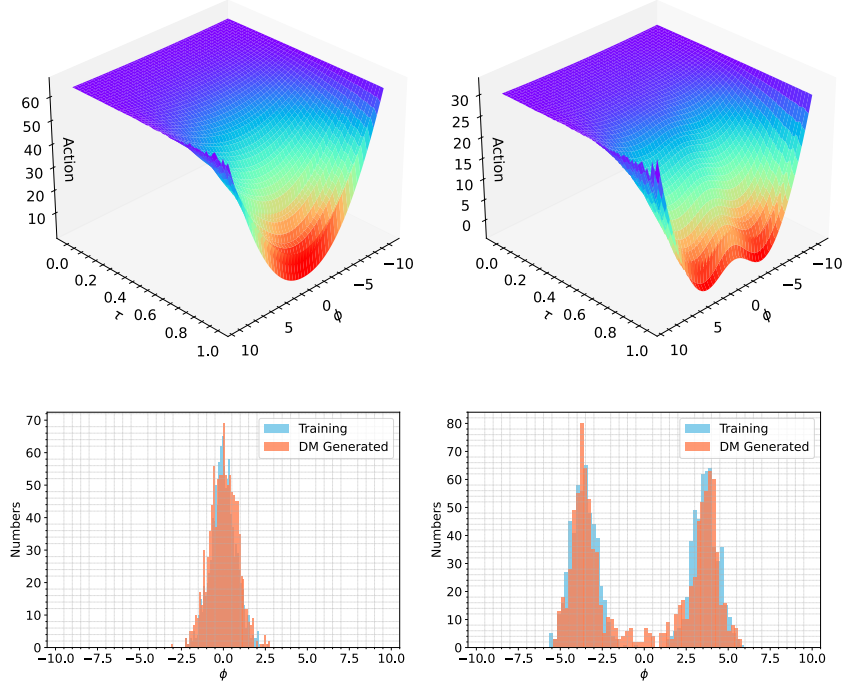


Figure 2: **(Upper panel)** The flow of the effective action, $S_{\text{DM}}(\phi, \tau)$, for various values of time $0 \leq \tau \leq T = 1$ during the stochastic process, learned by the diffusion model as a function of ϕ for both the single-well (left column) and double-well (right column) actions, using the relation $\nabla_{\phi} S_{\text{DM}}(\phi, \tau) = -\nabla_{\phi} \log p_{\tau}(\phi) \approx \mathbf{s}_{\theta}(\phi, \tau)$. **(Bottom panel)** The first 1024 training samples (blue histograms) and 1024 generated samples (orange histograms) for both the single-well (left column) and double-well (right column) actions.

Scalar Lattice Field Theory We consider a real scalar field in d *Euclidean* dimensions with the dimensionless action,

$$S_E = \sum_x \left[-2\kappa \sum_{\nu=1}^d \phi(x)\phi(x+\hat{\nu}) + (1-2\lambda)\phi^2(x) + \lambda\phi^4(x) \right], \quad (4)$$

where κ is the hopping parameter, and λ denotes the dimensionless coupling constant describing field interactions. Both parameters are positive, more details can be found in Ref. [29].

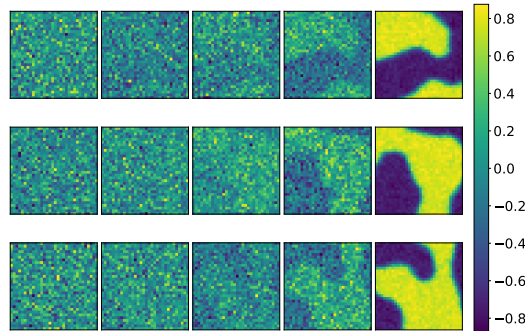


Figure 3: Generation of four independent configurations from a well-trained diffusion model in the broken phase. Each row in the figure represents a different sample, and each column represents a different time point ($\tau \in [0, 0.25, 0.5, 0.75, T = 1]$) during the denoising process.

In the broken phase ($\kappa = 0.5, \lambda = 0.022$), field configurations behave like large clusters. We demonstrate that the clustering behavior of field configurations in a $d = 2$ dimensional case, in which

it can be successfully captured by the well-trained DM. Fig. 3 visualizes the denoising process. The first column represents noise samples randomly drawn from the prior normal distribution, while the fifth column represents the generated samples obtained by denoising. Training set-ups and more quantitative evaluations both in the broken phase and symmetric phase can be found in Ref. [26].

4 Conclusion and Outlook

In this contribution, a novel method is introduced to generate quantum field configurations using generative diffusion models. The connection with stochastic quantization is highlighted, an approach to quantize field theories based on a stochastic process in a fictitious time direction. In DMs, the drift term is learned from prepared configurations in a forward process, whereas in SQ it is known and derived from the physical action. The approach is demonstrated in a toy model and a two-dimensional scalar ϕ^4 field theory. Future directions include further exploring the connection between DMs and SQ, training DMs without a training data set, and using DMs to incorporate the effect of fermions in QCD. Additionally, combining DMs with complex Langevin dynamics may be used to generate configurations for theories with a sign problem, such as e.g. QCD at nonzero baryon density [30].

Acknowledgments and Disclosure of Funding

We thank Profs. Tetsuo Hatsuda, Shuzhe Shi and Xu-Guang Huang for helpful discussions. We thank ECT* and the ExtreMe Matter Institute EMMI at GSI, Darmstadt, for support during the ECT*/EMMI workshop *Machine learning for lattice field theory and beyond* in June 2023 during the preparation of this paper. The work is supported by (i) the BMBF funded KISS consortium (05D23R11) in the ErUM-Data action plan (K.Z.), (ii) the AI grant of SAMSON AG, Frankfurt (K.Z. and L.W.), (iii) Xidian-FIAS International Joint Research Center (L.W.), (iv) STFC Consolidated Grant ST/T000813/1 (G.A.). K.Z. also thanks the donation of NVIDIA GPUs from NVIDIA Corporation.

References

- [1] Kai Zhou, Lingxiao Wang, Long-Gang Pang, and Shuzhe Shi. Exploring QCD matter in extreme conditions with Machine Learning, 2023. arXiv:2303.15136(hep-ph).
- [2] Kyle Cranmer, Gurtej Kanwar, Sébastien Racanière, Danilo J. Rezende, and Phiala E. Shanahan. Advances in machine-learning-based sampling motivated by lattice quantum chromodynamics. *Nature Rev. Phys.*, 5(9):526–535, 2023.
- [3] Kai Zhou, Gergely Endrődi, Long-Gang Pang, and Horst Stöcker. Regressive and generative neural networks for scalar field theory. *Phys. Rev. D*, 100(1):011501, July 2019.
- [4] Jan M. Pawłowski and Julian M. Urban. Reducing autocorrelation times in lattice simulations with generative adversarial networks. *Mach. Learn.: Sci. Technol.*, 1(4):045011, October 2020.
- [5] Lingxiao Wang, Yin Jiang, Lianyi He, and Kai Zhou. Continuous-Mixture Autoregressive Networks Learning the Kosterlitz-Thouless Transition. *Chin. Phys. Lett.*, 39(12):120502, 2022.
- [6] M. S. Albergo, G. Kanwar, and P. E. Shanahan. Flow-based generative models for Markov chain Monte Carlo in lattice field theory. *Phys. Rev. D*, 100(3):034515, August 2019.
- [7] Gurtej Kanwar, Michael S. Albergo, Denis Boyda, Kyle Cranmer, Daniel C. Hackett, Sébastien Racanière, Danilo Jimenez Rezende, and Phiala E. Shanahan. Equivariant Flow-Based Sampling for Lattice Gauge Theory. *Phys. Rev. Lett.*, 125(12):121601, September 2020.
- [8] Kim A. Nicoli, Christopher J. Anders, Lena Funcke, Tobias Hartung, Karl Jansen, Pan Kessel, Shinichi Nakajima, and Paolo Stornati. Estimation of Thermodynamic Observables in Lattice Field Theories with Deep Generative Models. *Phys. Rev. Lett.*, 126(3):032001, January 2021.
- [9] Pim de Haan, Corrado Rainone, Miranda C. N. Cheng, and Roberto Bondesan. Scaling Up Machine Learning For Quantum Field Theory with Equivariant Continuous Flows, 2021. arXiv:2110.02673(cs.LG).

- [10] Mathis Gerdes, Pim de Haan, Corrado Rainone, Roberto Bondesan, and Miranda C. N. Cheng. Learning Lattice Quantum Field Theories with Equivariant Continuous Flows, 2022. arXiv:2207.00283(hep-lat).
- [11] Shile Chen, Oleh Savchuk, Shiqi Zheng, Baoyi Chen, Horst Stoecker, Lingxiao Wang, and Kai Zhou. Fourier-flow model generating Feynman paths. *Phys. Rev. D*, 107(5):056001, 2023.
- [12] Daniel C. Hackett, Chung-Chun Hsieh, Michael S. Albergo, Denis Boyda, Jiunn-Wei Chen, Kai-Feng Chen, Kyle Cranmer, Gurtej Kanwar, and Phiala E. Shanahan. Flow-based sampling for multimodal distributions in lattice field theory, 2021. arXiv:2107.00734(hep-lat).
- [13] Ryan Abbott et al. Aspects of scaling and scalability for flow-based sampling of lattice QCD, 2022. arXiv:2211.07541(hep-lat).
- [14] Kim A. Nicoli, Christopher J. Anders, Tobias Hartung, Karl Jansen, Pan Kessel, and Shinichi Nakajima. Detecting and Mitigating Mode-Collapse for Flow-based Sampling of Lattice Field Theories, 2023. arxiv:2302.14082(hep-lat).
- [15] Ling Yang, Zhilong Zhang, Yang Song, Shenda Hong, Runsheng Xu, Yue Zhao, Wentao Zhang, Bin Cui, and Ming-Hsuan Yang. Diffusion models: A comprehensive survey of methods and applications. *ACM Comput. Surv.*, sep 2023.
- [16] Vinicius Mikuni and Benjamin Nachman. Score-based generative models for calorimeter shower simulation. *Phys. Rev. D*, 106(9):092009, 2022.
- [17] Vinicius Mikuni, Benjamin Nachman, and Mariel Pettee. Fast point cloud generation with diffusion models in high energy physics. *Phys. Rev. D*, 108(3):036025, 2023.
- [18] G. Parisi and Y. S. Wu. Perturbation theory without gauge fixing. *Sci. China, A*, 24(ASITP-80-004):483, August 1980.
- [19] Poul H. Damgaard and Helmuth Hüffel. Stochastic quantization. *Phys. Rept.*, 152(5):227–398, August 1987.
- [20] M. Namiki, I. Ohba, K. Okano, Y. Yamanaka, A. K. Kapoor, H. Nakazato, and S. Tanaka. *Stochastic quantization*, volume 9. Springer Berlin Heidelberg, 1992.
- [21] Max Welling and Yee Whye Teh. Bayesian learning via stochastic gradient langevin dynamics. In *Proc. 28th Int. Conf. Int. Conf. Mach. Learn.*, ICML’11, pages 681–688, Madison, WI, USA, June 2011. Omnipress.
- [22] Jakob Robnik and Uroš Seljak. Microcanonical Langevin Monte Carlo, 2023. arXiv:2303.18221(hep-lat).
- [23] Jordan Cotler and Semon Rezhikov. Renormalizing Diffusion Models, 2023. arXiv:2308.12355(hep-th).
- [24] Michael Samuel Albergo and Eric Vanden-Eijnden. Building Normalizing Flows with Stochastic Interpolants. In *The Eleventh International Conference on Learning Representations(ICLR 2023)*, March 2023.
- [25] Brian D. O. Anderson. Reverse-time diffusion equation models. *Stochastic Processes and their Applications*, 12(3):313–326, May 1982.
- [26] Lingxiao Wang, Gert Aarts, and Kai Zhou. Diffusion Models as Stochastic Quantization in Lattice Field Theory, 2023. arXiv:2309.17082(hep-lat).
- [27] Giorgio Parisi. On complex probabilities. *Physics Letters B*, 131(4):393–395, November 1983.
- [28] Gert Aarts, Erhard Seiler, and Ion-Olimpiu Stamatescu. Complex Langevin method: When can it be trusted? *Phys. Rev. D*, 81(5):054508, March 2010.
- [29] J. Smit. *Introduction to quantum fields on a lattice: A robust mate*, volume 15. Cambridge University Press, 1 2011.
- [30] Gert Aarts. Introductory lectures on lattice QCD at nonzero baryon number. *J. Phys. Conf. Ser.*, 706(2):022004, 2016.