

Machine learning for four-dimensional SU(3) lattice gauge theories

Urs Wenger^{a,*}

*^aInstitute for Theoretical Physics
Albert Einstein Center for Fundamental Physics
University of Bern
CH-3012 Bern
Switzerland
E-mail: wenger@itp.unibe.ch*

In this review I summarize how machine learning can be used in lattice gauge theory simulations and what approaches are currently available to improve the sampling of gauge field configurations, with a focus on applications in four-dimensional SU(3) gauge theories. These include approaches based on generative machine-learning models such as (stochastic) normalizing flows and diffusion processes, and an approach based on renormalization group (RG) transformations, more specifically the machine learning of RG-improved gauge actions using gauge-equivariant convolutional neural networks. In particular, I present scaling results for a machine-learned fixed-point action in four-dimensional SU(3) gauge theory towards the continuum limit. The results include observables based on the classically perfect gradient-flow scales, which are free of tree-level lattice artefacts to all orders, and quantities related to the static potential and the deconfinement transition.

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

*Speaker

1. Introduction and motivation

Machine learning is having an ever increasing impact on simulations of lattice gauge theories. In these proceedings I review some of the machine-learning applications with a focus on those enabling efficient generation of SU(3) gauge field configurations in four spacetime dimensions. I will concentrate the review on those generative machine-learning approaches which are new and either have the potential for or have already demonstrated promising scaling towards useful lattices.

In lattice gauge theories we typically consider partition functions of the form

$$Z(\beta) = \int \mathcal{D}U \exp\{-\beta A[U]\} \quad (1)$$

where $A[U]$ is the gauge action depending on the SU(N) gauge field U , $\beta = 2N/g^2$ is the (inverse) gauge coupling, and $\mathcal{D}U$ is the Haar integration measure. Expectation values for observables take the form

$$\langle \mathcal{O}_\xi \rangle_\beta = \frac{1}{Z(\beta)} \int \mathcal{D}U \exp\{-\beta A[U]\} \mathcal{O}_\xi[U] \quad (2)$$

where ξ denotes the characteristic physical length scale associated with the given observable \mathcal{O} . Expressed in units of the lattice spacing a the length scale ξ becomes a dimensionless quantity ξ/a which diverges towards the continuum limit $a \rightarrow 0$. The lattice spacing is determined by the gauge coupling β and for asymptotically free field theories, such as four-dimensional SU(N) gauge theory, the continuum limit $a \rightarrow 0$ is achieved by taking $\beta \rightarrow \infty$ or equivalently $g^2 \rightarrow 0$, as illustrated in Fig. 1. From the lattice perspective, the continuum limit $\xi/a \rightarrow \infty$ corresponds to a second-order (continuous) phase transition where the physical correlation length ξ diverges while the lattice spacing is constant. As a consequence, simulations of lattice field theories typically suffer from severe *critical slowing down* towards the continuum limit resulting in a dramatic increase of autocorrelation times. For lattice gauge theories, critical slowing down may manifest itself in *topological freezing*. In such a situation, the simulations are stuck in sectors of fixed topological charge, leading to meta-stabilities, non-ergodicity of the simulations and eventually a failure to reach the true equilibrium of the system.

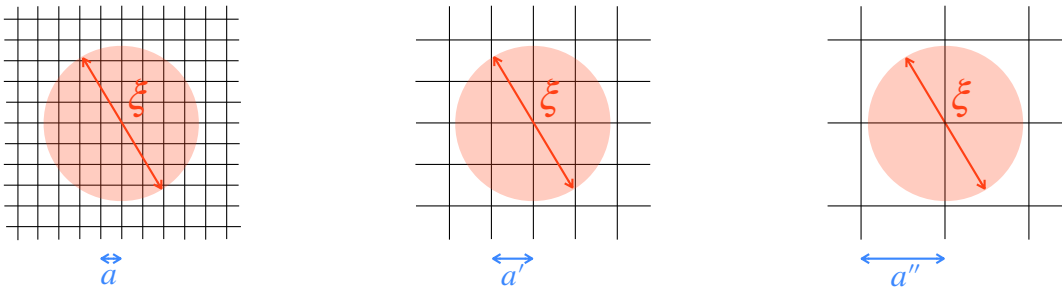


Figure 1: Illustration of the continuum limit for asymptotically free lattice field theories: the lattice spacing decreases from right to left, $a < a' < a''$, as the coupling decreases, $g < g' < g''$, or equivalently $\beta > \beta' > \beta''$ increases. In the limit $\beta \rightarrow \infty$ the lattice spacing $a \rightarrow 0$ vanishes, which for a fixed physical length scale ξ is equivalent to $\xi/a \rightarrow \infty$, i.e., the continuum limit is realized as a second-order (continuous) phase transition.

This provides the motivation for applying machine-learning methods in lattice gauge theories: they try to avoid—in one way or another—the critical slowing down encountered in Monte Carlo simulations towards the continuum limit. Taking stock one can identify two complementary directions, either to overcome critical slowing down by employing generative machine-learning models at fine lattice spacings, or to avoid critical slowing down by simulating at coarse lattice spacing using a machine-learned action with highly suppressed lattice artefacts based on renormalization group transformations (RGTs):

- **Generative machine-learning models:**
These approaches try to generate uncorrelated gauge field configurations at fine lattice spacings. Currently there are two different procedures available based on mapping gauge field configurations from a prior distribution, for which it is easy to draw uncorrelated samples, to the target distribution. The machine-learned maps are based on 1) reversible normalising flows, or 2) backward diffusion processes. A new method 3) is based on non-equilibrium Markov Chain Monte Carlo (NE-MCMC) and machine-learned stochastic normalising flows. I discuss these three approaches in some detail in in Secs. 2.1, 2.2 and 2.3, respectively.
- **Machine learning RGTs:**
In this approach one employs RGTs to relate fine to coarse lattices in order to construct effective RGT-improved lattice actions. Here, the idea is that uncorrelated lattice configurations can be generated at coarse lattice spacings, where there is no critical slowing down, while at the same time large lattice artefacts are avoided. Then the main challenge is to learn or parametrize the inverse RGT. I review this approach in Sec. 3.

Apart from these *generative* machine-learning approaches, on which this review is focused, there also exist machine-learning strategies to improve the evaluation of observables. One strategy tries to enhance the signal-to-noise ratios by employing, e.g., control variates, surrogate variables, or contour deformations, and optimizing them using machine-learning. These strategies have been reviewed and discussed in detail by Scott Lawrence in a plenary talk at last year's lattice conference [1]. An intriguing idea is to use the Feynman-Hellmann theorem and employ operator insertions, derivative observables and machine-learned normalizing flows which, in combination, can achieve a significant variance reduction [2]. Another interesting example is the exploration of gauge-fixing schemes using machine learning [3]. Let me also point out the efforts to optimise wave functions and ground-state operators. One example concerns the parametrization of interpolators for a static quark-antiquark pair, in order to optimize the overlap with the ground state [4]. Another example is the determination of the energy spectrum of $SU(N)$ gauge theories in the Hamiltonian formulation using physics-informed neural networks [5, 6]. The variety of these applications demonstrates the versatility of machine learning as a tool to improve and enhance simulations of lattice gauge theories.

2. Generative machine-learning models

2.1 Normalizing flows

The approach of using normalizing flows for generative machine-learning models, or more generally flow-based samplers, has been reviewed by Gurtej Kanwar in his plenary talk at the

lattice conference in 2023 [7], including an extensive discussion of the progress at the time and the prospects for the future. The review provided here is therefore kept very brief, yet it is interesting to understand what has happened in the meantime and what the current status is.

The general idea is to learn a map from a prior distribution of field configurations to the target one, either using a continuous or a discrete flow of the underlying field variables. More formally, given a (simple) prior distribution $r(U)$ on the gauge fields U , the (complicated) target distribution is $q(U') = q(f(U))$ where the diffeomorphism $f(U)$ is found and described by a machine-learning approach,

$$r(U) \xrightarrow{\text{ML}} q(U') = q(f(U)). \quad (3)$$

Apart from finding suitable gauge-equivariant maps, the main problem lies in the fact that the transformations require the computation of invertible Jacobians J_f and their determinants,

$$q(f(U)) = r(U) \left| \det \frac{\partial f(U)}{\partial U} \right|^{-1}. \quad (4)$$

The construction of computationally tractable transformations is usually achieved through a discrete set of coupling layers containing gauge-equivariant functions g_i , such that

$$f \equiv g_m \circ \dots \circ g_1 \quad (5)$$

with g_i being easily invertible. The Jacobian determinant of f can then be calculated efficiently by

$$\det J_f = \det J_{g_1} \cdot \dots \cdot \det J_{g_m}. \quad (6)$$

This can be accomplished, for example, by transforming only a subset of the degrees of freedom conditioned on the complementary subset. Most effectively, a simple triangular Jacobian is obtained by a suitable decoupling of the variables. The maps are self-trained by minimizing a target loss function based on the reverse Kullback-Leibler divergence

$$D_{KL}(q||p) \equiv \int \mathcal{D}U q(U) [\log q(U) - \log p(U)] \quad (7)$$

which is available because the gauge field action $S[U]$ and hence the exact target distribution $p(U)$ is explicitly known. (Recall that $q(U)$ is the machine-learned model output distribution and can be regarded as a variational parametrization ansatz for the target distribution.)

The symmetries of the gauge action can be taken into account by incorporating them into the flows. In particular, it is crucial to take gauge symmetry into account, such as in Ref. [8, 9] where this approach has been pioneered for U(1) and SU(N) gauge theories in two spacetime dimensions. For theories with dynamical fermions, flow-based sampling including pseudofermions has also been studied [10]. In Ref. [11] the framework has been extended to SU(3) gauge theory in four spacetime dimensions with the latest advancements reported in [12, 13]. New developments include architectural progress and the use of the correlated ensemble method [14].

Reviewing the developments over the last couple of years it is probably safe to state that the scaling of the normalizing-flow approach in the context of generative machine-learning models to four spacetime dimensions, SU(N) gauge theories, large volumes and fine lattice spacing all at the same time has turned out to be very challenging [15]. As a consequence, progress in this direction has somewhat slowed down and more recent normalizing-flow applications seem to be focusing on complementary directions [2, 12, 14].

2.2 Diffusion models

Generative diffusion models have become very popular in recent years and have enabled many exciting applications in text-to-image and text-to-video generation. For a discussion of the core principles I refer to the review [16] which also discusses the relation of score- or energy-based diffusion models with the flow-based models.

In a nutshell, generative diffusion models are based on adding noise to the degrees of freedom ϕ_0 , e.g., a gauge field $\phi_0 = U_0$ distributed according to a target distribution $p_0(U_0)$, using a stochastic differential equation. The stochastic evolution of the degrees of freedom in the forward process up to the fictitious time T produces new configurations $\phi_T = U_T$ distributed according to a simple, easy-to-sample prior distribution $p_T(U_T)$. The forward process can be reversed using a different, but related stochastic differential equation mapping p_T back to p_0 . This backward denoising process is used to generate new samples $\phi_0 = U'$ distributed according to p_0 . The forward and backward processes are illustrated in Fig. 2 taken from Ref. [17]. One of the challenges for diffusion-based generative models is to guarantee exactness in the sense that the generated samples are asymptotically distributed according to the target distribution.

To be more formal, the stochastic differential equation for the forward process is

$$\frac{d\phi}{dt} = f(\phi, t) + g(t)\eta(t) \quad (8)$$

where $f(\phi, t)$ is a time-dependent drift term, $g(t)$ the time-dependent diffusion coefficient, and $\eta(t)$ Gaussian noise. The denoising backward process is described by the stochastic differential equation

$$\frac{d\phi}{dt} = [f(\phi, t) - g(t)^2 \nabla_\phi \log p_t(\phi)] + g(t) \bar{\eta}(t) \quad (9)$$

where the drift term in square brackets now contains the gradient of $\log p_t(\phi)$ defining the score function $\nabla_\phi \log p_t(\phi)$ to be learned with a machine-learning approach.

One can gain an interesting insight if one sets the drift term $f(\phi, t) = 0$ in Eqs. (8) and (9). In this case, the forward process corresponds to a so-called variance-expanding scheme because the added noise is not regulated by a drift term. On the other hand, the backward process becomes a variant of stochastic quantization, see Ref. [18] for a pedagogical review. This relation can then be

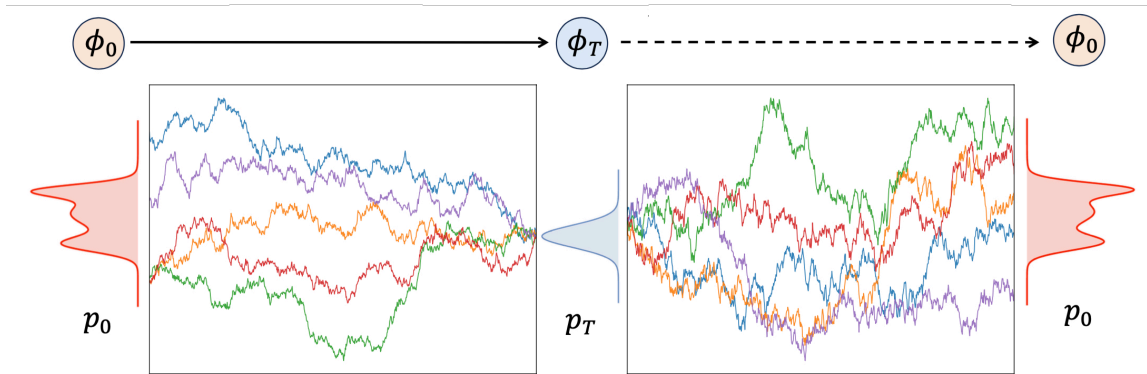


Figure 2: Illustration of the forward diffusion process where stochastic noise is added and the backward denoising process employed in generative diffusion models. Figure taken from Ref. [17].

used as a physical condition for sampling configurations along the backward process, as suggested in Ref. [19]. There, such a physics-conditioned diffusion model is discussed in the context of U(1) gauge theory in two spacetime dimensions combined with Metropolis-adjusted annealed Langevin dynamics in order to guarantee the exactness of the sampling and to enhance efficiency. Finally, we note that the connection to stochastic quantisation has also been exploited for complex-valued actions, using energy-based diffusions models, see Ref. [20], although this is for two-dimensional ϕ^4 theory.

Another example of a diffusion-based machine-learning approach just appeared before the conference [21]. In this work, care was taken to construct symmetry-preserving diffusion models based on score matching. Moreover, the loss function is augmented with a regularization term containing the force which, in general, is analytically known for gauge field theories. Consequently, two avenues are suggested to make the diffusion-based sampling exact, one based on reweighting and the other on resampling.

While these diffusion-based approaches are very promising and show great potential, they come with the caveat that so far they have been employed in two-dimensional gauge theories only, mainly for U(1). Exceptions are Ref. [22] for SU(3) gauge theory on a 4^2 lattice and the recent work in Ref. [23] for a single generic SU(N) gauge matrix. Furthermore, the extensions to two-dimensional U(2) and SU(2) gauge theories in Ref. [24, 25] demonstrate that progress is fast in this area. However, experience has shown that successful machine-learning approaches for two-dimensional gauge theories do not easily transfer to SU(3) and four spacetime dimensions.

2.3 Stochastic normalizing flows

A very interesting approach involving machine learning for lattice gauge theory has been put forward in Ref. [26] and is applied in the context of four-dimensional SU(3) gauge theory in Ref. [27]. The approach is based on the fact that topological freezing can be mitigated by employing open boundary conditions (OBC), e.g., in the direction of Euclidean time. Instead of opening the boundaries on a full spatial timeslice, one can open it on a spatially localized defect only. In order to avoid the problems with the loss of translational invariance in time (and space in case a spatially localized defect is used), one can apply parallel tempering to connect the ensemble with the defect to the one with fully periodic boundary conditions (PBC) [28, 29]. However, instead of using expensive parallel tempering, one can employ out-of-equilibrium evolutions based on Jarzynski's equality [30, 31]. This is illustrated in the left plot of Fig. 3 where the black squares at the bottom represent a number of Monte Carlo steps on the lattice with OBC enabling an efficient update of topological modes. After every n_{between} updates the configuration so obtained on the OBC lattice is subject to n_{steps} of out-of-equilibrium update steps (red squares) transforming it into a configuration on the PBC lattice. So from an ensemble of topologically decorrelated OBC configurations one obtains an ensemble of PBC configurations. The latter needs to be reweighted with a statistical weight obtained from the work spent along the non-equilibrium evolution. This evolution follows a specific protocol with transition probabilities $P_i, i = 1, \dots, N = n_{\text{steps}}$ satisfying detailed balance, but the evolution does not reach equilibrium. Connecting the non-equilibrium Markov Chain Monte Carlo (NE-MCMC) to the theoretical framework of stochastic normalizing flows [33] eventually enables the application of machine-learning techniques [34, 35]. More specifically, the NE-MCMC evolution can be enhanced by combining it with a class of deep generative normalizing-flow models

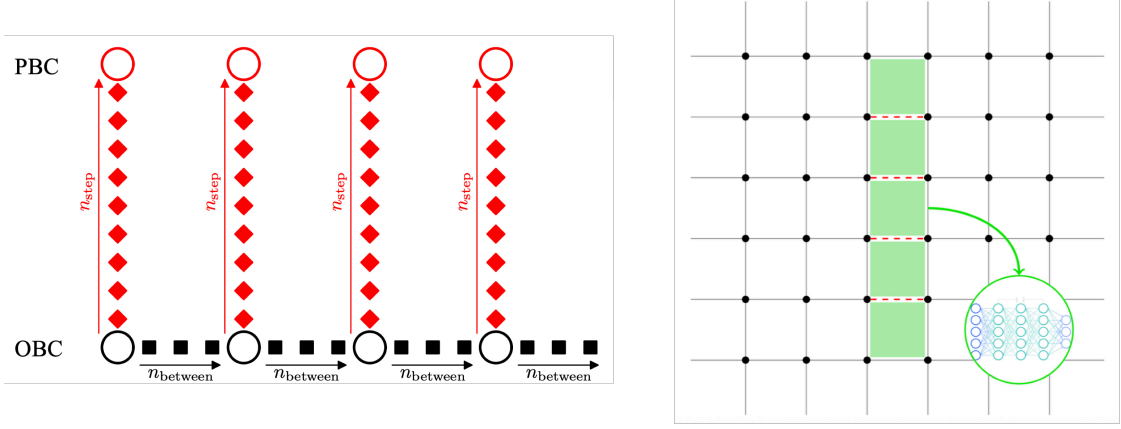


Figure 3: Illustration of the stochastic non-equilibrium MCMC scheme. *Left plot:* Configurations obtained on a lattice with open boundary conditions (OBC) using standard MCMC updates are connected via non-equilibrium MCMC updates to configurations on a lattice with periodic boundary conditions (PBC). *Right plot:* Connecting OBC to PBC on a localized defect using machine-learned stochastic normalizing flows on the links and plaquettes affected by the defect (plot taken from [32]).

parametrized by g_θ ,

$$U_0 \longrightarrow g_\theta^1(U_0) \xrightarrow{P_1} U_1 \longrightarrow g_\theta^2(U_1) \xrightarrow{P_2} \dots \xrightarrow{P_N} U_N = U. \quad (10)$$

This defines a particular instance of a stochastic normalizing flow where the normalizing-flow layers are trained to bring the NE-MCMC as close as possible to equilibrium, thereby reducing the variance of the work done along the given NE-MCMC protocol and hence improving the effectiveness of reweighting the configurations on the PBC lattice. Applying the normalizing flow only on the gauge links localized around the defect [36], as illustrated in the right plot of Fig. 3, simplifies the machine-learning task considerably. By combining the normalizing flow with the NE-MCMC one can achieve a speed-up of about a factor ~ 3 compared to just the standard NE-MCMC [32].

As demonstrated in Ref. [37] this combined strategy scales well for four-dimensional SU(3) gauge theories, both when decreasing the lattice spacing and increasing the number of degrees of freedom. For example, one finds that scaling n_{between} with a^{-2} keeps autocorrelations roughly fixed, while $n_{\text{step}} \propto a^{-3}$ maintains the efficiency of the flow towards the continuum limit. Indeed, scaling up to $\beta = 6.4$ on a 34^4 lattice has been reported at this conference [32] which underlines the potential of this approach.

3. Inverse renormalization group transformations

This machine-learning approach [38, 39] is based on the idea that at coarse lattice spacings uncorrelated lattice configurations can be easily generated without critical slowing down, while at the same time large lattice artefacts are avoided by machine learning a highly improved gauge action based on the renormalization group.

3.1 The fixed-point action

A real-space renormalization group transformation (RGT) can be defined as

$$\exp \{-\beta' A'[V]\} = \int \mathcal{D}U \exp \{-\beta (A[U] + T[U, V])\} \quad (11)$$

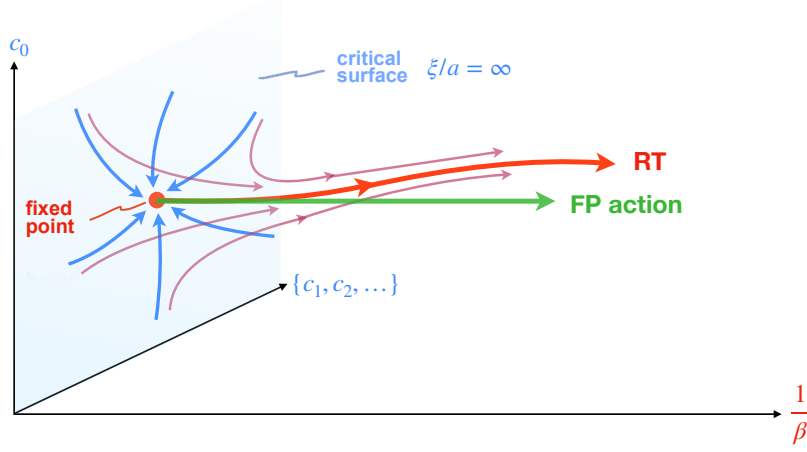


Figure 4: Illustration of the RGT flow. The plane at $\beta \rightarrow \infty$ represents the critical surface where all the irrelevant couplings $\{c_\alpha\}$ flow to the fixed point (FP). An action perturbed from the FP in direction of the only relevant coupling $g^2 \sim 1/\beta$ flows, under repeated (continuous) RGTs, along the renormalized trajectory (RT) defining *quantum perfect actions* with no lattice artefacts at finite lattice spacing. The action along the straight line in direction $1/\beta$ emanating from the FP defines a *classically perfect action* with no tree-level lattice artefacts $O(a^{2n})$ for all n .

where the blocking kernel $T[U, V]$ couples the gauge field V on a coarse lattice with lattice spacing a' to the gauge field U on a fine lattice with lattice spacing a . This can be realized, for example, by the kernel

$$T[U, V] = -\frac{\kappa}{N} \sum_{x', \mu} \left\{ \text{ReTr} \left(V_\mu(x') \cdot Q_\mu^\dagger(x') \right) - \mathcal{N}_\mu^\beta \right\} \quad (12)$$

where x' are the coordinates of the coarse lattice, $Q_\mu(x')$ is a blocked link constructed from the fine links $U(x)$, and \mathcal{N}_μ^β is a normalization constant. The parameter κ as well as additional parameters in the blocking function for the blocked link Q_μ define a specific RGT and are kept fixed. They can however be tuned for particular properties of the resulting effective action $\beta' A'[V]$, e.g., optimizing its locality. Each RGT step decreases the resolution of the lattice, essentially moving from left to right in Fig. 1, while keeping the long-distance physics intact. This is guaranteed by choosing the normalization factor \mathcal{N}_μ^β such that $Z(\beta') = Z(\beta)$. As a consequence, the long-distance physics at any lattice spacing a' is related to the continuum physics through the RGTs. The physics is encoded in the effective action $\beta' A'[V]$ which, however, involves infinitely many couplings $\{c'_\alpha\}$.¹ This is illustrated in Fig. 4 where we show the flow of the couplings under iterated (continuous) RGTs. Fortunately, for asymptotically free gauge theories there is only one (marginally) relevant coupling, namely the gauge coupling g , while all other couplings $\{c_0, c_1, c_2 \dots\}$ are irrelevant. The continuum limit $g^2 \rightarrow 0$, or equivalently $1/\beta \rightarrow 0$ as in Fig. 4, defines the critical surface with $\xi/a = \infty$. On that surface the (irrelevant) couplings flow to the fixed point (FP), where they are reproduced under the RGT, $\{c_\alpha^{\text{FP}}\} \xrightarrow{\text{RGT}} \{c_\alpha^{\text{FP}}\}$. The universal properties in the neighbourhood of the FP guarantee the universality of the continuum limit for different gauge actions. Lattice actions at a very small, but finite lattice spacing are situated just above the critical surface. Under iterated RGTs they flow towards the renormalised trajectory (RT) which emanates from the FP in the direction of the only relevant coupling $g^2 \sim 1/\beta$. The couplings on the RT describe *quantum perfect actions*.

¹From now on we discard the primes to denote RG-transformed quantities.

These are effective actions which have no lattice artefacts at all at any lattice spacing, because they are directly connected to the FP on the critical surface via the RGTs. The quantum perfect actions constitute the holy grail of Symanzik's improvement programme: they are improved to all orders in g^2 and a^2 and a simulation at a single coarse lattice spacing reproduces exactly the continuum long-distance physics.

Finding and constructing such an effective action is of course very difficult and presents two practical challenges: firstly, how to parametrize the RT, i.e., which (finite) set of operators and coefficients $\{c_\alpha\}$ to choose, and secondly, how to determine the coefficients $\{c_\alpha^{\text{RT}}\}$ or $\{c_\alpha^{\text{FP}}\}$. The first question has been answered long ago by Hasenfratz and Niedermayer [40]. They realized that for $\beta \rightarrow \infty$ (on the critical surface) the RGT becomes a classical saddle point problem and reduces to the FP equation

$$A^{\text{FP}}[V] = \min_{\{U\}} \{A^{\text{FP}}[U] + T[U, V]\} . \quad (13)$$

The action A^{FP} defines an action for all values of β , cf. the straight line in Fig. 4. Hasenfratz and Niedermayer also realized that the FP action is *classically perfect*: it has no lattice artefacts on solutions of the classical equations of motion and hence no tree-level lattice artefacts $O(a^{2n})$ to all orders in n . This is so because the classical FP equation preserves all classical properties since they are connected back to the continuum through iterations of the FP equation. The iterated FP equation essentially realises an inception procedure for the classical properties of the gauge field configurations on coarse lattices. In addition to the absence of tree-level lattice artefacts, the lattice artefacts $O(g^2 a^{2n})$, $n = 1, 2, \dots$ induced by quantum effects are expected to be substantially reduced because they are suppressed by the small coupling g^2 . This holds close to the continuum where the FP action follows closely the RT, cf. Fig. 4. From the perfect classical properties it follows that the FP action has scale invariant instanton solutions [41–44]. More importantly, it also enables a classically perfect gradient flow as recently discussed in Ref. [45, 46].

In summary, the FP equation solves the first of the two practical problems mentioned above, namely how to determine the couplings $\{c_\alpha^{\text{FP}}\}$. As a solution to the second problem, namely, how to choose a specific parametrization of the FP action in practice, in a collaboration with Kieran Holland, Andreas Ipp and David Müller we have recently proposed to use a machine-learning approach [38, 39] as described in the next section.

3.2 Machine learning the fixed-point action

In order to parametrize the FP action accurately and in a most efficient way, it is crucial to choose a well-tailored finite set of Wilson loops, together with the corresponding coefficients $\{c_\alpha^{\text{FP}}\}$. In the past, a sum of powers of traces of Wilson loops has been used, starting from a small set of simple Wilson loops, such as the plaquette and rectangular loops [42–44, 47], and increasing the complexity by including smeared Wilson loops [48]. More recently, it has been proposed in Ref. [38, 39] to use a machine-learning approach. It employs a lattice gauge equivariant convolutional neural network (L-CNN) which is capable of generating arbitrarily shaped Wilson loops in a systematic way [49]. The coefficients of these loops can then be machine learned to reproduce arbitrary gauge equivariant or gauge invariant functions of the gauge fields. The main ingredients of the network are illustrated in Fig. 5. The lattice-convolutional layers (L-Conv) take gauge covariant objects W and parallel transport them from positions $x + k\hat{\mu}$ to x using the gauge

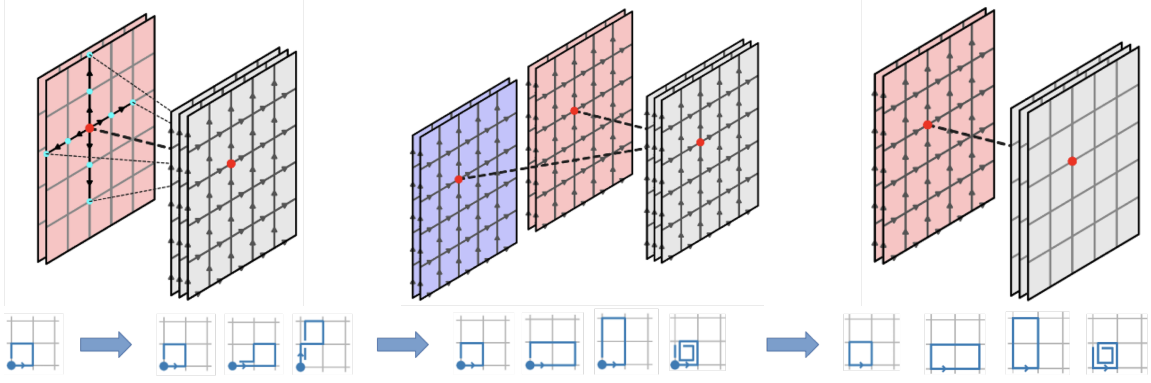


Figure 5: Illustrations of the L-CNN: the convolutional layer (L-Conv) in the left panel parallel transports gauge-covariant objects to a common position (red dot), where they are combined in the bilinear layer (L-Bilin) in the middle panel and eventually mapped to gauge invariant objects with the trace layer (L-Tr) in the right panel. The lower row depicts simple examples of applying the sequence of the layers above.

links U producing new, gauge covariant objects $W'_{x,i} = \sum_{j,\mu,k} w_{i,j,\mu,k} U_{x,k \cdot \hat{\mu}} W_{x+k \cdot \hat{\mu},j} U_{x,k \cdot \hat{\mu}}^\dagger$ where $U_{x,k \cdot \hat{\mu}} = U_{x,\mu} \cdot \dots \cdot U_{x+k-1 \cdot \hat{\mu}}$. The lattice-bilinear layers (L-Bilin) take two gauge covariant objects W and W' and produce bilinear combinations $W''_{x,i} = \sum_{j,j'} \alpha_{i,j,j'} W_{x,j} W'_{x,j'}$. The trace layer (L-Tr) eventually produces gauge invariant objects $w_{x,i} = \text{Tr} W_{x,j} \in \mathbb{C}$ and the network can be further enhanced by supplementing additional activation layers. The ranges of the indices i, j, k are part of the architecture choice, while the coefficients $\{w_{i,j,\mu,k}\}$ and $\{\alpha_{i,j,j'}\}$ are the trainable parameters independent of x making the network translationally invariant.

The data set for the supervised machine learning of the trainable parameters is produced as follows. A large range of gauge field ensembles are generated with the Wilson gauge action (or any other action) covering field fluctuations from very smooth to very coarse. For each configuration V one then solves the FP Eq. (13) by finding the minimizing configuration U_{\min} on the RHS using a sufficiently good approximation of $A^{\text{FP}}[U]$, e.g., as provided in [48]. This yields the exact FP action value $A^{\text{FP}}[V]$ and in addition also the exact derivatives w.r.t. to each gauge link,

$$\frac{\delta A^{\text{FP}}[V]}{\delta V_{x,\mu}^a} = \frac{\delta T[U_{\min}, V]}{\delta V_{x,\mu}^a} = -\kappa \text{Re} \text{Tr} (it^a V_{x,\mu} Q_{x,\mu}^\dagger [U_{\min}]), \quad (14)$$

where x runs over all sites of the coarse lattice, $\mu = 1, \dots, 4$ over all spacetime directions, and $a = 1, \dots, 8$ over the colour indices. In total, this yields $4 \times 8 \times (\text{volume}) + 1$ learning data per configuration. The direct parametrization of the derivatives is of course most useful for simulating the action with the HMC algorithm, or for constructing the gradient flow.

An important step in any machine-learning approach is to find an appropriate network architecture. The outcome of such a search for an optimal architecture is shown in Fig. 6 where the distributions of the relative action errors and the derivative errors of the L-CNN output w.r.t. to the exact FP values in Eqs. (13) and (14) are represented in terms of box plots. From the dependence on the number of combined convolutional and bilinear layer pairs (left row), the total number of output channels (middle row) and kernel sizes (right row) one can identify good architectures. The one chosen in Ref. [39] consists of three combined layers with 12, 24, 24 output channels and kernel sizes $k = 2, 2, 1$, respectively, yielding a total of about 443k parameters. Since the number of layers

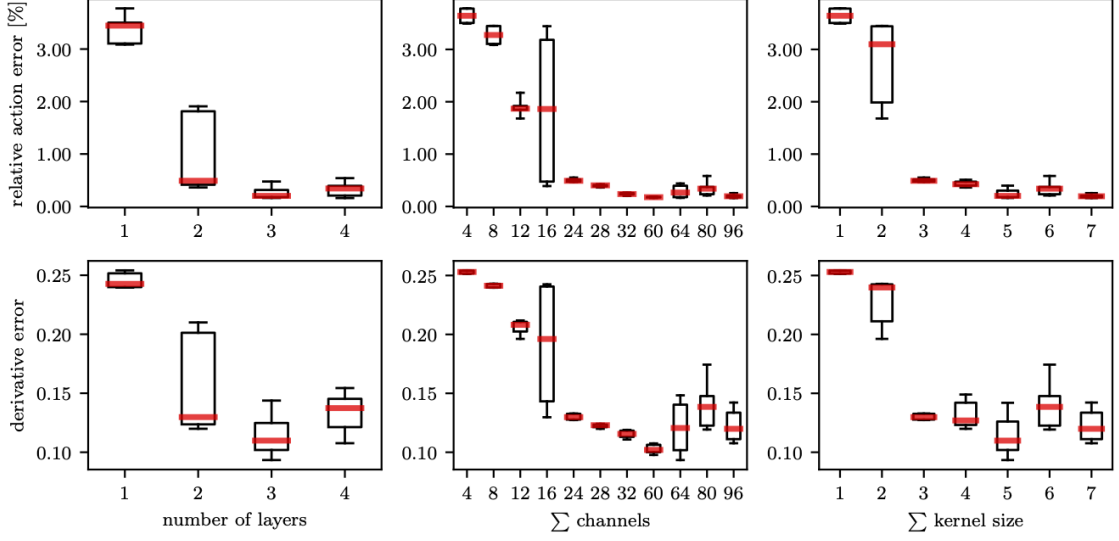


Figure 6: Illustrations of the architecture search. Shown are the relative action error (*top row*) and derivative error (*bottom row*) of the L-CNN w.r.t. to the exact FP values as a function of the architecture parameters, i.e., the number of layers (*left plots*), the total number of channels (*middle plots*), and the kernel size (*right plots*).

is finite, the L-CNN produces an ultralocal action and therefore a truncated approximation of the FP action. However, seeing that the effective couplings between two gauge links in the parametrized action decay exponentially fast, at least as $\exp(-3.13r)$ with their separation r , cf. [39], the truncation error is expected to be small. In any case, what matters for the successful description of the FP action is of course the total parametrization error. Eventually, the quality of the parametrization can only be checked in actual simulations.

3.3 The fixed-point action in action

While the FP action has a very complicated structure in terms of a plethora of extended Wilson loops, the derivatives with respect to the gauge links are directly obtained as an output from the L-CNN. As a consequence, HMC simulations and the calculations of gradient-flow observables are straightforward. In fact, the gradient flow provides an ideal test case for the FP approach and the quality of the machine-learned parametrization of the FP action, because gradient-flow observables can be measured very precisely and entail essentially no systematic errors. They are therefore ideal candidates to quantify lattice artefacts and test the scaling towards the continuum limit. In particular, one can consider physical reference scales t_x and w_x [50, 51] defined by

$$t^2 \langle E \rangle |_{t=t_x} = x, \quad t \frac{d}{dt} \left(t^2 \langle E \rangle \right) |_{t=w_x^2} = x \quad (15)$$

yielding dimensionless ratios t_x/w_x^2 or t_x/t_y as scaling quantities with well defined continuum limits. Because the FP action is classically perfect, no lattice artefacts are introduced through the gradient flow or the measurement of the action density $E(t)$, and the only deviations from perfect scaling are either due to quantum lattice artefacts of order $O(g^2 a^{2n})$, or due to the imperfect parametrization of the FP action by the machine-learned L-CNN. The results for the dimensionless ratios $t_{0.5}/t_{0.3}$ (left plot) and $t_{0.3}/w_{0.3}$ (right plot) as a function of the lattice spacing $a^2/t_{0.3}$ are

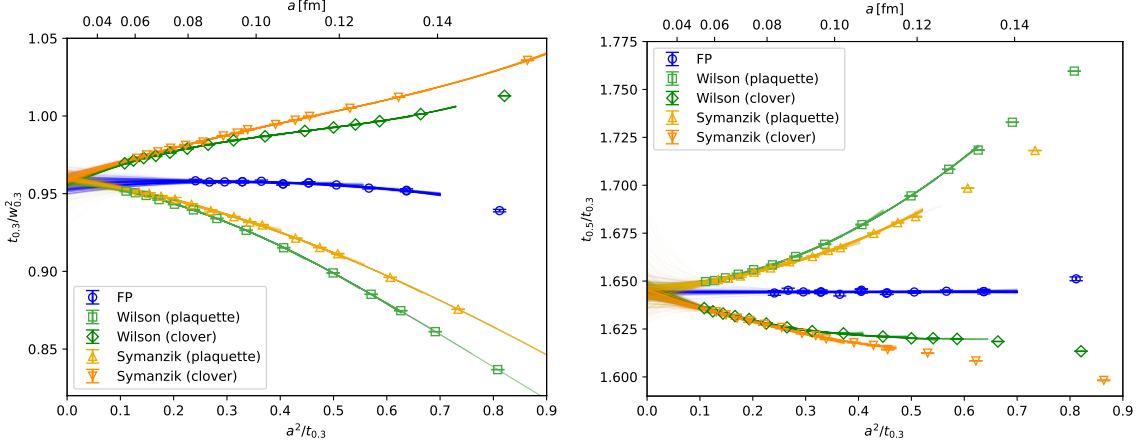


Figure 7: Continuum-limit extrapolations for the ratios $t_{0.3}/w_{0.3}^2$ and $t_{0.5}/t_{0.3}$. Results from Wilson and Symanzik MC simulations are shown using plaquette and clover discretizations of the action density.

shown in Fig. 7 using MC simulations of the FP, Wilson and the tree-level Symanzik-improved gauge actions. For the latter two, the action density is evaluated with either the plaquette or clover operator. The leading lattice artefacts are expected to be $O(a^2)$ for the Wilson data, $O(a^4)$ for the Symanzik data, and $O(g^2 a^2)$ for the FP data. The lines represent bootstrap samples of various possible fits describing the continuum extrapolations. We note that the leading $O(a^4)$ behaviour of the Symanzik data is masked by the $O(a^2)$ discretization effects from the flow action (Wilson) and the action density (plaquette and clover). This can be remedied by employing the Zeuthen flow introduced in Ref. [52] yielding $O(a^2)$ improved gradient-flow observables. In contrast, the improvement with the machine-learned FP action is to all orders in the lattice spacing at tree level. Indeed, the observed lattice artefacts for the FP action are less than 1% up to lattice spacings of 0.14 fm, allowing continuum physics to be extracted for SU(3) gauge theory in four spacetime dimensions from coarse lattices. This is further exemplified in Fig. 8 where the plot on the left

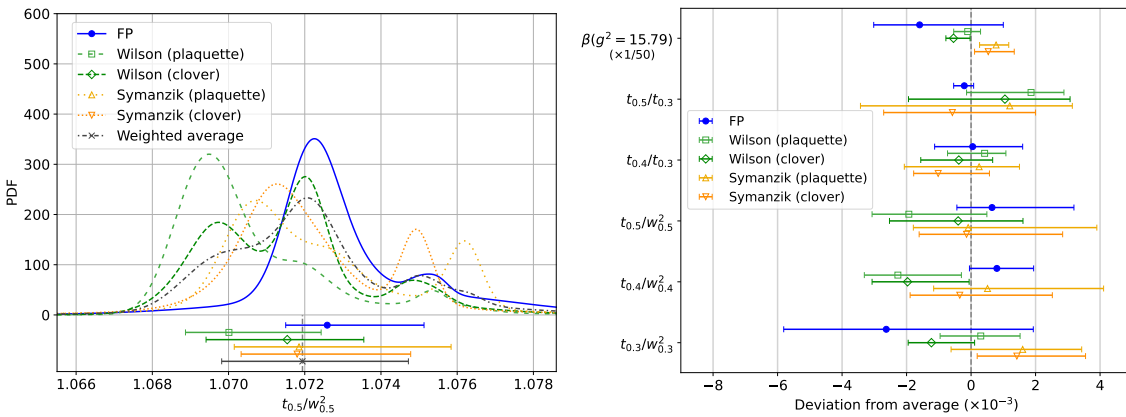


Figure 8: *Left plot:* Comparison of various PDFs obtained from AIC-weighted continuum limits of the ratio $t_{0.5}/w_{0.5}^2$ using a variety of fit functions and ranges. *Right plot:* Comparison of continuum predictions for four-dimensional SU(3) gauge theory from MC simulations using either the FP, Wilson or tree-level Symanzik improved lattice action. The β -function results are rescaled by a factor of 50 for visibility.

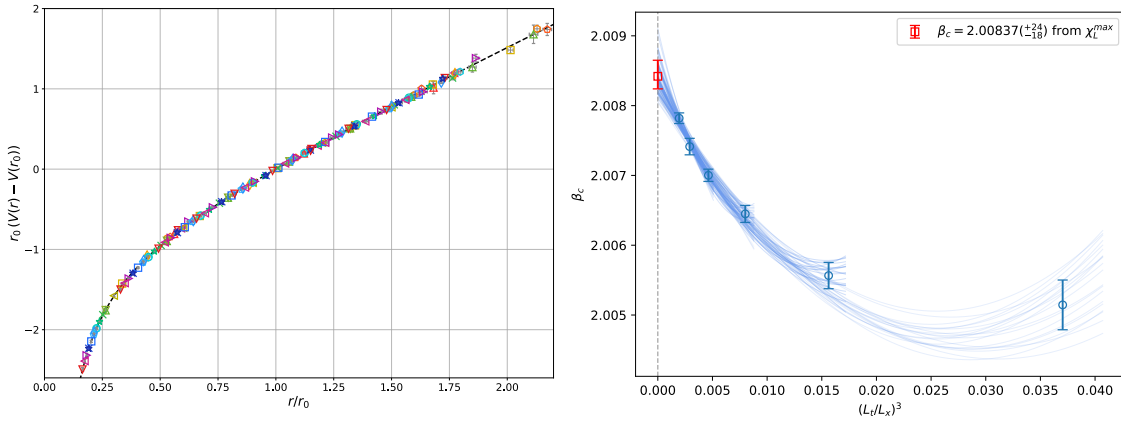


Figure 9: *Left plot:* Static quark-antiquark potential obtained on lattices with lattice spacings ranging between $0.08 \text{ fm} \leq a \leq 0.3 \text{ fm}$. *Right plot:* Thermodynamic limit of the critical coupling β_c obtained from the position of the Polyakov loop susceptibility at a lattice spacing of $a \simeq 0.3 \text{ fm}$ corresponding to a temporal lattice extent of $L_t = 2$.

shows the AIC-weighted PDFs of the continuum limits for the ratio $t_{0.5}/w_{0.5}^2$ using a variety of fit functions and ranges for the FP, Wilson and Symanzik action. The plot shows how the continuum values of the gradient-flow observables are dominated by the systematic effects from the continuum extrapolations. It underlines the importance of controlling lattice artefacts well. The right plot shows a comparison of continuum predictions for a variety of gradient-flow observables, including the β -function at the gradient-flow coupling $g^2 = 15.78$. There is clear consistency between FP, Wilson and Symanzik results demonstrating universality and the successful implementation of the machine-learned L-CNN for the FP approach.

The classically perfect improvement is not limited to gradient-flow observables, but in fact also applies to spectral observables. This is illustrated in the left plot of Fig. 9 where results for the static quark-antiquark potential are shown including simulations at lattice spacings as coarse as $a \simeq 0.3 \text{ fm}$. There are essentially no lattice artefacts visible even at this coarse resolution. In contrast, the right plot is not about lattice artefacts, but rather about finite-volume effects. It shows the extrapolation of the critical coupling β_c to the thermodynamic limit. The critical coupling is obtained from the position of the peak of the Polyakov loop susceptibility at a lattice spacing of $a \simeq 0.3 \text{ fm}$ corresponding to a temporal lattice extent of $L_t = 2$. Simulating at such a coarse lattice spacing has the advantage that it is particularly easy to reach large aspect ratios. As a consequence, also in this situation the FP approach is most suited and the machine-learned FP action indeed appears to perform well.

4. Summary and conclusions

Since the application of machine-learning techniques in lattice gauge theories is a rather new field, many developments are still in an exploratory state. Hence, there are many opportunities for making progress. It is interesting to see the many creative combinations of often complementary approaches and tools that are employed. There is a large variety of different ideas, many of which will probably not be successful in the end, but nevertheless need to be tried out.

The efforts of applying machine learning for lattice gauge theories are often motivated by the urge to overcome problems related to critical slowing down towards the continuum, in particular topological freezing. One lesson to be learned from this brief review is that it is in general not sufficient to just adapt a given machine-learned generative model to a gauge theory, but instead it requires the employment of additional physics-based concepts in order to enable efficient sampling of gauge field configurations. Examples for such physics-driven machine-learning approaches are the physics-conditioned diffusion models [19], the stochastic-normalizing-flow models enhanced by non-equilibrium dynamics [35, 37], and the machine-learned FP actions based on renormalization group transformations [39, 46], as discussed in this review. Another interesting proposal in that direction has been put forward very recently in Ref. [53] where physics-informed renormalization group flows are used to transform the generative process into that of solving differential equations for the kernels of the RG transformation. The necessity of including physics information into the machine learning may be a hint why approaches solely based on normalizing flows without any physics input have not yet been able to scale to realistic lattice volumes in four spacetime dimensions.

In fact, it is probably safe to state that the step from low-dimensional applications with simple degrees of freedom to four-dimensional large-volume applications is most challenging. So far, not many applications have successfully mastered this step, with the notable exceptions being [35, 37] and [39, 46]. However, given the variety of ideas available, there is no doubt that there will soon be many more approaches establishing machine learning as a transformative tool for overcoming computational challenges in lattice gauge theories.

Acknowledgments

I would like to thank K. Holland, A. Ipp, and D.I. Müller for a most enjoyable collaboration. I would also like to thank G. Aarts, L. Backfried, C. Bonanno, E. Cellini, G. Kanwar, J. Mayer-Stedte, A. Nada, F. Romero-López, S. Romiti, and L. Verzhicelli for useful discussions, and S. Thommen for providing the data in the left plot in Fig. 9.

This work has been supported by the Swiss National Science Foundation (SNSF) through grant No. 200020_208222 and No. 200021-232288, the Platform for Advanced Scientific Computing (PASC) under the project Alpenglue, and the Albert Einstein Center for Fundamental Physics at the University of Bern. The computational results presented in the context of the FP action have been achieved in part using the Vienna Scientific Cluster (VSC) and LEONARDO at CINECA, Italy, via an AURELEO (Austrian Users at LEONARDO supercomputer) project, and UBELIX, the HPC cluster at the University of Bern.

References

- [1] S. Lawrence, *Machine-learning approaches to accelerating lattice simulations*, *PoS LATTICE2024* (2025) 010, [2502.02670].
- [2] R. Abbott, D. Boyda, Y. Fu, D. C. Hackett, G. Kanwar, F. Romero-López et al., *Variance reduction in lattice QCD observables via normalizing flows*, 2603.02984.

- [3] W. Detmold, G. Kanwar, Y. Lin, P. E. Shanahan and M. L. Wagman, *Exploring gauge-fixing conditions with gradient-based optimization*, in *41st International Symposium on Lattice Field Theory*, 10, 2024. [2410.03602](#).
- [4] V. Bellscheidt, N. Brambilla, A. S. Kronfeld and J. Mayer-Studte, *Wilson loops with neural networks*, [2602.02436](#).
- [5] T. Spriggs, E. Greplova, J. Carrasquilla and J. Nys, *Accurate Ground States of $SU(2)$ Lattice Gauge Theory in 2+1D and 3+1D*, *Phys. Rev. Lett.* **136** (2026) 101902, [[2509.12323](#)].
- [6] S. Romiti, *$SU(N)$ lattice gauge theories with physics-informed neural networks*, *Phys. Rev. D* **113** (2026) 054511, [[2510.26904](#)].
- [7] G. Kanwar, *Flow-based sampling for lattice field theories*, in *40th International Symposium on Lattice Field Theory*, 1, 2024. [2401.01297](#).
- [8] G. Kanwar, M. S. Albergo, D. Boyda, K. Cranmer, D. C. Hackett, S. Racanière et al., *Equivariant flow-based sampling for lattice gauge theory*, *Phys. Rev. Lett.* **125** (2020) 121601, [[2003.06413](#)].
- [9] D. Boyda, G. Kanwar, S. Racanière, D. J. Rezende, M. S. Albergo, K. Cranmer et al., *Sampling using $SU(N)$ gauge equivariant flows*, *Phys. Rev. D* **103** (2021) 074504, [[2008.05456](#)].
- [10] R. Abbott et al., *Gauge-equivariant flow models for sampling in lattice field theories with pseudofermions*, *Phys. Rev. D* **106** (2022) 074506, [[2207.08945](#)].
- [11] R. Abbott et al., *Normalizing flows for lattice gauge theory in arbitrary space-time dimension*, [2305.02402](#).
- [12] R. Abbott, D. Boyda, D. C. Hackett, G. Kanwar, F. Romero-López, P. E. Shanahan et al., *Practical applications of machine-learned flows on gauge fields*, *PoS LATTICE2023* (2024) 011, [[2404.11674](#)].
- [13] R. Abbott, D. Boyda, G. Kanwar, F. Romero-López, D. C. Hackett, P. E. Shanahan et al., *Progress in Normalizing Flows for 4d Gauge Theories*, *PoS LATTICE2024* (2025) 066, [[2502.00263](#)].
- [14] R. Abbott, A. Botev, D. Boyda, D. C. Hackett, G. Kanwar, S. Racanière et al., *Applications of flow models to the generation of correlated lattice QCD ensembles*, *Phys. Rev. D* **109** (2024) 094514, [[2401.10874](#)].
- [15] R. Abbott et al., *Aspects of scaling and scalability for flow-based sampling of lattice QCD*, *Eur. Phys. J. A* **59** (2023) 257, [[2211.07541](#)].
- [16] C.-H. Lai, Y. Song, D. Kim, Y. Mitsufuji and S. Ermon, *The principles of diffusion models*, 2025.

- [17] L. Wang, G. Aarts and K. Zhou, *Diffusion models as stochastic quantization in lattice field theory*, *JHEP* **05** (2024) 060, [2309.17082].
- [18] K. Fukushima, S. Kamata and Y. Hirono, *Stochastic Quantization and Diffusion Models*, *J. Phys. Soc. Jap.* **94** (2025) 031010, [2411.11297].
- [19] Q. Zhu, G. Aarts, W. Wang, K. Zhou and L. Wang, *Physics-conditioned diffusion models for lattice gauge theory*, *JHEP* **03** (2026) 111, [2502.05504].
- [20] 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].
- [21] O. Vega, J. Komijani, A. El-Khadra and M. Marinkovic, *Group-Equivariant Diffusion Models for Lattice Field Theory*, 2510.26081.
- [22] A. Lou, M. Xu and S. Ermon, *Scaling riemannian diffusion models*, 2023.
- [23] G. Kanwar and O. Vega, *Spectral Diffusion for Sampling on $SU(N)$* , in *42th International Symposium on Lattice Field Theory*, 12, 2025. 2512.19877.
- [24] 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.
- [25] H. Alharazin, J. Y. Panteleeva and B. D. Sun, *Diffusion Models for $SU(2)$ Lattice Gauge Theory in Two Dimensions*, 2602.09045.
- [26] C. Bonanno, A. Nada and D. VDACCHINO, *Mitigating topological freezing using out-of-equilibrium simulations*, *JHEP* **04** (2024) 126, [2402.06561].
- [27] D. VDACCHINO, A. Nada and C. Bonanno, *Topological susceptibility of $SU(3)$ pure-gauge theory from out-of-equilibrium simulations*, *PoS LATTICE2024* (2025) 415, [2411.00620].
- [28] M. Hasenbusch, *Fighting topological freezing in the two-dimensional CPN-1 model*, *Phys. Rev. D* **96** (2017) 054504, [1706.04443].
- [29] C. Bonanno, C. Bonati and M. D’Elia, *Large- N $SU(N)$ Yang-Mills theories with milder topological freezing*, *JHEP* **03** (2021) 111, [2012.14000].
- [30] C. Jarzynski, *Nonequilibrium equality for free energy differences*, *Phys. Rev. Lett.* **78** (Apr, 1997) 2690–2693, [cond-mat/9610209].
- [31] M. Caselle, G. Costagliola, A. Nada, M. Panero and A. Toniato, *Jarzynski’s theorem for lattice gauge theory*, *Phys. Rev. D* **94** (2016) 034503, [1604.05544].
- [32] C. Bonanno, A. Bulgarelli, E. Cellini, A. Nada, D. Panfalone, D. VDACCHINO et al., *A scalable flow-based approach to mitigate topological freezing*, in *42th International Symposium on Lattice Field Theory*, 1, 2026. 2601.20708.
- [33] M. Caselle, E. Cellini, A. Nada and M. Panero, *Stochastic normalizing flows as non-equilibrium transformations*, *JHEP* **07** (2022) 015, [2201.08862].

- [34] A. Bulgarelli, E. Cellini and A. Nada, *Sampling $SU(3)$ pure gauge theory with Stochastic Normalizing Flows*, *PoS LATTICE2024* (2025) 040, [2409.18861].
- [35] 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].
- [36] A. Bulgarelli, E. Cellini, K. Jansen, S. Kühn, A. Nada, S. Nakajima et al., *Flow-Based Sampling for Entanglement Entropy and the Machine Learning of Defects*, *Phys. Rev. Lett.* **134** (2025) 151601, [2410.14466].
- [37] C. Bonanno, A. Bulgarelli, E. Cellini, A. Nada, D. Panfalone, D. Vadicchino et al., *Scaling flow-based approaches for topology sampling in $SU(3)$ gauge theory*, 2510.25704.
- [38] K. Holland, A. Ipp, D. I. Müller and U. Wenger, *Fixed point actions from convolutional neural networks*, *PoS LATTICE2023* (2024) 038, [2311.17816].
- [39] K. Holland, A. Ipp, D. I. Müller and U. Wenger, *Machine learning a fixed point action for $SU(3)$ gauge theory with a gauge equivariant convolutional neural network*, *Phys. Rev. D* **110** (2024) 074502, [2401.06481].
- [40] P. Hasenfratz and F. Niedermayer, *Perfect lattice action for asymptotically free theories*, *Nucl. Phys. B* **414** (1994) 785–814, [hep-lat/9308004].
- [41] M. Blatter, R. Burkhalter, P. Hasenfratz and F. Niedermayer, *Instantons and the fixed point topological charge in the two-dimensional $O(3)$ sigma model*, *Phys. Rev. D* **53** (1996) 923–932, [hep-lat/9508028].
- [42] T. A. DeGrand, A. Hasenfratz, P. Hasenfratz and F. Niedermayer, *The Classically perfect fixed point action for $SU(3)$ gauge theory*, *Nucl. Phys. B* **454** (1995) 587–614, [hep-lat/9506030].
- [43] T. A. DeGrand, A. Hasenfratz, P. Hasenfratz and F. Niedermayer, *Nonperturbative tests of the fixed point action for $SU(3)$ gauge theory*, *Nucl. Phys. B* **454** (1995) 615–637, [hep-lat/9506031].
- [44] T. A. DeGrand, A. Hasenfratz, P. Hasenfratz and F. Niedermayer, *Fixed point actions for $SU(3)$ gauge theory*, *Phys. Lett. B* **365** (1996) 233–238, [hep-lat/9508024].
- [45] U. Wenger, K. Holland, A. Ipp and D. I. Müller, *HMC and gradient flow with machine-learned classically perfect fixed point actions*, *PoS LATTICE2024* (2025) 466, [2502.03315].
- [46] K. Holland, A. Ipp, D. I. Müller and U. Wenger, *Machine-Learned Renormalization-Group-Improved Gauge Actions and Classically Perfect Gradient Flows*, *Phys. Rev. Lett.* **136** (2026) 031901, [2504.15870].
- [47] M. Blatter and F. Niedermayer, *New fixed point action for $SU(3)$ lattice gauge theory*, *Nucl. Phys. B* **482** (1996) 286–304, [hep-lat/9605017].

- [48] F. Niedermayer, P. Rufenacht and U. Wenger, *Fixed point gauge actions with fat links: Scaling and glueballs*, *Nucl. Phys. B* **597** (2001) 413–450, [[hep-lat/0007007](#)].
- [49] M. Favoni, A. Ipp, D. I. Müller and D. Schuh, *Lattice Gauge Equivariant Convolutional Neural Networks*, *Phys. Rev. Lett.* **128** (2022) 032003, [[2012.12901](#)].
- [50] M. Lüscher, *Properties and uses of the Wilson flow in lattice QCD*, *JHEP* **08** (2010) 071, [[1006.4518](#)].
- [51] BMW collaboration, S. Borsányi, S. Dürr, Z. Fodor, C. Hoelbling, S. D. Katz, S. Krieg et al., *High-precision scale setting in lattice QCD*, *JHEP* **09** (2012) 010, [[1203.4469](#)].
- [52] A. Ramos and S. Sint, *Symanzik improvement of the gradient flow in lattice gauge theories*, *Eur. Phys. J. C* **76** (2016) 15, [[1508.05552](#)].
- [53] F. Ihssen, R. Kapust and J. M. Pawłowski, *Generative sampling with physics-informed kernels*, [2510.26678](#).