PROCEEDINGS OF SCIENCE

Generative sampling with physics-informed kernels

Friederike Ihssen,^{a} **Renzo Kapust**^{a,*} **and Jan M. Pawlowski**^{a,b}

^{a}*Institut für Theoretische Physik, Universität Heidelberg, Philosophenweg 16, 69120 Heidelberg, Germany*

^{b}*ExtreMe Matter Institute EMMI, GSI, Planckstr. 1, 64291 Darmstadt, Germany*

E-mail: `kapust@thphys.uni-heidelberg.de`

We construct a generative network for Monte-Carlo sampling in lattice field theories and beyond, for which the layerwise propagation is computed and optimised independently on each layer. The architecture uses physics-informed renormalisation group flows that provide access to the layerwise propagation step from one layer to the next in terms of a simple first order partial differential equation for the respective renormalisation group kernel through a given layer. Thus, it transforms the generative task into that of solving once the set of independent and linear differential equations for the kernels of the transformation. As these equations are analytically known, the kernels can be refined iteratively. This allows us to structurally tackle out-of-domain problems generally encountered in generative models and opens the path to further optimisation. We illustrate the practical feasibility of the architecture within simulations in scalar field theories.

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

---

[*] Speaker

https://pos.sissa.it/

PoS(LATTICE2025)055

1. Introduction

Efficiently computing high-dimensional integrals poses a central task in lattice field theory calculations. This task is commonly conducted with Markov chain Monte Carlo approaches, where configurations are drawn from the respective Boltzmann distribution

$$p(\varphi) = \frac{1}{\mathcal{N}} e^{-\hat{S}(\varphi)} = e^{-S(\varphi)} \qquad \text{with} \qquad S(\varphi) = \hat{S}(\varphi) + \log \mathcal{N} \,, \tag{1}$$

where we absorbed the normalisation $\mathcal{N}$ into the action $S(\varphi)$. However, as the phenomena of critical slowing down and topological freezing show, traditional sampling algorithms commonly lose their efficiency as one approaches the continuum limit of the theory [1–3]. In these scenarios, the autocorrelation time of the Markov chains diverge, requiring prohibitively large amounts of computational resources. These problems may be overcome within generative models. These models aim to find transformations that move configurations from a theory with efficient sampling to configurations of the targeted theory, see e.g. [4–7].

However, generative Machine Learning models for Monte-Carlo simulations of lattice field theories face two out-of-domain (OOD) problems that hamper progress in this area. The most obvious one is related to the high dimensionality of the transformation between the two theories, which grows with the number of lattice sites and typically renders the learned distribution inaccurate. Roughly speaking, this originates in the fact that interpolation problems in high dimensions are equally ill-conditioned as extrapolation problems in low dimensions. Consequently, the computational costs of the final sampling steps with the true distribution scale badly with the dimension. This is a manifestation of the curse of dimensionality.

The second problem is a standard extrapolation problem and may be explained as follows: assume that the generative model has been trained with a given samples size related to a limited information content: the sample size is in one-to-one correspondence to the access to a specific set of moments or cumulants of the distribution with the orders $n < N_{\text{max}}$, where N_{max} increases with the sample size but scales very badly. Further, sampling/training gives access to higher order moments/cumulants with $n > N_{\text{max}}$. However, the information carried by the higher order moments, or rather their irreducible parts, is not contained in the lower ones. Without any additional structure in the moments of the theory, this information is provided by the final accept-reject step with the true distribution which hence is increasingly expensive. We coin this OOD problem, the curse of limited information. Together, these two OOD problems lead to the bad scaling properties in the limit of large dimensions observed in many generative models.

Hence, the general situation asks for the construction of generative models or sampling algorithms for which the problems above are solved constructively rather than hoping for an underlying simplicity of the considered theory. Here, we describe a generative architecture based on *physics-informed kernels* (PIKs) put forward in [8], based on physics-informed renormalisation group flows (PIRGs) [9].

The PIRG approach allows the construction of infinitesimal field transformations along the RG-time t governed by a set of kernels $\dot{\phi}_t$ on the basis of the PIRG pair $(S_t, \dot{\phi}_t)$. The kernels can be directly computed from the respective RG equation for the pair. The key point of the PIRG in this approach is to analytically fix the action S_t along this transformation. Then, the underlying renormalisation group equation is a simple linear differential equation for the kernel $\dot{\phi}_i$ at each time step. Importantly, this differential equation does not contain information from the previous times, thus completely decoupling the propagation steps from each other. This possible trivial parallelisation as well as the nature of the analytically known differential equation for the kernel at a given action is aimed to structurally resolve both of the OOD problems mentioned in the beginning.

In the present work, we detail this approach based on PIKs in Section 2, where it is argued that PIKs allows for the construction of 'truly' generative models. The conceptional developments are showcased within a zero-dimensional lattice field theory. In Section 4, we conclude with a summary and an outlook of the many interesting further applications of the PIK-architecture.

2. Sampling architecture with physics-informed kernels

Renormalisation group flows describe the scale dependence and general reparametrisations of a given statistical distribution. Both aspects are governed by the infinitesimal change of the distribution with the RG-time t. For the present purpose, we take the RG-time in a normalised time interval $t \in I$ with $I = [0, 1]$. So, instead of the field φ in (1), we consider fields ϕ_t that are related to φ with a general non-linear transformation. This transformation comes with a change of the distribution $p_t(\phi)$. The respective differential equation is either derived for the normalised distribution $p_t(\phi)$ or its Laplace or Fourier transform,

$$Z(J) = \int D\phi \, p_t(\phi) \, e^{\phi_i J_i} \,, \qquad \text{with} \qquad D\phi = \prod_{i\in\mathcal{D}} \int_{\mathcal{T}_t} d\phi_i \,, \tag{2}$$

where $\mathcal{D}$ denotes the lattice with lattice sites labeled by i where the field value ϕ_i on that site takes values in $\mathcal{T}_t$. Here as for the remainder of this work, we use the Einstein sum convention. The most general scale and reparametrisation RG transformation is accommodated by the Wegner flow [10],

$$\frac{dp_t(\phi)}{dt} = -\frac{\partial}{\partial\phi_i} \left[\dot{\phi}_{t,i}(\phi) p_t(\phi) \right] . \tag{3}$$

For related work in generative models and on its relation to optimal transport, see e.g. [11]. The field transformation related to (3) reads

$$\frac{d\phi_{t,i}(\phi)}{dt} = \dot{\phi}_{t,i}(\phi) \,. \tag{4}$$

Here, the vector field $\dot{\phi}_t$ denotes the kernel of the transformation at the time t. The flow in (3) is a total derivative and hence, in the absence of boundary terms, leaves the total distribution $Z(J = 0) = 1$ invariant.

A key idea underlying the physics-informed RG flows [9] is to exploit the full generality of the Wegner flow (3) for the distribution or related RG equations for other generating functions: Instead of solving the flows for the distribution $p_t(\phi)$, the rate function $\Gamma_t(\phi)$ or other generating functions F_t, we keep the coordinates ϕ general and solve the equations for pairs

$$\left(F_t(\phi), \dot{\phi}_t(\phi) \right) . \tag{5}$$

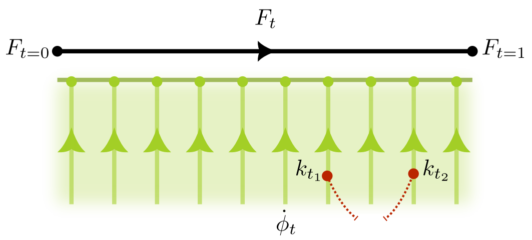

Figure 1: Setup of the PIK-architecture, with fully accessible the path of F_t (black) and kernels $\dot\phi_t$ (green) with completely independent parameters k_t (red).

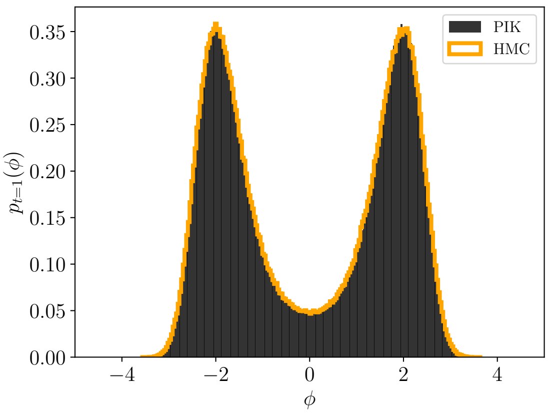

Figure 2: Comparison of the targeted (orange, HMC) and modelled (black, PIK) zero-dimensional ϕ^4-theory distribution.

As mentioned in Section 1, the key novel feature of the architecture proposed here is a structural property of the physics-informed RG flows: If the generating function $F_t \in \{p_t, \ldots\}$ is fully specified and known, the associated map $t \to F_t$ is completely independent of the previous ones,

$$\dot\phi_t(\phi) = \dot\phi_t(\phi; F_t)\,. \tag{6}$$

Put differently, the kernel $\dot\phi_t$ of the map is *physics-informed*. This perspective has also been used in [12], where (6) is solved within an architecture with a global loss function. In the present work, we use that (6) leaves us with independent tasks of solving $\dot\phi_t$ for the whole time line from $t = 0$ to $t = 1$, which can be trivially parallelised. This is depicted in Figure 1, where $\dot\phi_t$ with parameters k_t is computed vertically (green) to the evolution of F_t (black).

Focusing on flows of the distribution $p_t(\phi) = e^{-S(\phi)}$, we consider the situation where we parametrised a path for the action S_t which connects the theory of targeted action S_1 with a more tractable one, S_0. For the action, the Wegner equation can easily be rewritten as

$$\frac{dS_t(\phi)}{dt} = \left[\frac{\partial}{\partial\phi_i} - \frac{\partial S_t(\phi)}{\partial\phi_i}\right]\dot\phi_{t,i}(\phi)\,. \tag{7}$$

For a given S_t, this is a simple linear first order partial differential equation (PDE) for the physics-informed kernel $\dot\phi_t$ that has to be solved in order to transform the fields according to (4).

As indicated by (1), the action given as $S_t(\phi) = \hat S_t(\phi) + \log\mathcal{N}_t$. Here, $\hat S_t(\phi)$ contains all field-dependent parts and $\log\mathcal{N}_t$ is field-independent and ensures normalisation. We parametrise

$$\hat S_t(\phi) = \sum_i c_{i,t} O_{i,t}(\phi)\,, \tag{8}$$

with scale-dependent coefficients $c_{i,t}$ and a set of analytically known basis functions $O_{i,t}$, which in principle can also be chosen time dependent. For the present work, we drop the operator's explicit time-dependence, for a short discussion see [8]. At the initial and final scale, $\{c_i, O_i\}$ are chosen such that the action matches to the known base and target distributions. In between, the only constraint for the trajectory is its differentiability, and the ability to compute the field-derivatives $\partial S_t(\phi)/\partial\phi_i$. However, secondary constraints can be used for the optimisation of the sampling process, e.g. learning optimal base distributions.

The final ingredient in the PIK-architecture is the normalisation of the flowing distribution p_t. For PIKs, the flow of the normalisation $\mathcal{N}_t$ can be directly computed from that of $\hat{S}_t$ by

$$\frac{d\log\mathcal{N}_t}{dt} = -\int D\phi\, p_t(\phi)\,\frac{d\hat{S}_t(\phi)}{dt}\,. \tag{9}$$

Actually computing (or approximating) this constant to the required precision is a complicated task which has to be avoided in practice. To this end, we rewrite the Wegner equation (7) as

$$\frac{d\log\mathcal{N}_t}{dt} = \left[\frac{\partial}{\partial\phi_i} - \frac{\partial\hat{S}_t(\phi)}{\partial\phi_i}\right]\dot{\phi}_{t,i}(\phi) - \frac{d\hat{S}_t(\phi)}{dt}\,. \tag{10}$$

Equation (10) allows us to compute the flow of the normalisation from the right hand side. Importantly, this circumvents the expensive sampling task (9) and is key to the efficient sampling in the PIK-architecture. Furthermore, the kernel $\dot{\phi}$ can be computed by considering differences of Wegner equations for different field configurations ϕ and χ. This leaves us with

$$\frac{d\hat{S}_t(\phi)}{dt} - \frac{d\hat{S}_t(\chi)}{dt} = \left[\frac{\partial}{\partial\phi_i} - \frac{\partial\hat{S}_t(\phi)}{\partial\phi_i}\right]\dot{\phi}_t(\phi) - \left[\frac{\partial}{\partial\phi_i} - \frac{\partial\hat{S}_t(\phi)}{\partial\phi_i}\right]\dot{\phi}_t(\phi)\Big|_{\phi=\chi} =: \mathcal{L}_\chi\dot{\phi}_t(\phi)\,, \tag{11}$$

where we introduced the linear operator $\mathcal{L}_\chi$ for notational convenience. Note that here all terms are analytically known and can be directly computed. Hence, one must not estimate or learn the change in the normalisation [12, 13] but obtains an efficient condition for the kernel $\dot{\phi}_t$.

In the following we consider solution strategies for the Wegner equation (3), based on a parametrisation of the kernel $\dot{\phi}_t$ in some set of basis elements $\{K_{j,t}\}_{j=1}^M$,

$$\dot{\phi}_t(\phi) = \sum_{j=1}^{M} k_{j,t}K_{j,t}(\phi)\,, \tag{12}$$

with the respective set of expansion coefficients $\{k_{j,t}\}_{j=1}^M$. With (11), this reduces the complicated problem of finding a transformation between two distributions to the problem of solving an independent set of linear systems of the form $A_t k_t = b_t$. Here, the matrix A_t and vector b_t are directly determined by the chosen basis $\{K_{j,t}\}_{j=1}^M$ and the parametrised S_t.

We conclude this outline of the PIK-architecture with highlighting three key properties that follow from its structure and discuss them in the subsequent sections.

(i) *Independent Kernels:* PIKs convert the global task of finding a precise invertible map from the base distribution p_0 to the final distribution p_1 within a deep network into independent tasks of computing the infinitesimal maps $\dot{\phi}_t(\phi; F_t)$ at each time step. For the pair $(S_t, \dot{\phi}_t)$ this amounts to solving (7) for $\dot{\phi}_t$.

(ii) *OOD resolution:* The Wegner equation (7) provides us with an error control of this map. This is also at the root of systematic improvements in the sampling process beyond the initial training time. In particular, curing the OOD problem with the PIRG pair $(S_t, \dot{\phi}_t)$ boils down to locally improving the solution $\dot{\phi}_t$ to a known linear differential equation, whose coefficients are determined by $\hat{S}_t$.

(iii) *Optimisation:* The properties (i) and (ii) are relevant for the setup and solution of PIKs for a given pair $(F_t, \dot{\phi}_t)$. The independent and analytically given structure of (7) offers further freedom of optimisation for the pair, given external constraints. In particular, this includes a new kind of parameter conditional flow, systematic ways to improve inaccurate solutions of (7) and the choice of the used generating function F_t.

3. Physics-informed kernels at work

We proceed with selected applications of the PIK-architecture to the commonly used benchmark system of a zero-dimensional ϕ^4-theory, illustrating the above three key properties.

3.1 Independent kernels

In order to illustrate the independent kernel structure of the PIK-architecture, we require a path for the action S_t that connects a simple base distribution to the targeted one. For the zero-dimensional ϕ^4-theory, an intuitive choice is given by

$$\hat{S}_t(\phi) = \frac{1}{2} m_t^2 \phi^2 + \frac{1}{4} \lambda_t \phi^4 \quad \text{with} \quad m_t^2 = m_0^2 + t\,(m_1^2 - m_0^2)\,, \quad \lambda_t = \lambda_0 + t\,(\lambda_1 - \lambda_0)\,. \tag{13}$$

Here, the mass m_t and coupling λ_t are chosen to be time-dependent, continuously connecting a base distribution with parameters (m_0, λ_0) to another theory with parameters (m_1, λ_1). More specifically, we use $(m_0^2 = 1,\ \lambda_0 = 0)$ and $(m_1^2 = -2,\ \lambda_1 = 1/2)$, which transports a Gaußian into a multimodal interacting theory. For multimodal theories, PIKs circumvent the problem of mode-collapse [14–16] by enforcing a particular path S_t, which fixes the way the unimodal distribution is morphed into the multimodal one.

During the flow, the field ϕ_t is propagated in time according to (4). This ODE in RG-time is solved numerically by a standard fourth order Runge-Kutta scheme with a step size of $\Delta t = 1/25$, leaving us with $N_t = 51$ independent kernels $\dot{\phi}_t$ to be determined. Each kernel is expressed in a basis with $K_{1,t} = \phi$ and $K_{j,t} = \sin(\omega_{j,t}\phi)$ for $j \geq 2$ and frequencies $\omega_{j,t}$, which respects the $\mathbb{Z}_2$ symmetry of the action [17, 18]. Using the efficient condition (11), the parameters of each kernel can be optimised independently without computing the change in the normalisation explicitly.

Figure 2 shows a comparison of the targeted (orange) and modelled (black) distribution at the final time $t = 1$. Here, the targeted distribution is obtained using 10^6 samples from a standard hybrid Monte Carlo algorithm (HMC). The modelled distribution was generating by pushing 10^6 samples from the initial Gaußian distribution through the PIK flow. We find an excellent agreement between both distributions.

3.2 OOD resolution

Having showcased that PIKs have the capacity to compute each kernel $\dot{\phi}_t$ independently without the requirement of computing the change in the normalisation, we know turn ot the property (ii) *OOD resolution*. Instead of the explicit volume scaling, which is part of ongoing research, we illustrate the systematic improveability of the PIK sampling process after its initial setup. This is naturally enabled by the fact that we have analytical access to the Wegner equation (7) and (11)

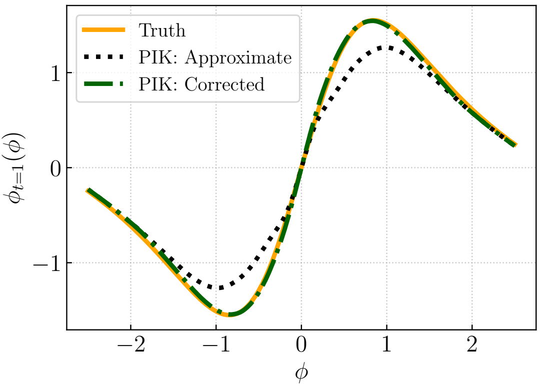

Figure 3: Comparison of the true (orange), the imperfect (black), and the corrected (green) kernel $\dot{\phi}_{t=1}$ for the zero-dimensional ϕ^4-theory.

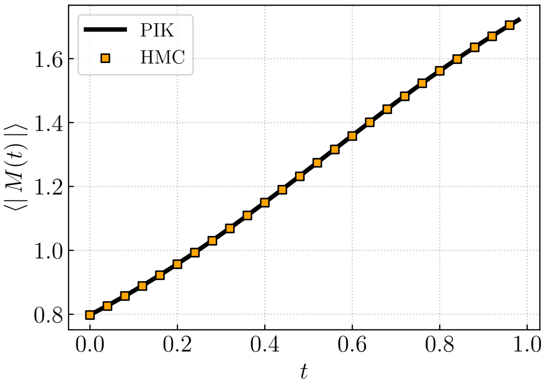

Figure 4: Magnetisation of the zero-dimensional ϕ^4-theory computed with the PIK-architecture in a continuous manner (black) and HMC with individual simulations (orange).

that has to be satisfied by each kernel individually. It is further boosted by the fact that the Wegner equation is linear in $\dot{\phi}_t$, making iterative improvements directly accessible.

In order to mimic the OOD problem of the curse of limited information, we purposefully only compute the final kernel on a dataset limited to $50k$ samples. This approximate kernel is referred to as $\dot{\phi}_t^{(0)}$. This kernel (black) together with the true solution (orange) is depicted in Figure 3. Note, that also for other continuous generative models, see e.g. [18–21], the Wegner equation at the last time step must be solved exactly (7) in order to be truly generative.

In order to systematically improve the sampling, we must improve the solution of the kernel $\dot{\phi}_t^{(0)}$ such that it provides a more accurate solution of the Wegner equation. There are many ways to achieve this goal. Here, we use the linear structure of (11). The residual of (11) with the approximate kernel $\dot{\phi}_t^{(0)}$ is then readily computed as

$$r_\chi^{(0)}(\phi) = \frac{d\hat{S}_t(\phi)}{dt} - \frac{d\hat{S}_t(\chi)}{dt} - \mathcal{L}_\chi \dot{\phi}_t^{(0)}(\phi)\,. \tag{14}$$

Naturally, a correction $\Delta\dot{\phi}_t^{(0)}$, that satisfies $\mathcal{L}_\chi \Delta\dot{\phi}_t^{(0)}(\phi) = r_\chi^{(0)}(\phi)$, will lead to a corrected kernel $\dot{\phi}_t^{(1)} = \dot{\phi}_t^{(0)} + \Delta\dot{\phi}_t^{(0)}$ that solves the Wegner equation (11) accurately. This procedure can also be iterated as the Wegner equation is linear in $\dot{\phi}_t$, hence the superscripts, indicating the first iteration.

In order not to spoil the well-fitted regions of the approximate kernel, we solve for the correction with a collocation approach using compactly supported radial basis functions (RBFs), more concretely antisymmetrised Wendland functions. As the RBFs have a compact support, the resultant linear system is sparse and readily solved.

In Figure 3, we show the corrected kernel $\dot{\phi}_t^{(1)}$ in green. We find that the correction indeed leads to a kernel that is very close to the true one and corrects the approximate kernel in the regions with the largest deviations. While the correction mechanisms can be readily improved, this example illustrates the general systematic workflow of PIKs from (a) an (imprecise) global solution to (b) a determination of OOD regions during sampling or using information of the residual to (c) a residual-driven correction of the kernel. This is natural for PDE approaches but the proposed method is to our knowledge novel in the context of generative models.

3.3 Optimisation

The above sections showcased the setting, computation, and systematic improvement within the PIK-architecture. In this section, we comment on the property (iii) *Optimisation* of PIKs. As outlined in Section 2, there are various ways to optimise the pair PIRG pair. These reach from the choice of the generating function to optimal paths for the same, improved base distributions and to the optimisation of the solution method of the flow equation. Here, we illustrate one such optimisation at the example of a parameter-conditional flow.

With PIKs we have direct control over the distribution at each time step. Here, we use this freedom to enforce the same functional form of the action at each time step. The only change of the distribution in the flow are the parameters (m^2, λ). This allows us to make PIKs parameter conditional without inflating the network, which is the usual strategy to make the parameter of the generative model conditional [18, 22]. Moreover, this directly allows us to compute observables continuously without sampling the configurations for each parameter.

We exemplify this PIK approach in Figure 4 for the magnetisation

$$\langle |M(t)| \rangle = \int D\phi \, p_t(\phi) \, |\phi| \, , \tag{15}$$

For the PIK-architecture (black), this observable can be computed by pushing 10^6 Gaußian samples from the initial distribution through the flow and computing the magnetisation at each intermediate step. For comparison, we also show the same observable computed with individual HMC simulations (orange), where each point corresponds to a separate simulation.

4. Summary and outlook

We have described physics-informed kernels (PIKs) as a new architecture for generative models as constructed in [8]. The PIK-architecture aims at solving OOD problems in generative models as well as implementing a systematic error control. At its heart, PIKs reduce the task of finding transformations between two distributions into a decoupled chain of layerwise transformations that are given by an analytically known renormalisation group equation. For the distribution itself, this is the Wegner equation (3), for the rate function it is the generalised flow equation [23], and similar ones exist for other generating functions. Viewed as PDEs for the kernels, these equations boil down to independent linear differential equations. Importantly, their solution can be systematically corrected at each time step without requiring additional steps in the Markov chain, which is different to most prominent generative models like normalising flows or diffusion models, see e.g. [6, 18–21, 24–26]. As a paradigmatic example, we have considered PIKs for the distribution with completely determined action paths $S_t(\phi)$ that lead to simple Wegner equations.

This setup paves the way for many interesting applications and improvements of the PIK-architecture. The first prominent direction is the application of PIKs to higher-dimensional field theories as well as the extension to field theories with gauge or fermionic degrees of freedom. Moreover, as the PIK-architecture enables general transformations of the measure it is tailor-made to tackle sign problems, where the action becomes complex. This is particularly relevant for real-time simulations and finite density field theories and is part of a forthcoming work [27]. Furthermore, the PIK-architecture enables further optimisations of the sampling process, e.g. by computing optimal paths for S_t, including optimal base distributions. Lastly, going beyond sampling, the PIK-architecture is also envisioned to infer the action or Hamiltonian in a symbolic way from data, which is also part of ongoing research.

Acknowledgments

We thank Gert Aarts, Kenji Fukushima, Sander Hummerich, Thore Kolja Joswig, Ullrich Köthe, Timoteo Lee, Manfred Salmhofer, Robert Scheichl, Lingxiao Wang and Kai Zhou for discussions and collaborations on related subjects. We acknowledge the DM-LFT Collaboration for stimulating discussions and continuous support throughout this work. It is funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under Germany's Excellence Strategy EXC 2181/1 - 390900948 (the Heidelberg STRUCTURES Excellence Cluster). We also acknowledge support by the state of Baden-Württemberg through bwHPC.

References

[1] U. Wolff, *CRITICAL SLOWING DOWN*, *Nucl. Phys. B Proc. Suppl.* **17** (1990) 93.

[2] B. Alles, 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`].

[3] L. Del Debbio, G.M. Manca and E. Vicari, *Critical slowing down of topological modes*, *Phys. Lett. B* **594** (2004) 315 [`hep-lat/0403001`].

[4] M. Luscher, *Trivializing maps, the Wilson flow and the HMC algorithm*, *Commun. Math. Phys.* **293** (2010) 899 [`0907.5491`].

[5] D.J. Rezende and S. Mohamed, *Variational Inference with Normalizing Flows*, `1505.05770`.

[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** (2019) 034515 [`1904.12072`].

[7] K.A. Nicoli, C.J. Anders, L. Funcke, T. Hartung, K. Jansen, P. Kessel et al., *Estimation of Thermodynamic Observables in Lattice Field Theories with Deep Generative Models*, *Phys. Rev. Lett.* **126** (2021) 032001 [`2007.07115`].

[8] F. Ihssen, R. Kapust and J.M. Pawlowski, *Generative sampling with physics-informed kernels*, `2510.26678`.

[9] F. Ihssen and J.M. Pawlowski, *Physics-informed renormalisation group flows*, *Annals Phys.* **481** (2025) 170177 [`2409.13679`].

[10] F.J. Wegner, *Some invariance properties of the renormalization group*, *J. Phys. C* **7** (1974) 2098.

[11] J. Cotler and S. Rezchikov, *Renormalizing Diffusion Models*, `2308.12355`.

[12] M.S. Albergo and E. Vanden-Eijnden, *NETS: A Non-Equilibrium Transport Sampler*, `2410.02711`.

[13] B. Máté and F. Fleuret, *Learning Interpolations between Boltzmann Densities*, `2301.07388`.

[14] D.C. Hackett, C.-C. Hsieh, S. Pontula, M.S. Albergo, D. Boyda, J.-W. Chen et al., *Flow-based sampling for multimodal and extended-mode distributions in lattice field theory*, `2107.00734`.

[15] S. Chen, O. Savchuk, S. Zheng, B. Chen, H. Stoecker, L. Wang et al., *Fourier-flow model generating Feynman paths*, *Phys. Rev. D* **107** (2023) 056001 [`2211.03470`].

[16] K.A. Nicoli, C.J. Anders, T. Hartung, K. Jansen, P. Kessel and S. Nakajima, *Detecting and mitigating mode-collapse for flow-based sampling of lattice field theories*, *Phys. Rev. D* **108** (2023) 114501 [`2302.14082`].

[17] J. Köhler, L. Klein and F. Noé, *Equivariant flows: Exact likelihood generative learning for symmetric densities*, `2006.02425`.

[18] 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`].

[19] R.T.Q. Chen, Y. Rubanova, J. Bettencourt and D. Duvenaud, *Neural Ordinary Differential Equations*, `1806.07366`.

[20] A. Bulgarelli, E. Cellini and A. Nada, *Scaling of stochastic normalizing flows in SU(3) lattice gauge theory*, *Phys. Rev. D* **111** (2025) 074517 [`2412.00200`].

[21] G. Aarts, D.E. Habibi, A. Ipp, D.I. Müller, T.R. Ranner, L. Wang et al., *Generalizable Equivariant Diffusion Models for Non-Abelian Lattice Gauge Theory*, `2601.19552`.

[22] A. Singha, D. Chakrabarti and V. Arora, *Conditional normalizing flow for Markov chain Monte Carlo sampling in the critical region of lattice field theory*, *Phys. Rev. D* **107** (2023) 014512 [`2207.00980`].

[23] J.M. Pawlowski, *Aspects of the functional renormalisation group*, *Annals Phys.* **322** (2007) 2831 [`hep-th/0512261`].

[24] 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`.

[25] Y. Lipman, R.T.Q. Chen, H. Ben-Hamu, M. Nickel and M. Le, *Flow matching for generative modeling*, `2210.02747`.

[26] A. Singha, E. Cellini, K.A. Nicoli, K. Jansen, S. Kühn and S. Nakajima, *Multilevel Generative Samplers for Investigating Critical Phenomena*, `2503.08918`.

[27] F. Ihssen, R. Kapust and J.M. Pawlowski, *Solving sign problems with physics-informed kernels*, *in preparation* (2026) .