PROCEEDINGS OF SCIENCE

Sampling from a complex distribution using an energy-based diffusion model

Diaa E. Habibi,[a,*] **Gert Aarts,**[a] **Lingxiao Wang**[b,c] **and Kai Zhou**[d,e]

[a] *Centre for Quantum Fields and Gravity, Department of Physics, Swansea University, Swansea, SA2 8PP, United Kingdom*

[b] *Center for Interdisciplinary Theoretical and Mathematical Sciences (iTHEMS), RIKEN, Wako, Saitama 351-0198, Japan*

[c] *Institute for Physics of Intelligence, Graduate School of Science, The University of Tokyo, Bunkyo-ku, Tokyo 113-0033, Japan*

[d] *School of Science and Engineering, The Chinese University of Hong Kong, Shenzhen (CUHK-Shenzhen), Guangdong, 518172, China*

[e] *School of Artificial Intelligence, The Chinese University of Hong Kong, Shenzhen (CUHK-Shenzhen), Guangdong, 518172, China*

E-mail: n.e.habibi@swansea.ac.uk, g.aarts@swansea.ac.uk, lingxiao.wang@riken.jp, zhoukai@cuhk.edu.cn

For theories with a sign problem, complex Langevin dynamics effectively samples from a real-valued probability distribution that is *a priori* unknown and notoriously hard to predict. In generative AI, diffusion models have proven capable of learning distributions from data. We explore their ability to capture the distributions sampled by a complex Langevin process, comparing score-based and energy-based diffusion models, and outline potential applications.

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

*Speaker

PoS(LATTICE2025)050

1. Introduction

Theories with a complex Boltzmann weight suffer from sign and overlap problems, and simulating them using conventional methods based on importance sampling is difficult. A well-known example is QCD at non-zero baryon density, for which the quark determinant is complex-valued for real quark chemical potential, see e.g. Refs. [1, 2] for reviews. A potential solution is to rely on analytical continuation and extend the theory into a complexified manifold, e.g. via complex Langevin dynamics [3, 4], Lefschetz thimbles [5], holomomorphic flow [6], and variations thereof.

In this contribution we focus on complex Langevin (CL) dynamics, in which a complexified manifold is explored via a stochastic process after analytical continuation [3, 4]. We refer to Ref. [7] for an extensive list of references. An important role in CL dynamics is played by the (real and semi-positive) distribution on this manifold. This distribution is typically not known, as the Fokker-Planck equation (FPE) associated to the CL process cannot be solved in general. However, characterising this distribution is important: once it is determined, it would enable new avenues to link averages over complex-valued Boltzmann weights $\rho(x)$ to statistical averages over real, semi-positive probability distributions $p(x, y)$, via the relation

$$\int dx\, \rho(x) O(x) = \int dxdy\, p(x, y) O(x + iy), \tag{1.1}$$

written here for one degree of freedom. In particular, if the distribution on the RHS is unambiguously determined, one can use importance sampling and the sign problem is evaded. Besides the specific application to CL dynamics, this is an open problem worth studying more generally [8, 9].

Diffusion models [10–13] are a class of generative AI methods, which learn distributions from data. A recent comprehensive review can be found in Ref. [14]. The formulation in terms of stochastic differential equations (SDEs) is closely related to stochastic quantisation [15], using a learned and time-dependent drift term, instead of a drift derived from the action [16, 17]. Diffusion models have been applied to lattice theories for scalar [16, 18, 19] and U(1) gauge fields [20–22], and very recently also to SU(N) one-link models [23], two-dimensional U(2) and SU(2) gauge theories [24] using lattice gauge equivariant convolutional neural networks [25], and two-dimensional SU(2) gauge theory without imposing equivariance [26].

Since diffusion models learn probability distributions directly from data, they can be employed to learn the distribution sampled during the CL process [27]. Subsequently, the trained diffusion models can be used to generate more configurations and study properties of the distribution. Here we focus on the difference between so-called score-based and energy-based diffusion models. This contribution is based on Ref. [7].

2. Distributions sampled by complex Langevin dynamics

We consider one degree of freedom with a complex weight $\rho(x) \sim \exp[-S(x)] \in \mathbb{C}$ throughout. The CL process, with real noise, reads

$$\begin{aligned} \dot{x}(t) &= K_x[x(t), y(t)] + \sqrt{2}\eta(t), & K_x(x, y) &= \mathrm{Re}\, \frac{d}{dz} \log \rho(z)\Big|_{z\to x+iy}, \\ \dot{y}(t) &= K_y[x(t), y(t)], & K_y(x, y) &= \mathrm{Im}\, \frac{d}{dz} \log \rho(z)\Big|_{z\to x+iy}, \end{aligned} \tag{2.1}$$

with Gaussian noise $\eta \sim \mathcal{N}(0,1)$. The corresponding FPE reads

$$\partial_t p(x,y;t) = \left[\partial_x\left(\partial_x - K_x\right) - \partial_y K_y\right] p(x,y;t). \tag{2.2}$$

The assumption is that in the long-time limit averages over trajectories, using Eq. (2.1), yield expectation values with respect to the original weight, i.e.,

$$\lim_{t\to\infty} \langle O[x(t)+iy(t)]\rangle_\eta = \lim_{t\to\infty} \int dxdy\, p(x,y;t) O(x+iy) \stackrel{?}{=} \int dx\, \rho(x) O(x). \tag{2.3}$$

It is well known that this is non-trivial and that convergence to wrong solutions may occur [28, 29]. Here we assume that the process converges to the correct solution and that Eq. (2.3) holds. The FPE (2.2) cannot be solved in general, even in the stationary limit. One reason is that $\partial_x K_y(x,y) \neq \partial_y K_x(x,y)$. Instead we use diffusion models to learn the distribution directly from data.

3. Score- and energy-based diffusion models

Diffusion models are probabilistic generative models that learn from data [10–13]. In our application data generated using CL dynamics form the training set and it is assumed that these are representative of the target distribution $p_0(\mathbf{x}) \equiv \lim_{t\to\infty} p(\mathbf{x},t)$, where $\mathbf{x} = (x,y)$.

Diffusion models consist of a forward process, during which noise is applied to the data, and a backward *denoising* process, during which new data are generated. In the so-called *variance expanding* scheme, using the description in terms of SDEs [13], the forward and backward processes are given by

$$\begin{aligned} &\text{forward:} && \dot{\mathbf{x}}(t) = g(t)\boldsymbol{\eta}(t), \\ &\text{backward:} && \dot{\mathbf{x}}(t) = -g^2(t)\nabla \log p(\mathbf{x},t) + g(t)\boldsymbol{\eta}(t). \end{aligned} \tag{3.1}$$

The forward process runs between $0 \leq t \leq T$, while for the backward process time is reversed. The diffusion coefficient $g(t)$ sets the time-dependent noise strength and $\boldsymbol{\eta} \sim \mathcal{N}(0,1)$ is again Gaussian noise. Initial conditions for the forward process are determined by the target distribution $\mathbf{x}(0) = \mathbf{x}_0 \sim p_0(\mathbf{x}_0)$ and taken from a simple prior for the backward process.

The additional term in the backward process, $\nabla \log p(\mathbf{x},t)$, is the so-called *score*. It is not known *a priori* and is modelled by a neural network $\mathbf{s}_\theta(\mathbf{x},t)$, where θ denotes all the trainable network parameters. It is learnt during the forward process by first writing

$$p(\mathbf{x},t) = \int d\mathbf{x}_0\, p(\mathbf{x},t|\mathbf{x}_0) p_0(\mathbf{x}_0), \tag{3.2}$$

and then minimising the loss function [13],

$$\mathcal{L}(\theta) = \frac{1}{2}\int_0^T dt\, \mathbb{E}_{p(\mathbf{x},t)}\left[\lambda(t) \left\| \mathbf{s}_\theta(\mathbf{x},t) - \nabla \log p(\mathbf{x},t|\mathbf{x}_0)\right\|^2\right], \tag{3.3}$$

where the weight $\lambda(t)$ is chosen to be the variance of the noise at time t. The expectation includes the average over the initial distribution $p_0(\mathbf{x})$, obtained by summing over the data set. After the diffusion model is trained, new samples are generated by numerically solving the backward process. This formulation is known as score-based diffusion models. For a detailed analytical study in a language familiar to lattice field theorists, see Ref. [19].

We are particularly interested in further analysing the distribution obtained after training at the end of the backward process, since this should describe our unknown target distribution, i.e., the distribution sampled with the CL process,

$$\lim_{t\to 0} \mathbf{s}_{\theta^*}(\mathbf{x}, t) \approx \lim_{t\to 0} \nabla \log p(\mathbf{x}, t) \approx \nabla \log p_0(\mathbf{x}), \tag{3.4}$$

where θ^* denote a (near-)optimal set of network parameters (from now on we drop the * as well as the $_0$ on the target distribution $p_0(\mathbf{x})$).

In a first attempt, one may simply try to integrate the score to obtain $p(\mathbf{x})$, up to the normalisation. However, there is no guarantee that the score is conservative. Indeed, in general one finds that the score contains both a conservative and a non-conservative component (see below),

$$\mathbf{s}_\theta(\mathbf{x}, t) = \nabla\Phi_\theta(\mathbf{x}, t) + \mathbf{r}_\theta(\mathbf{x}, t), \qquad \nabla \cdot \mathbf{r}_\theta(\mathbf{x}, t) = 0. \tag{3.5}$$

This fact is known in the ML literature, see e.g. Refs. [30–32], although it is usually not emphasised.

If one insists on learning the data distribution directly, it is more useful to turn to energy-based diffusion models [31–33] and learn instead the energy,

$$E_\theta(\mathbf{x}, t) \approx -\log p(\mathbf{x}, t). \tag{3.6}$$

We follow Ref. [33] and write the energy function as

$$E_\theta(\mathbf{x}, t) = \frac{1}{2}\left\| \mathbf{v}_\theta(\mathbf{x}, t) \right\|^2. \tag{3.7}$$

The score is conservative by construction,

$$\partial_i \log p(\mathbf{x}, t) \approx -\partial_i E_\theta(\mathbf{x}, t) = -\mathbf{v}_\theta(\mathbf{x}, t) \cdot \partial_i \mathbf{v}_\theta(\mathbf{x}, t), \tag{3.8}$$

and is used in the loss function (3.3) during the training process. Once the energy is determined, the distribution follows immediately,

$$p(\mathbf{x}) \approx \lim_{t\to 0} \exp\left[-E_\theta(\mathbf{x}, t)\right], \tag{3.9}$$

up to the overall normalisation. Some comments are in order. First, the choice of energy parametrisation is not unique. Moreover, the need for additional derivatives makes this method more expensive than the score-based approach. Finally, even though in the latter the score is not conservative, this does not affect the effectiveness of score-based diffusion models. For details of the numerical implementation, we refer to Ref. [7].

4. Complex-valued quartic model

We demonstrate this approach in a simple but well-studied example, for which CL dynamics is known to converge correctly, namely the quartic model with a complex mass parameter [34],

$$S = \frac{1}{2}\sigma_0 x^2 + \frac{1}{4}\lambda x^4, \qquad \sigma_0 = A + iB. \tag{4.1}$$

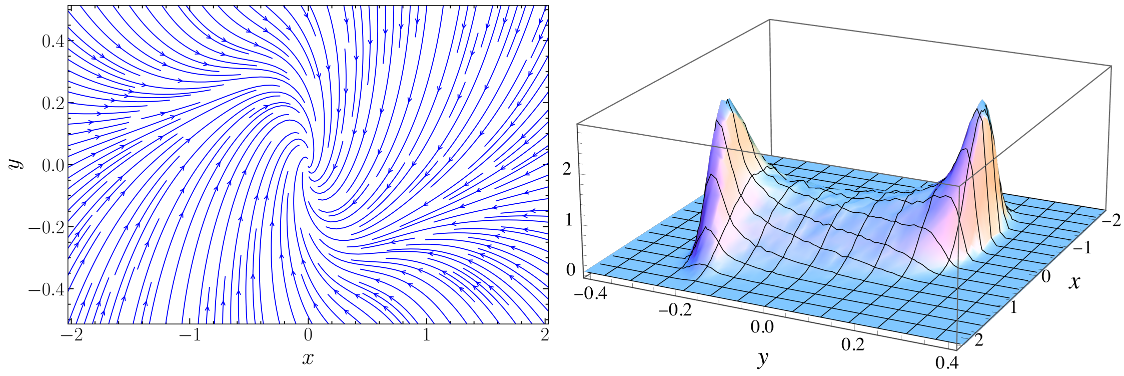

Figure 1: Complex-valued quartic model with parameters $\sigma_0 = 1 + i$ and $\lambda = 1$: CL drift in the complex plane (left) and histogram $p(x, y)$ obtained by sampling the CL process (right) [34].

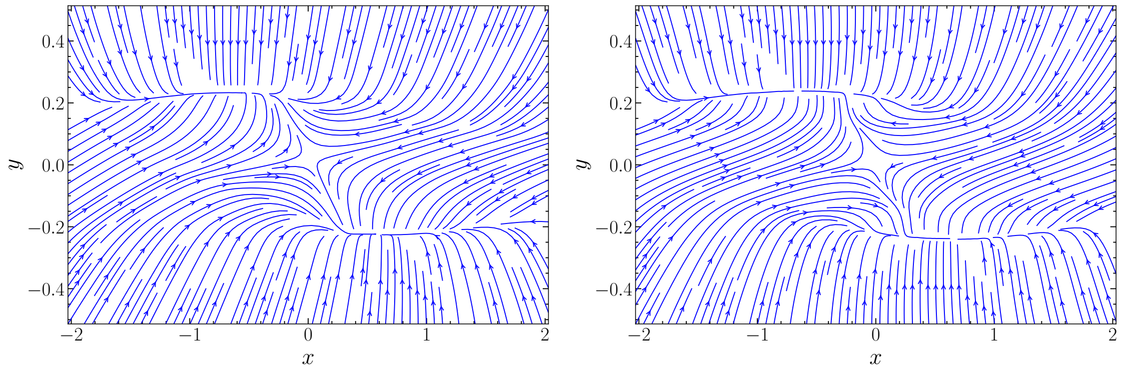

Figure 2: Learned scores in the quartic model at the end of the backward process, using a score-based (left) and energy-based (right) diffusion model. The drift on the left (right) is not conservative (conservative).

CL dynamics converges correctly when $3A^2 - B^2 \geq 0$: since the CL process then takes place in a strip $-y_- < y < y_-$, with

$$y_-^2 = \frac{A}{2\lambda}\left(1 - \sqrt{1 - \frac{B^2}{3A^2}}\right), \qquad y_- > 0, \tag{4.2}$$

the criteria for correctness apply [28, 29]. For more details on this aspect we refer to Ref. [34]. Below we take $A = B = \lambda = 1$, for which $y_- = 0.3029$. In Fig. 1 (left) the CL drift is shown. The boundaries at $|y| = y_-$ are visible: all arrows point inward and hence the process cannot escape the strip $|y| < y_-$. The CL process is solved using a higher-order discretisation, see Ref. [7] for details. Collecting the data in a histogram yields the 'empirical' distribution shown in Fig. 1 (right). The distribution is zero when $|y| > y_-$, consistent with the argument based on the drift. It is this distribution that we aim to learn.

We have trained the diffusion models using the CL data as the training set. In Fig. 2 we show the learned scores at the end of the backward process in the score-based (left) and energy-based (right) formulations. Even though they look very similar, the one on the left is not conservative, as demonstrated in Fig. 3 (left), where the curl of the score, $\partial_x s_{y,\theta}(\mathbf{x}) - \partial_y s_{x,\theta}(\mathbf{x})$, is shown, averaged over 10 independently trained models. The score on the right is conservative by construction.

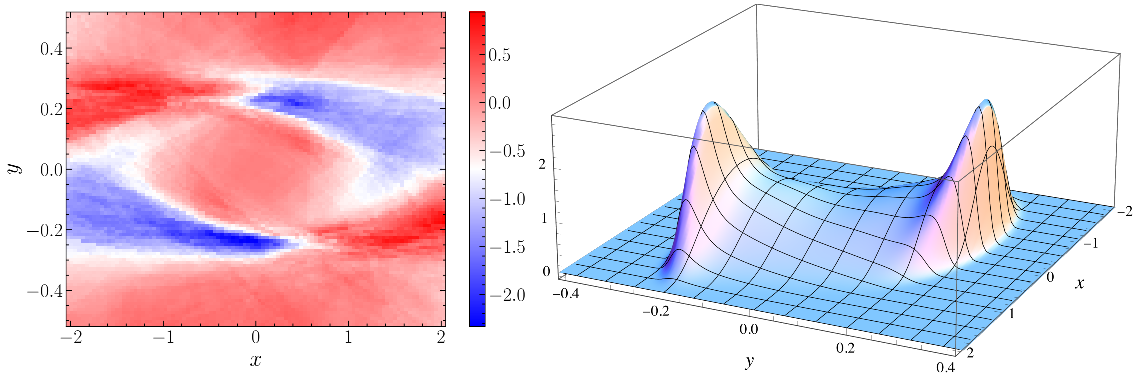

Figure 3: Left: curl of the score in the score-based formulation, averaged over 10 independently trained models. Right: distribution $p_\theta(\mathbf{x}) \sim \exp[-E_\theta(\mathbf{x})]$ in the energy-based formulation.

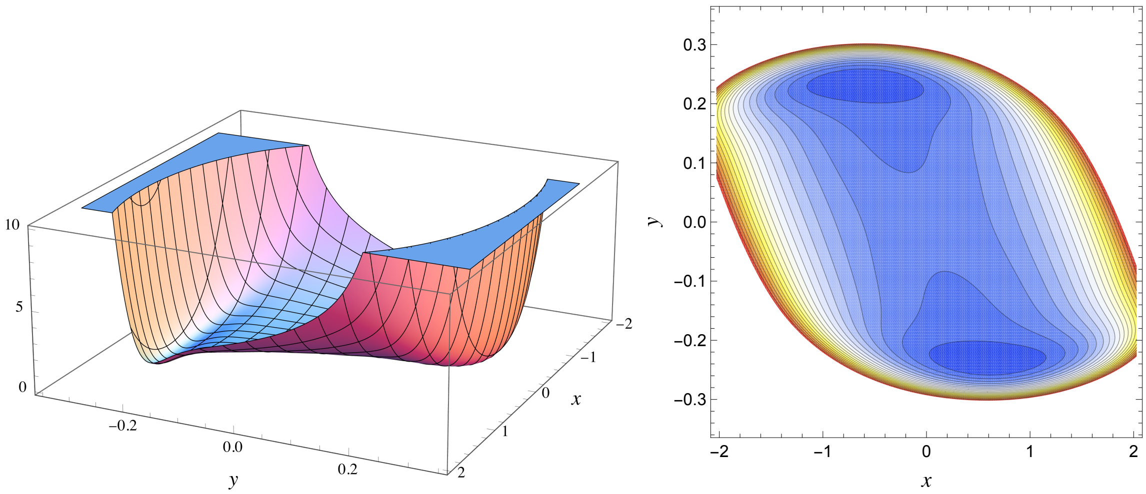

Figure 4: Energy $E_\theta(\mathbf{x})$ learnt in the energy-based model, in a three-dimensional plot and as a contour plot.

Before moving on to the distribution itself, we note that the learned scores in Fig. 2 are very different from the CL drift of Fig. 1 (left). There are a number of reasons for this. First recall that the CL drift is used in the CL equation with real noise only, whereas in the diffusion models noise is applied in both directions. Second, the CL drift is not integrable, since $\partial_x K_y(\mathbf{x}) \neq \partial_y K_x(\mathbf{x})$, and has an attractive fixed point at the origin. Instead, in the diffusion models the origin is a saddle point, while the two peaks, at $\mathbf{x}_{\text{peak}} \sim \pm(0.6, -0.25)$, are attractive. In fact, the scores in the diffusion models are closer to the flow relevant for Lefschetz thimbles [35].

In the energy-based formulation, the energy $E_\theta(\mathbf{x})$ is learnt directly. In Fig. 4 we show the energy, as a three-dimensional plot, restricted to $E_\theta(\mathbf{x}) \leq 10$, and using a contour plot. We note the two (shallow) minima around $\mathbf{x}_{\text{peak}}$ and a rapidly increasing energy outside the main region of interest. We now have direct access to the distribution, simply by exponentiating the learned energy. The result is shown in Fig. 3 (right). The undetermined prefactor is determined by normalising the distribution to 1. The distribution indeed resembles the empirical one constructed in Fig. 1 (right). We emphasise that this is the first time a parametrisation of the distribution sampled by a CL process is obtained in a non-trivial case without explicitly constructing a histogram. A quantitative

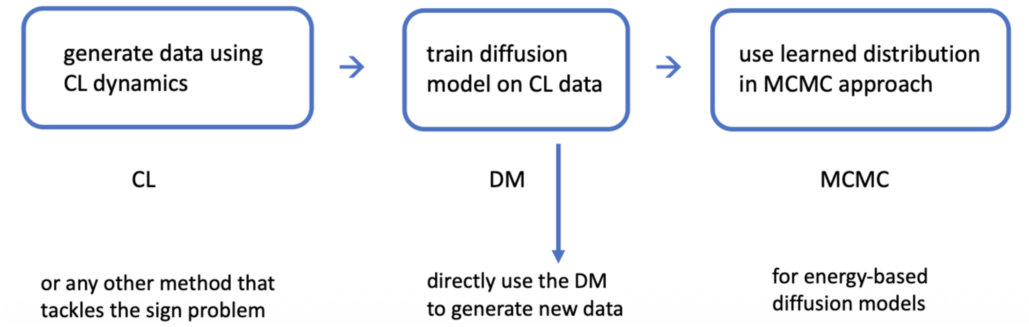

Figure 5: Flow chart indicating a three-stage approach to move from theories with a sign problem to MCMC sampling. It is assumed that the first stage, represented here by complex Langevin dynamics, gives a correct representation of the underlying theory.

comparison between the CL data and results obtained with the score-based and energy-based diffusion models can be found in Ref. [7].

We can now make the final step. Having learnt the energy, we can forget its origin and directly sample from the corresponding distribution, using a Metropolis-Hastings Markov Chain Monte Carlo (MCMC) simulation with the learned energy in the accept/reject criterion. This setup is outlined in the flow chart in Fig. 5. This third step was not feasible before, as the probability distribution was not expressed in terms of an energy function $E_\theta(\mathbf{x})$, which can be evaluated at arbitrary points within the data domain. Of course, this approach depends crucially on the first step providing a reliable approach to generating data in theories with a sign problem. Hence the application of the diffusion model does not solve the sign problem by itself. Instead, it provides a way to consider the equality in Eq. (1.1), with the distribution on the RHS learnt from data. For a comparison of the results obtained using the Metropolis-Hastings MCMC simulation as outlined here with the other approaches, we refer to Ref. [7].

Finally, for completeness we also note that systematic uncertainties in the diffusion models can be estimated by creating an ensemble of N independently trained models, with the same set of (near-)optimal hyperparameters. We found that the systematic uncertainty dominates, reflecting the stochastic nature of the training procedure [7].

5. Summary and outlook

We have demonstrated that diffusion models can be used to study the distribution sampled by a complex Langevin process for theories with a sign problem. Since this distribution is typically not known *a priori*, this provides a new perspective on understanding CL dynamics and complex-valued distributions more generally. Both score-based and energy-based diffusion models can learn the score, i.e., the log derivative of the distribution, and subsequently be used to generate additional configurations. In score-based models, the learned score is not conservative, as this is not imposed. The score can then not be integrated to yield the distribution. In energy-based models, it is conservative by construction. Hence, if the interest is in the distribution itself and not just in the generation of additional samples, energy-based models are preferred.

With direct access to the (unnormalised) distribution, a number of directions can be explored. One of them is a detailed study of properties of the distribution in regions where CL encounters problems. The direction we pursued here is to use the learned energy in an MCMC simulation. In this setup, CL is used to provide training data and the diffusion model is used to learn the energy. Afterwards, an MCMC simulation is used to generate additional configurations, without reference to the original CL process or diffusion model. To verify the results, it may be useful to compute the configurational temperature [36, 37]. Of course such an approach is only useful when the original CL process converges correctly; as stated, the diffusion models as employed here will not solve the sign problem when CL fails. So far we have explored these aspects in a simple model. The obvious next step is to extend this to field theories.

Acknowledgements – GA thanks KZ and his group for the kind hospitality at CUHK-Shenzhen during the completion of this work. This visit was supported in part by the Royal Society International Exchanges 2024 Global Round 2 IES\R2\242215. GA is further supported by STFC Consolidated Grant ST/X000648/1 and by a Royal Society Leverhulme Trust Senior Research Fellowship. DEH is supported by the UKRI AIMLAC CDT EP/S023992/1. LW thanks the DEEP-IN working group at RIKEN-iTHEMS for its support. LW is also supported by the RIKEN TRIP initiative (RIKEN Quantum), JSPS KAKENHI Grant No. 25H01560, and JST-BOOST Grant No. JPMJBY24H9. KZ is supported by the CUHK-Shenzhen university development fund under grant No. UDF01003041 and UDF03003041, Shenzhen Peacock fund under No. 2023TC0179 and NSFC grant under No. 92570117. We acknowledge the support of the Supercomputing Wales and AccelerateAI projects, which are part-funded by the European Regional Development Fund (ERDF) via Welsh Government.

Research Data and Code Access – The code and data are available in Ref. [38].

References

[1] P. de Forcrand, *Simulating QCD at finite density*, *PoS* **LAT2009** (2009) 010 [1005.0539].

[2] G. Aarts, *Introductory lectures on lattice QCD at nonzero baryon number*, *J. Phys. Conf. Ser.* **706** (2016) 022004 [1512.05145].

[3] G. Parisi, *On complex probabilities*, *Physics Letters B* **131** (1983) 393.

[4] J.R. Klauder, *Stochastic quantization*, in *Recent Developments in High-Energy Physics*, H. Mitter and C.B. Lang, eds., (Vienna), pp. 251–281, Springer Vienna, 1983.

[5] AuroraScience collaboration, *New approach to the sign problem in quantum field theories: High density QCD on a Lefschetz thimble*, *Phys. Rev. D* **86** (2012) 074506 [1205.3996].

[6] A. Alexandru, G. Basar, P.F. Bedaque, G.W. Ridgway and N.C. Warrington, *Sign problem and Monte Carlo calculations beyond Lefschetz thimbles*, *JHEP* **05** (2016) 053 [1512.08764].

[7] G. Aarts, D.E. Habibi, L. Wang and K. Zhou, *Combining complex Langevin dynamics with score-based and energy-based diffusion models*, *JHEP* **12** (2025) 160 [2510.01328].

[8] D. Weingarten, *Complex probabilities on* $\mathbb{R}^N$ *as real probabilities on* $\mathbb{C}^N$ *and an application to path integrals*, *Phys. Rev. Lett.* **89** (2002) 240201 [quant-ph/0210195].

[9] L.L. Salcedo and E. Seiler, *Schwinger–Dyson equations and line integrals*, *J. Phys. A* **52** (2019) 035201 [1809.06888].

[10] J. Sohl-Dickstein, E.A. Weiss, N. Maheswaranathan and S. Ganguli, *Deep unsupervised learning using nonequilibrium thermodynamics*, in *Proc. 32nd Int. Conf. Int. Conf. Mach. Learn. - Vol. 37*, pp. 2256–2265, 2015 [1503.03585].

[11] J. Ho, A. Jain and P. Abbeel, *Denoising diffusion probabilistic models*, in *Proc. 34th Int. Conf. Neural Inf. Process. Syst.*, pp. 6840–6851, 2020 [2006.11239].

[12] Y. Song and S. Ermon, *Generative Modeling by Estimating Gradients of the Data Distribution*, 1907.05600.

[13] 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.

[14] C.-H. Lai, Y. Song, D. Kim, Y. Mitsufuji and S. Ermon, *The Principles of Diffusion Models*, 2510.21890.

[15] G. Parisi and Y.S. Wu, *Perturbation theory without gauge fixing*, *Sci. China, A* **24** (1980) 483.

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

[17] K. Fukushima and S. Kamata, *Stochastic quantization and diffusion models*, 2411.11297.

[18] L. Wang, G. Aarts and K. Zhou, *Generative Diffusion Models for Lattice Field Theory*, in *37th Conference on Neural Information Processing Systems*, 2023 [2311.03578].

[19] G. Aarts, D.E. Habibi, L. Wang and K. Zhou, *On learning higher-order cumulants in diffusion models*, *Mach. Learn. Sci. Tech.* **6** (2025) 025004 [2410.21212].

[20] Q. Zhu, G. Aarts, W. Wang, K. Zhou and L. Wang, *Diffusion models for lattice gauge field simulations*, in *38th conference on Neural Information Processing Systems*, 2024 [2410.19602].

[21] Q. Zhu, G. Aarts, W. Wang, K. Zhou and L. Wang, *Physics-Conditioned Diffusion Models for Lattice Gauge Theory*, 2502.05504.

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

[23] G. Kanwar and O. Vega, *Spectral Diffusion for Sampling on* SU(N), in *42th International Symposium on Lattice Field Theory*, 2025 [2512.19877].

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

[25] M. Favoni, A. Ipp, D.I. Müller and D. Schuh, *Lattice Gauge Equivariant Convolutional Neural Networks*, *Phys. Rev. Lett.* **128** (2022) 032003 [2012.12901].

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

[27] D.E. Habibi, G. Aarts, L. Wang and K. Zhou, *Diffusion models learn distributions generated by complex Langevin dynamics*, *PoS* **LATTICE2024** (2025) 039 [2412.01919].

[28] G. Aarts, E. Seiler and I.-O. Stamatescu, *Complex Langevin method: When can it be trusted?*, *Phys. Rev. D* **81** (2010) 054508.

[29] G. Aarts, F.A. James, E. Seiler and I.-O. Stamatescu, *Complex Langevin: Etiology and Diagnostics of its Main Problem*, *Eur. Phys. J. C* **71** (2011) 1756 [1101.3270].

[30] S. Saremi, *On approximating ∇f with neural networks*, 1910.12744.

[31] T. Salimans and J. Ho, *Should EBMs model the energy or the score?*, in *International Conference on Learning Representations (ICLR 2021)*, 2021, https://openreview.net/forum?id=9AS-TF2jRNb.

[32] C. Horvat and J.-P. Pfister, *On gauge freedom, conservativity and intrinsic dimensionality estimation in diffusion models*, 2402.03845.

[33] Y. Du, C. Durkan, R. Strudel, J.B. Tenenbaum, S. Dieleman, R. Fergus, J. Sohl-Dickstein, A. Doucet et al., *Reduce, Reuse, Recycle: Compositional Generation with Energy-Based Diffusion Models and MCMC*, in *International Conference on Learning Representations (ICLR 2023)*, 2023 [2302.11552].

[34] G. Aarts, P. Giudice and E. Seiler, *Localised distributions and criteria for correctness in complex Langevin dynamics*, *Annals Phys.* **337** (2013) 238 [1306.3075].

[35] G. Aarts, *Lefschetz thimbles and stochastic quantization: Complex actions in the complex plane*, *Phys. Rev. D* **88** (2013) 094501 [1308.4811].

[36] A. Joseph and A. Kumar, *Thermodynamic Diagnostics for Complex Langevin Simulations: The Role of Configurational Temperature*, 2509.08287.

[37] A. Joseph and A. Kumar, *Configurational Temperature as a Diagnostic for Complex Langevin Dynamics in the 3D XY Model*, 2509.13314.

[38] D.E. Habibi, G. Aarts, L. Wang and K. Zhou, *Combining complex langevin dynamics with score-based and energy-based diffusion models - data release & workflow*, Nov., 2025. 10.5281/zenodo.17725665.

PoS(LATTICE2025)050