

## Parallel tempered metadynamics: Overcoming potential barriers without surfing or tunneling

Timo Eichhorn<sup>Ⓧ,\*</sup>, Gianluca Fuwa<sup>Ⓧ,†</sup>, Christian Hoelbling<sup>Ⓧ,‡</sup>, and Lukas Varnhorst<sup>Ⓧ,§</sup>  
*Department of Physics, University of Wuppertal, Gaußstraße 20, 42119 Wuppertal, Germany*

 (Received 7 November 2023; accepted 23 April 2024; published 7 June 2024)

At fine lattice spacings, Markov chain Monte Carlo simulations of QCD and other gauge theories with or without fermions are plagued by slow modes that give rise to large autocorrelation times. This can lead to simulation runs that are effectively stuck in one topological sector, a problem known as topological freezing. Here, we demonstrate that for a relevant set of parameters, metadynamics can be used to unfreeze four-dimensional SU(3) gauge theory. However, compared to local update algorithms and the Hybrid Monte Carlo algorithm, the computational overhead is significant in pure gauge theory, and the required reweighting procedure may considerably reduce the effective sample size. To deal with the latter problem, we propose modifications to the metadynamics bias potential and the combination of metadynamics with parallel tempering. We test the new algorithm in four-dimensional SU(3) gauge theory and find that it can achieve topological unfreezing without compromising the effective sample size, thereby reducing the autocorrelation times of topological observables by at least 2 orders of magnitude compared to conventional update algorithms. Additionally, we observe significantly improved scaling of autocorrelation times with the lattice spacing in two-dimensional U(1) gauge theory.

DOI: [10.1103/PhysRevD.109.114504](https://doi.org/10.1103/PhysRevD.109.114504)

### I. INTRODUCTION

In recent years, physical predictions based on lattice simulations have reached subpercent accuracies [1]. With ever-shrinking uncertainties, the demand for precise extrapolations to the continuum grows, which in turn necessitates finer lattice spacings. Current state-of-the-art methods for simulations of lattice gauge theories either rely on a mixture of heat bath [2–6] and overrelaxation [7–9] algorithms for pure gauge theories or molecular-dynamics-based algorithms like the Hybrid Monte Carlo algorithm (HMC) [10] for simulations including dynamical fermions. For all of these algorithms, the computational effort to carry out simulations dramatically increases at fine lattice spacings due to critical slowing down. While the exact behavior depends on a number of factors, such as the update algorithms, the exact discretization of the action, and the choice of boundary conditions, the scaling of the integrated

autocorrelation times with the inverse lattice spacing can usually be described by a power law.

In addition to the general diffusive slowing down, topologically nontrivial gauge theories may exhibit topological freezing [11–18]. This effect appears due to the inability of an algorithm to overcome the action barriers between topological sectors, which can lead to extremely long autocorrelation times of topological observables and thus an effective breakdown of ergodicity.

Over the years, several strategies have been developed to deal with this situation. On the most basic level, it has become customary in large scale simulations to monitor the topological charge of the configurations in each ensemble, thus avoiding regions of parameter space which are affected by topological freezing [19–21]. Another possibility to circumvent the problem consists in treating fixed topology as a finite volume effect and either correcting observables for it [22,23] or increasing the physical volume sufficiently to derive the relevant observables from local fluctuations [24]. It is also possible to use open boundary conditions in one lattice direction [25], which invalidates the concept of an integer topological charge for the price of introducing additional boundary artifacts and a loss of translational symmetry.

Despite the success of these strategies in many relevant situations, the need for a genuine topology changing update algorithm is still great. This is evident from the large number and rather broad spectrum of approaches that are currently being investigated in this direction. Some of these

\*timo.eichhorn@protonmail.com  
 †gianluca.fuwa@uni-wuppertal.de  
 ‡hch@uni-wuppertal.de  
 §varnhorst@uni-wuppertal.de

*Published by the American Physical Society under the terms of the Creative Commons Attribution 4.0 International license. Further distribution of this work must maintain attribution to the author(s) and the published article's title, journal citation, and DOI. Funded by SCOAP<sup>3</sup>.*

approaches address critical slowing down, in general, whereas others focus particularly on topological freezing. These approaches include parallel tempering [26–28], modified boundary conditions [29], and combinations of both [30,31]; multiscale thermalization [32–34], instanton(-like) updates [35–40], metadynamics [41,42], multicanonical simulations [43,44], density of states [45], Fourier acceleration [46–50], and trivializing maps [51–53]. Additionally, recent years have seen multitudinous efforts to utilize generative models to sample configurations more efficiently [54–75]. For a recent review of both conventional and machine learning approaches see [76].

In this work we propose a new update algorithm, parallel tempered metadynamics (PT-MetaD), and demonstrate its efficiency in four-dimensional SU(3) at parameter values where conventional update algorithms suffer from topological freezing. In its basic variant, which we present here, PT-MetaD consists of two update streams simulating the same physical system. One of the streams uses any suitable conventional algorithm, while the other one includes a bias potential that facilitates tunneling between topological sectors. At regular intervals, swaps between the two streams are suggested, so that the good topological sampling from the bias potential stream carries over to the other stream. The algorithm thus combines ideas from parallel tempering [77], metadynamics [78], and multicanonical simulations [43], leading to an efficient sampling of topological sectors while avoiding the problem of small effective sample sizes, which is usually associated with techniques involving reweighting, such as metadynamics or multicanonical simulations. Additionally, applying PT-MetaD to theories including fermions is conceptually straightforward.

This paper is organized as follows. We start out by giving a general introduction to metadynamics in Sec. II. Afterwards, Sec. III describes our simulation setup, including our choice of actions, observables, and update algorithms. Some details on the application of metadynamics in the context of SU(3) gauge theory are also given. In Sec. IV, we present baseline results obtained with conventional update algorithms, including a rough determination of gradient flow scales for the doubly blocked Wilson (DBW2) action. In Sec. V we present results obtained with pure metadynamics for four-dimensional SU(3) and discuss several possible improvements. In Sec. VI we introduce parallel tempered metadynamics and show some scaling tests of the new algorithm in two-dimensional U(1) gauge theory, as well as exploratory results in four-dimensional SU(3). Finally, in Sec. VII, we conclude with a summary and outlook on the application of the new algorithm to full QCD. Appendices A–D contain mathematical conventions and derivations related to the metadynamics force calculation, and Appendix E summarizes exact results for observables in two-dimensional U(1) gauge theory.

## II. METADYNAMICS

Consider a system described by a set of degrees of freedom  $\{U\}$ , where the states are distributed according to the probability density

$$p(U) = \frac{1}{Z} e^{-S(U)}, \quad (1)$$

with the partition function  $Z$  defined as

$$Z = \int \mathcal{D}[U] e^{-S(U)}. \quad (2)$$

The expectation value of an observable  $O$  is defined as

$$\langle O \rangle = \int \mathcal{D}[U] p(U) O(U). \quad (3)$$

In the context of lattice gauge theories, the integration measure  $\mathcal{D}[U]$  is usually the product of Haar measures for each link variable, but more generally  $\mathcal{D}[U]$  may be understood as a measure on the configuration space of the system.

Metadynamics [78] is an enhanced-sampling method, based on the addition of a history-dependent bias potential  $V_t(s(U))$  to the action  $S(U)$ , where  $t$  is the current simulation time. The dynamics of the modified system are governed by  $S_t^M(U) = S(U) + V_t(s(U))$ , and now explicitly depend on a number of observables  $s_i(U)$ , with  $i \in \{1, \dots, N\}$ , that are referred to as collective variables (CVs). These CVs span a low-dimensional projection of the configuration space of the system and may generally be arbitrary functions of the underlying degrees of freedom  $\{U\}$ . However, when used in combination with molecular-dynamics-based algorithms, such as the Hybrid Monte Carlo algorithm, the CVs need to be differentiable functions of the underlying degrees of freedom. During the course of a simulation, the bias potential is modified in such a way as to drive the system away from regions of configuration space that have been explored previously, eventually converging towards an estimate of the negative free energy as a function of the CVs, up to a constant offset [79,80]. Usually, this is accomplished by constructing the potential from a sum of Gaussians  $g(s)$ , so that at simulation time  $t$ , the potential is given by

$$V_t(s) = \sum_{t' \leq t} \prod_{i=1}^N g(s_i - s_i(t')). \quad (4)$$

The exact form of the Gaussians is determined by the parameters  $w$  and  $\delta s_i$ :

$$g(s_i) = w \exp\left(-\frac{s_i^2}{2\delta s_i^2}\right). \quad (5)$$

Both parameters affect the convergence behavior of the potential in a similar way: Increasing the height  $w$  or the

widths  $\delta s_i$  may accelerate the convergence of the potential during early stages of the simulation but lead to larger fluctuations around the equilibrium during later stages. Furthermore, the widths  $\delta s_i$  effectively introduce a smallest scale that can still be resolved in the space spanned by the CVs, which needs to be sufficiently small to capture the relevant details of the potential.

If the bias potential has reached a stationary state, i.e., its time dependence in the region of interest is just an overall additive constant, then the modified probability density, which we shall also refer to as target density, is given by

$$p'(U) = \frac{1}{Z'} e^{-S(U) - V(s(U))}, \quad (6)$$

with the modified partition function

$$Z' = \int \mathcal{D}[U] e^{-S(U) - V(s(U))}. \quad (7)$$

Expectation values with respect to the modified distribution can then be defined in the usual way, i.e., via

$$\langle O \rangle' = \int \mathcal{D}[U] p'(U) O(U). \quad (8)$$

On the other hand, expectation values with respect to the original, unmodified probability density can be written in terms of the modified probability distribution with an additional weighting factor. For a dynamic potential, different reweighting schemes have been put forward to achieve this goal [81], but if the potential is static, then the weighting factors are directly proportional to the exponential of the bias potential:

$$\langle O \rangle = \frac{\int \mathcal{D}[U] p'(U) O(U) e^{V(s(U))}}{\int \mathcal{D}[U] e^{V(s(U))}}. \quad (9)$$

The case of a static potential is thus essentially the same as a multicanonical simulation [43].

In situations where the evolution of the system is hindered by high action barriers separating relevant regions of configuration space, metadynamics can be helpful in overcoming those barriers, since the introduction of a bias potential modifies the marginal distribution over the set of CVs. For conventional metadynamics, the bias potential is constructed in such a way that the marginal modified distribution is constant:

$$p'(s_i) = \int \mathcal{D}[U] p'(U) \delta(s_i - s_i(U)) = \text{const.} \quad (10)$$

Conversely, for a given original distribution  $p(s)$  and a desired target distribution  $p'(s)$ , the required potential is given by

$$V(s) = \ln \left( \frac{p'(s)}{p(s)} \right). \quad (11)$$

Nevertheless, it is important to recognize that even if the bias potential completely flattens out the marginal distribution over the CVs, the simulation is still expected to suffer from other (diffusive) sources of critical slowing down as is common for Markov chain Monte Carlo simulations.

### III. SIMULATION SETUP AND OBSERVABLES

#### A. Choice of gauge actions

For our simulations of SU(3) gauge theory, we work on a four-dimensional lattice  $\Lambda$  with periodic boundary conditions. Configurations are generated using the Wilson [82] and the DBW2 [83] gauge actions, both of which belong to a one-parameter family of gauge actions involving standard  $1 \times 1$  plaquettes as well as  $1 \times 2$  planar loops, which may be expressed as

$$S_g = \frac{\beta}{3} \sum_{n \in \Lambda} \left( \sum_{\mu < \nu} c_0 (3 - \text{Re tr}[\mathcal{W}_{\mu,\nu}(n)]) + \sum_{\mu \neq \nu} c_1 (3 - \text{Re tr}[\mathcal{W}_{\mu,2\nu}(n)]) \right). \quad (12)$$

Here,  $\mathcal{W}_{k\mu,l\nu}(n)$  refers to a Wilson loop of shape  $k \times l$  in the  $\mu$ - $\nu$  plane originating at the site  $n$ . The coefficients  $c_0$  and  $c_1$  are constrained by the normalization condition  $c_0 + 8c_1 = 1$  and the positivity condition  $c_0 > 0$ , where the latter condition is sufficient to guarantee that the set of configurations with minimal action consists of locally pure gauge configurations [84]. For the Wilson gauge action ( $c_1 = 0$ ), only plaquette terms contribute, whereas the DBW2 action ( $c_1 = -1.4088$ ) also involves rectangular loops.

It is well known that the critical slowing down of topological modes is more pronounced for improved gauge actions than for the Wilson gauge action [12,14,15,17,18]: A larger negative coefficient  $c_1$  suppresses small dislocations, which are expected to be the usual mechanism mediating transitions between topological sectors on the lattice. Among the most commonly used gauge actions, this effect is most severely felt by the DBW2 action. In previous works [15,17], local update algorithms were found to be inadequate for exploring different topological sectors in a reasonable time frame at lattice spacings around 0.06 fm. Instead, the authors had to generate thermalized configurations in different topological sectors using the Wilson gauge action, before using these configurations as starting points for simulations with the DBW2 action. Thus, this action allows us to explore parameters where severe critical slowing down is visible, while avoiding very fine lattice

spacings and thereby limiting the required computational resources.

### B. Observables

The observables we consider here are based on various definitions of the topological charge and Wilson loops of different sizes at different smearing levels. The unrenormalized topological charge is defined using the clover-based definition of the field strength tensor:

$$Q_c = \frac{1}{32\pi^2} \sum_{n \in \Lambda} \epsilon_{\mu\nu\rho\sigma} \text{tr} [\hat{F}_{\mu\nu}^{\text{clov}}(n) \hat{F}_{\rho\sigma}^{\text{clov}}(n)]. \quad (13)$$

This field strength tensor is given by

$$\hat{F}_{\mu\nu}^{\text{clov}}(n) = -\frac{i}{8} (C_{\mu\nu}(n) - C_{\nu\mu}(n)), \quad (14)$$

where the clover term  $C_{\mu\nu}(n)$  is defined as

$$C_{\mu\nu}(n) = P_{\mu,\nu}(n) + P_{\nu,-\mu}(n) + P_{-\mu,-\nu}(n) + P_{-\nu,\mu}(n), \quad (15)$$

and  $P_{\mu,\nu}(n)$  denotes the plaquette:

$$P_{\mu,\nu}(n) = U_\mu(n) U_\nu(n + \hat{\mu}) U_\mu^\dagger(n + \hat{\nu}) U_\nu^\dagger(n). \quad (16)$$

Alternatively, the topological charge may also be defined via the plaquette-based definition, here denoted by  $Q_p$ :

$$Q_p = \frac{1}{32\pi^2} \sum_{n \in \Lambda} \epsilon_{\mu\nu\rho\sigma} \text{tr} [\hat{F}_{\mu\nu}^{\text{plaq}}(n) \hat{F}_{\rho\sigma}^{\text{plaq}}(n)]. \quad (17)$$

Similar to the clover-based field strength tensor, the plaquette-based field strength tensor is defined as

$$\hat{F}_{\mu\nu}^{\text{plaq}}(n) = -\frac{i}{2} (P_{\mu,\nu}(n) - P_{\nu,\mu}(n)). \quad (18)$$

Both  $Q_c$  and  $Q_p$  formally suffer from  $\mathcal{O}(a^2)$  artifacts, although the coefficient is typically smaller for the clover-based definition  $Q_c$ . The topological charge is always measured after  $\mathcal{O}(30)$  steps of stout smearing [85] with a smearing parameter  $\rho = 0.12$ . To estimate the autocorrelation times of the system, we consider the squared topological charge [18], the Wilson gauge action, and  $n \times n$  Wilson loops for  $n \in \{2, 4, 8\}$  at different smearing levels. We denote the latter two by  $S_w$  and  $\mathcal{W}_n$ , respectively.

### C. Update algorithms

Throughout this work, we employ a number of different update schemes: To illustrate critical slowing down of conventional update algorithms and to set a baseline for comparison with metadynamics-based algorithms, we use

standard Hybrid Monte Carlo updates with unit length trajectories (1HMC), a single heat bath sweep (1HB), five heat bath sweeps (5HB), and a single heat bath sweep followed by four overrelaxation sweeps (1HB+4OR). The local update algorithms are applied to three distinct SU(2) subgroups during each sweep [6], and the HMC updates use an Omelyan-Mryglod-Folk fourth-order minimum norm integrator [86] with a step size of  $\epsilon = 0.2$ , which leads to acceptance rates above 99% for the parameters used here.

We compare these update schemes to metadynamics HMC updates with unit length trajectories (MetaD-HMC), and a combination of parallel tempering with metadynamics (PT-MetaD) which is discussed in more detail in Sec. VI.

An important requirement for the successful application of metadynamics is the identification of appropriate CVs. In our case, the CV should obviously be related to the topological charge. However, it should not always be (close to) integer valued but rather reflect the geometry of configuration space with respect to the boundaries between topological sectors. On the other hand, the CV needs to track the topological charge closely enough for the algorithm to be able to resolve and overcome the action barriers between topological sectors. A straightforward approach is to apply only a moderate amount of some kind of smoothing procedure, such as cooling or smearing, to a gluonic definition of the topological charge, for which we choose  $Q_c$ . Since these smoothing procedures involve spatial averaging, the action will become less local, which complicates the use of local update algorithms. Therefore, we use the HMC algorithm to efficiently update the entire gauge field at the same time, which requires a differentiable smoothing procedure, such as stout [85] or hypercubic exponential (HEX) [87] smearing. Due to its simpler implementation compared to HEX smearing, we choose stout smearing here. Previous experience [88] seems to indicate that for  $Q_c$ , four to five stout smearing steps with a smearing parameter  $\rho = 0.12$  strike a reasonable balance between having a smooth CV and still representing the topological charge accurately. We found that using  $Q_p$  would require significantly more smearing steps, whereas some improved definitions involving more general rectangular loops would not reduce the necessary amount of smearing. The force contributed by the bias potential may be written in terms of the chain rule:

$$F_{\mu,\text{meta}}(n) = -\frac{\partial V_{\text{meta}}}{\partial Q_{\text{meta}}} \frac{\partial Q_{\text{meta}}}{\partial U_{\mu_s}^{(s)}(n_s)} \times \frac{\partial U_{\mu_s}^{(s)}(n_s)}{\partial U_{\mu_{s-1}}^{(s-1)}(n_{s-1})} \dots \frac{\partial U_{\mu_1}^{(1)}(n_1)}{\partial U_{\mu}^{(1)}(n)}. \quad (19)$$

Here we have introduced the notation  $V_{\text{meta}}$  for the bias potential and  $Q_{\text{meta}}$  for the CV to clearly distinguish it from other definitions of the topological charge. Note that there

is an implicit summation over the lattice sites  $n_i$  and the Lorentz indices  $\mu_i$ . The first term in the equation, corresponding to the derivative of the bias potential with respect to  $Q_{\text{meta}}$ , is trivial, but the latter two terms are more complicated: The derivative of  $Q_{\text{meta}}$  with respect to the maximally smeared field  $U^{(s)}$  is given by a sum of staples with clover term insertions, and the final terms correspond to the stout force recursion [85] that also appears during the force calculation when using smeared fermions. Note that in machine learning terminology, this operation is essentially a backpropagation [89] and may be computed efficiently using reverse mode automatic differentiation. More details on the calculation of the force can be found in Appendices B–D.

The bias potential is constructed from a sum of one-dimensional Gaussians, as described in Sec. II, and stored as a histogram. Due to charge conjugation symmetry, we can update the potential symmetrically. Values at each point are reconstructed by linearly interpolating between the two nearest bins, and the derivative is approximated by their finite difference. To limit the evolution to relevant regions of the phase space, we introduce an additional penalty term to the potential once the value of  $Q_{\text{meta}}$  has crossed certain thresholds  $Q_{\text{min}}$  and  $Q_{\text{max}}$ . If the system has exceeded the threshold, then the potential is given by the outermost value of the histogram, plus an additional term that scales quadratically with the distance to the outer limit of the histogram.

Unless mentioned otherwise, we have used the following values as default parameters for the potential:  $Q_{\text{max/min}} = \pm 8$ ,  $n_{\text{bins}} = 800$ ,  $w = 0.05$ , while  $\delta Q^2$  has always been set equal to the bin width, i.e.,  $(Q_{\text{max}} - Q_{\text{min}})/n_{\text{bins}}$ .

It is often convenient to build up a bias potential in one or several runs, and then simulate and measure with a static potential generated from the previous runs. In some sense, this can be thought of as a combination of metadynamics and multicanonical simulations.

#### IV. RESULTS WITH CONVENTIONAL UPDATE ALGORITHMS

To establish a baseline to compare our results to, we have investigated the performance of some conventional update algorithms using the Wilson and DBW2 gauge actions. Furthermore, we have made a rough determination of the gradient flow scales  $t_0$  and  $w_0$  for the DBW2 action. Some preliminary results for the Wilson action were already presented in [88].

##### A. Critical slowing down with Wilson and DBW2 gauge actions

In order to study the scaling of autocorrelations for different update schemes, we have performed a series of simulations with the Wilson gauge action on a range of lattice spacings. The parameters were chosen to keep the physical volume approximately constant at around  $(1.1 \text{ fm})^4$ , using

TABLE I. A summary of the simulation parameters for the Wilson gauge action runs using conventional update algorithms. The scale was set via the rational fit from [90] (where  $r_0 = 0.49 \text{ fm}$ ), which in turn used data from [91].

$\beta$	$L/a$	$a$ [fm]	$N_{\text{conf}}$
5.8980	10	0.1097	100000
6.0000	12	0.0914	100000
6.0938	14	0.0783	100000
6.1802	16	0.0686	100000
6.2602	18	0.0610	100000
6.3344	20	0.0549	100000
6.4035	22	0.0499	100000

the scale given by the rational fit function in [90], which was based on data from [91]. A summary of the simulation parameters can be found in Table I.

Since autocorrelation times near second-order phase transitions are expected to be described by a power law, we use the following fit ansatz in an attempt to parametrize the scaling:

$$\tau_{\text{int}} = c \left( \frac{r_0}{a} \right)^z. \quad (20)$$

All autocorrelation times and their uncertainties are estimated following the procedure described in [92]. Figure 1 shows the scaling of the integrated autocorrelation times of  $2 \times 2$  Wilson loops  $\mathcal{W}_2$  and the square  $Q_c^2$  of the clover-based topological charge with the lattice spacing. Additionally, the figure also includes power law fits to the data and the resulting values for the dynamical critical exponents  $z(\mathcal{W}_2)$  and  $z(Q_c^2)$ . Both observables were measured after 31 stout smearing steps with a smearing parameter  $\rho = 0.12$ . While the integrated autocorrelation times of both observables increase towards finer lattice spacings and are adequately described by a power law behavior, the increase is much steeper for the squared topological charge than for the smeared  $2 \times 2$  Wilson loops. Below a crossover point at  $a \approx 0.08 \text{ fm}$ , the autocorrelation times of the squared topological charge start to dominate. They can be described by both a dynamical critical exponent  $z \approx 5$  or, alternatively, by an exponential increase that was first suggested in [16]. This behavior is compatible with the observations in [18].

In contrast, the autocorrelation time of Wilson loops is compatible with a much smaller exponent  $z \approx 1-2$ . As can be seen in Table II, the critical exponent does not change significantly with the size of the Wilson loop after 31 stout smearing steps. Generally, the integrated autocorrelation times of smeared Wilson loops slightly increase both with the size of the loops and the number of smearing levels. The only exception to this behavior occurs for larger loops, where a few steps of smearing are required to obtain a clean signal and not measure the autocorrelation of the noise instead.

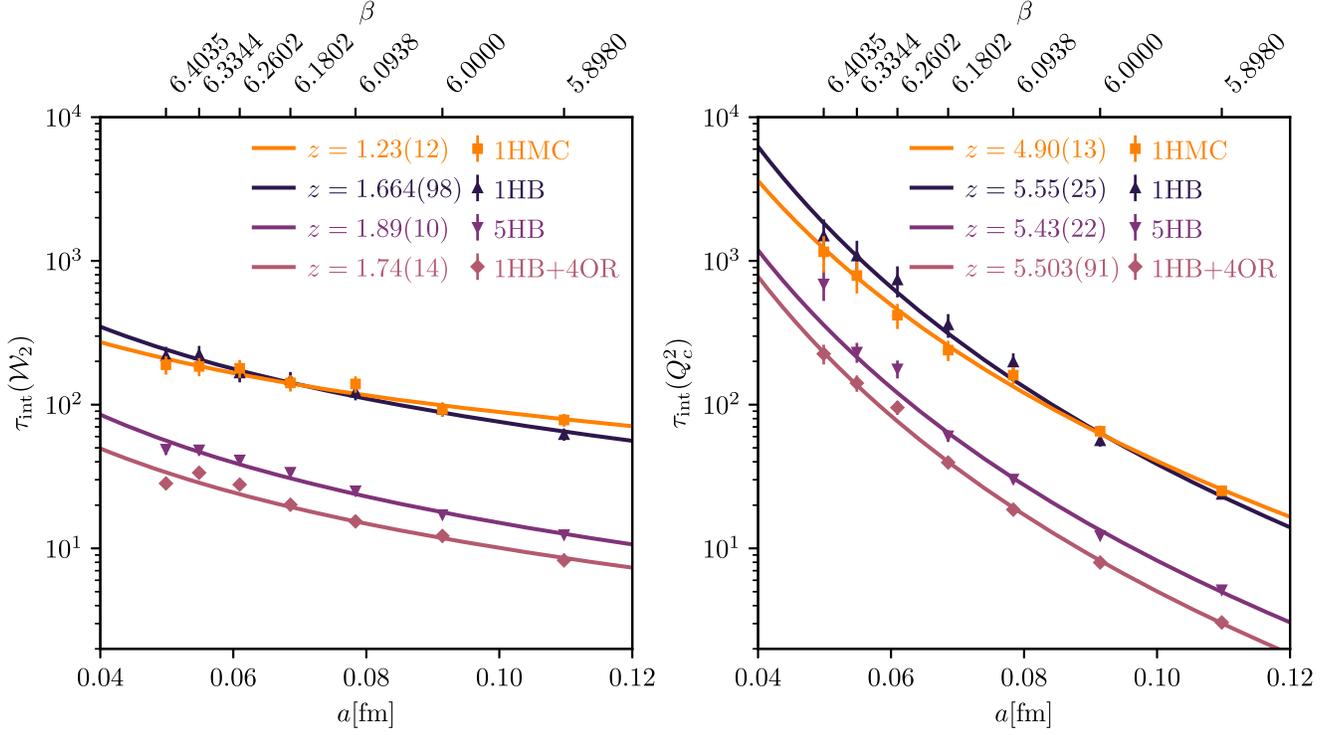


FIG. 1. Scaling of the integrated autocorrelation times of square  $2 \times 2$  Wilson loops  $\mathcal{W}_2$  (left) and the squared topological charge  $Q_c^2$  (right) for different update schemes using the Wilson gauge action. The scaling of both observables can be described using a power law fit [Eq. (20)] and is compatible with a dynamical critical exponent  $z \approx 2$  for the Wilson loops and  $z \approx 5$  for the squared topological charge. Details on the simulation parameters are listed in Table I.

Regarding the different update algorithms, the unit length HMC does show a somewhat better scaling behavior for all observables than the local update algorithms, but is also the most computationally expensive update scheme considered here (see Table III).<sup>1</sup> For all the local update algorithms, the critical exponents are very similar, but the combination of one heat bath and four overrelaxation steps has the smallest prefactor. It is interesting to note that this algorithm is also approximately twice as fast as the five-step heat bath update scheme, while still providing smaller autocorrelation times. The single step heat bath without overrelaxation, although numerically cheapest, exhibits the worst prefactor of the local update algorithms. Note that the reported numbers for the critical exponents differ from those in [88] due to a different fit ansatz (in the proceedings, the ansatz included an additional constant term).

For the DBW2 action, the problem is more severe. Figure 2 shows the time series of the topological charge for two runs using the 1HMC and the 1HB+4OR update scheme. Both simulations were carried out on  $16^4$  lattices at  $\beta = 1.25$  using the DBW2 action.

<sup>1</sup>Since we are ultimately interested in dynamical fermion simulations, we do not consider the more efficient local HMC variant presented in [93], as it is applicable to local bosonic actions only.

TABLE II. The dynamical critical exponents obtained from power law fits to the integrated autocorrelation times of  $Q_c^2$ ,  $S_w$ , and Wilson loops of different sizes after 31 stout smearing steps for different update schemes and the Wilson gauge action. Notably, the dynamical critical exponents associated with  $Q_c^2$  are much larger than those associated with the smeared action or smeared Wilson loops of different sizes.

Update scheme	$z(Q_c^2)$	$z(S_w)$	$z(\mathcal{W}_2)$	$z(\mathcal{W}_4)$	$z(\mathcal{W}_8)$
1HMC	4.90(13)	1.27(12)	1.23(12)	1.16(12)	1.29(16)
1HB	5.55(25)	1.69(10)	1.66(10)	1.64(9)	1.82(12)
5HB	5.43(22)	1.92(11)	1.89(10)	1.85(10)	1.95(10)
1HB+4OR	5.50(9)	1.77(15)	1.74(14)	1.71(14)	1.85(13)

TABLE III. Relative performance of the update algorithms used in our scaling runs. The results cited here are taken from simulations on  $22^4$  lattices. Note that the performance of the heat bath algorithm is slightly better for larger  $\beta$  [4,5].

Update scheme	Relative time
1HMC	6.98
1HB	1.00
5HB	4.99
1HB+4OR	2.02

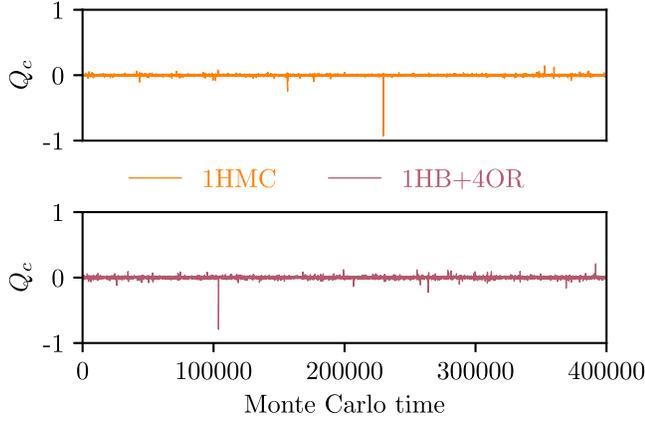


FIG. 2. Time series of the topological charge for  $V = 16^4$ ,  $\beta = 1.25$  using the DBW2 action. The configurations were generated with the 1HMC (top) and the 1HB+4OR (bottom) update schemes. Out of a total of 400000 configurations each, only a single configuration during the 1HB+4OR run and two (successive) configurations during the 1HMC run fulfill the condition  $|Q_c| > 0.5$ .

Evidently, both update schemes are unable to tunnel between different topological sectors in a reasonable time. Only a single configuration during the 1HB+4OR run and two (successive) configurations during the 1HMC run fulfill the condition  $|Q_c| > 0.5$ .

### B. Scale setting for the DBW2 action

To the best of our knowledge, scales for the DBW2 action in pure gauge theory have only been computed based on simulations with  $\beta \leq 1.22$  [17,94], and interpolation formulas are only available based on data with  $\beta \leq 1.04$  [95]. Since here we perform simulations at  $\beta = 1.25$ , we compute approximate values for  $t_0$  [96] and  $w_0$  [97], which allows us to estimate our lattice spacings for comparison to the Wilson results. Both scales are based on the energy density  $E$ , which is defined as follows:

$$E = \frac{1}{4a^2V} \sum_{n \in \Lambda} F_{\mu\nu}^a(n) F_{\mu\nu}^a(n) = -\frac{1}{2a^2V} \sum_{n \in \Lambda} \text{tr}[F_{\mu\nu}(n) F_{\mu\nu}(n)]. \quad (21)$$

Similar to the topological charge definitions, we adopt a plaquette- and clover-based definition of the field strength tensor, with the only difference being that the components are also made traceless, and not just anti-Hermitian. The gradient flow scales  $t_0$  and  $w_0$  are both defined implicitly:

$$\mathcal{E}(t_0) = t^2 \langle E \rangle \Big|_{t=t_0} = 0.3, \quad (22)$$

$$W(w_0^2) = t \frac{d}{dt} \mathcal{E}(t) \Big|_{t=w_0^2} = 0.3. \quad (23)$$

TABLE IV. Results for different gradient flow scales for the DBW2 gauge action. These results should not be interpreted as an attempt at a precise scale determination but rather as an approximate estimate.

$\beta$	$N_t \times N_s^3$	$t_{0,\text{plaq}}/a^2$	$t_{0,\text{clov}}/a^2$	$w_{0,\text{plaq}}^2/a^2$	$w_{0,\text{clov}}^2/a^2$
1.04	$48 \times 32^3$	3.445(3)	3.647(3)	3.601(4)	3.641(4)
1.10	$48 \times 32^3$	4.483(6)	4.684(6)	4.675(9)	4.716(9)
1.15	$48 \times 32^3$	5.549(9)	5.751(10)	5.787(14)	5.827(14)
1.16	$48 \times 32^3$	5.761(9)	5.962(9)	5.992(15)	6.032(15)
1.17	$48 \times 32^3$	6.032(8)	6.234(8)	6.291(14)	6.332(13)
1.18	$48 \times 32^3$	6.269(13)	6.470(14)	6.525(24)	6.566(24)
1.19	$48 \times 32^3$	6.524(10)	6.726(10)	6.803(12)	6.844(12)
1.20	$48 \times 32^3$	6.798(15)	7.000(15)	7.082(20)	7.123(20)
1.21	$48 \times 32^3$	7.047(16)	7.248(16)	7.331(25)	7.372(25)
1.22	$48 \times 32^3$	7.386(23)	7.588(24)	7.710(35)	7.751(35)
1.23	$48 \times 32^3$	7.642(23)	7.844(24)	7.954(35)	7.995(35)
1.24	$48 \times 32^3$	7.963(23)	8.165(23)	8.293(35)	8.334(35)
1.25	$48 \times 32^3$	8.312(27)	8.515(28)	8.681(37)	8.721(37)

The flow equation was integrated using the third-order commutator free Runge-Kutta scheme from [96] with a step size of  $\epsilon = 0.025$ . Measurements of the clover-based energy density were performed every ten integration steps, and  $t^2 \langle E(t) \rangle$  was fitted with a cubic spline, which was evaluated with a step size of 0.001. For every value of  $\beta$ , two independent simulations with 100 measurements each were performed on  $48 \times 32^3$  lattices. Every measurement was separated by 200 update sweeps with the previously described 1HB+4OR update scheme, and the initial 2000 updates were discarded as thermalization phase. Our results are displayed in Table IV.

Using the physical value of  $\sqrt{t_0} = 0.1638(10)$  fm from [98], these results imply a physical volume of approximately  $(0.9 \text{ fm})^4$  and a temperature of around 219 MeV for the  $16^4$  lattice from the previous section.

In order to facilitate comparison with other results, we also provide an interpolation of our lattice spacing results. For this purpose, we use a rational fit ansatz with three fit parameters that is asymptotically consistent with perturbation theory [90] and has a sufficient number of degrees of freedom to describe our data well:

$$\ln(t_0/a^2) = \frac{8\pi^2}{33} \beta \frac{1 + d_1/\beta + d_2/\beta^2}{1 + d_3/\beta}. \quad (24)$$

For our reference, clover-based  $t_0$  scale setting, this results in a fit with  $\chi^2/\text{d.o.f.} \approx 1.31$  and parameters  $d_1 \approx 1.0351$ ,  $d_2 \approx -1.3763$ ,  $d_3 \approx 0.4058$ , which is displayed in Fig. 3. We want to emphasize that these results are not meant to be an attempt at a precise scale determination but rather only serve as an approximate estimate. Especially for the finer lattices, the proper sampling of the topological sectors

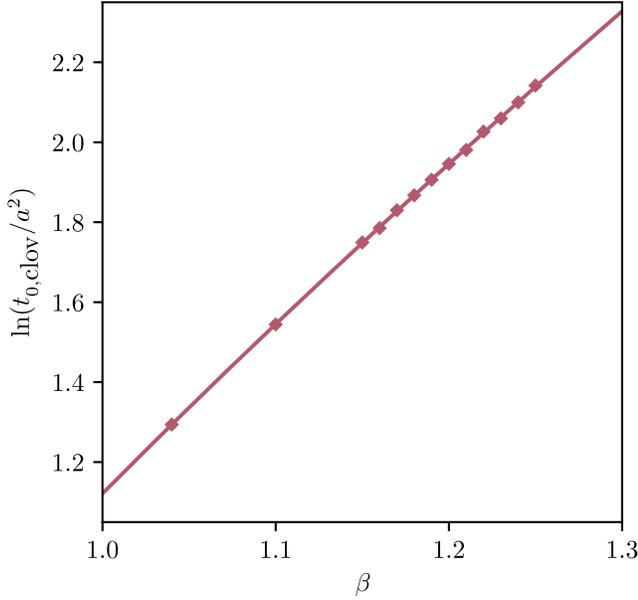


FIG. 3. Rational fit of the form Eq. (24) to the  $t_{0,\text{clov}}/a^2$  values presented in Table IV. The fit has  $\chi^2/\text{d.o.f.} \approx 1.31$  and parameters  $d_1 \approx 1.0351$ ,  $d_2 \approx -1.3763$ , and  $d_3 \approx 0.4058$ . Error bars are substantially smaller than the symbols.

cannot be guaranteed, and the comparatively small volumes may introduce non-negligible finite volume effects.

## V. RESULTS WITH METADYNAMICS

Figure 4 shows the time series of the topological charge obtained from simulations with the HMC and the MetaD-HMC with five and ten stout smearing steps on a  $22^4$  lattice at  $\beta = 6.4035$  using the Wilson gauge action.

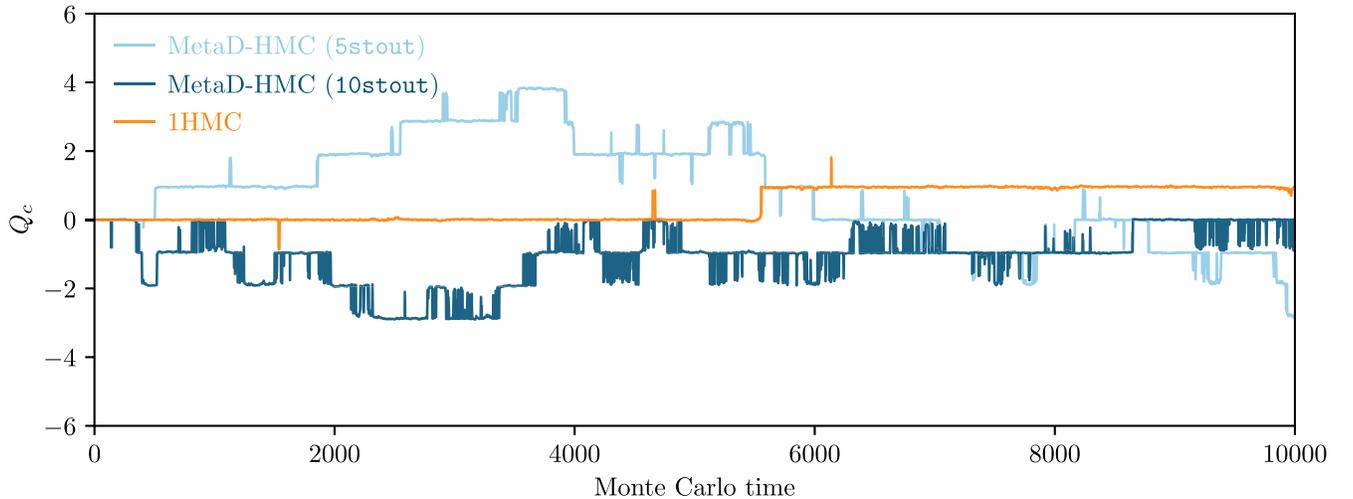


FIG. 4. Comparison of the time series of the topological charge between runs using the HMC algorithm and MetaD-HMC runs for  $V = 22^4$  and  $\beta = 6.4035$  using the Wilson gauge action. The bias potentials in the metadynamics runs were built up dynamically/from scratch during each run. Both metadynamics runs are able to transition between topological sectors numerous times, whereas the run using the conventional HMC is essentially stuck in two sectors.

Both MetaD-HMC runs tunnel multiple times between different topological sectors, whereas the conventional HMC run essentially displays only a single tunneling event between sectors  $Q = 0$  and  $Q = 1$ . The autocorrelation times of  $Q_c^2$  for the MetaD-HMC runs are comparable despite the different amounts of smearing used to define the CV: For the run with five smearing steps  $\tau_{\text{int}}(Q_c^2) = 586(142)$ , while  $\tau_{\text{int}}(Q_c^2) = 342(68)$  for the run with ten smearing steps. A noteworthy difference between the two MetaD-HMC runs is the increase of fluctuations with higher amounts of smearing. If too many smearing steps are used to define the CV, the resulting  $Q$  values will generally be closer to integers, which will eventually drive the system to coarser regions of configuration space. Since these regions do not contribute significantly to expectation values in the path integral, it is desirable to minimize the time that the algorithm spends there. This is directly related to the issue of small effective sample sizes after reweighting, which we will discuss in more detail in Sec. VB.

A similar comparison of topological charge time series for the DBW2 action can be seen in Fig. 5. Here, two MetaD-HMC runs with four and five stout smearing steps on a  $16^4$  lattice at  $\beta = 1.25$  are compared to the 1HMC and 1HB+4OR runs, which were already shown in Fig. 2. Both conventional update schemes are confined to the zero sector, whereas the two MetaD-HMC runs explore topological sectors up to  $|Q| = 6$ . More quantitatively, the integrated autocorrelation time of  $Q_c^2$  is estimated to be  $\tau_{\text{int}}(Q_c^2) = 5126(1500)$  for the run with four smearing steps and  $\tau_{\text{int}}(Q_c^2) = 4159(1137)$  for the run with five smearing steps. On the other hand, lower bounds for the autocorrelation times of the 1HMC and 1HB+4OR update schemes are  $4 \times 10^5$ , which is larger by more than a factor of 70.

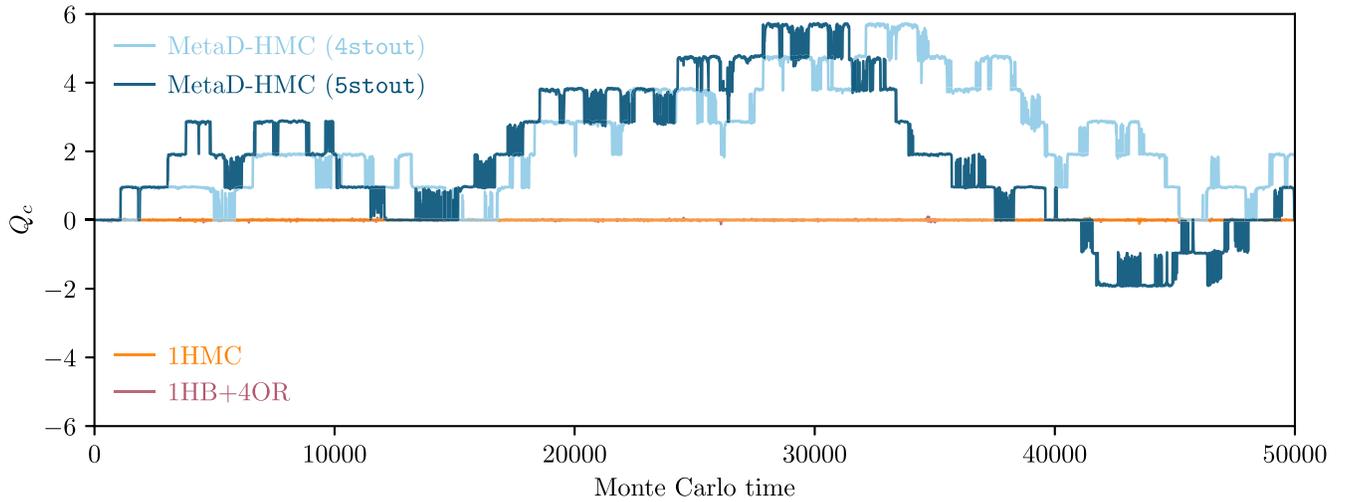


FIG. 5. Comparison of the time series of the topological charge between runs using conventional update algorithms (HMC and a combination of heat bath and overrelaxation updates) and MetaD-HMC runs for  $V = 16^4$  and  $\beta = 1.25$  using the DBW2 action. The bias potentials in the metadynamics runs were built up dynamically during each run. The results shown for the 1HMC and 1HB+4OR update schemes are from the same runs as the time series shown in Fig. 2. While the conventional update algorithms are unable to escape the  $Q = 0$  sector, both metadynamics runs frequently transition between different topological sectors.

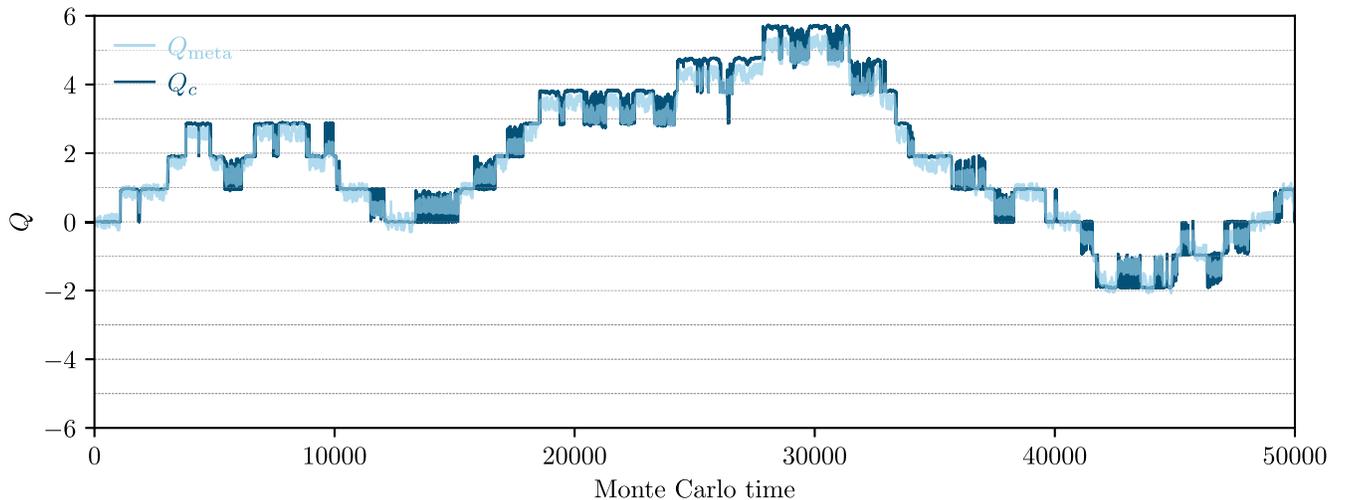


FIG. 6. Time series of the CV  $Q_{\text{meta}}$  and the topological charge  $Q_c$ , measured after five and 30 stout smearing steps with a smearing parameter of  $\rho = 0.12$ , respectively. The data are from the 5stout MetaD-HMC run shown in Fig. 5.

To illustrate the role of the CV  $Q_{\text{meta}}$ , it may be helpful to compare the time series of  $Q_{\text{meta}}$  and  $Q_c$ , as shown in Fig. 6. The two observables are clearly correlated, but  $Q_{\text{meta}}$  is distributed more evenly between integers.

### A. Computational overhead and multiple timescale integration

A fair comparison of the different update schemes also needs to take the computational cost of the algorithms into account. Table V shows the relative timings for the different update schemes used in this section, measured for simulations carried out on  $16^4$  lattices. While no significant

efforts were made to optimize the performance of our implementation of the MetaD-HMC, it is still clear that the additional overhead introduced by the computation of the metadynamics force contribution is significant for pure gauge theory. The relative overhead is especially large compared to local update algorithms, which are already more efficient than the regular HMC. Note, however, that due to the more nonlocal character of the DBW2 gauge action, the relative loss in efficiency when switching to metadynamics from either a local update algorithm or the HMC is already noticeably smaller.

Since the majority of the computational overhead comes from the metadynamics force contribution, and

TABLE V. Relative timings for the different update schemes measured for simulations carried out on  $16^4$  lattices. The significant computational overhead for the metadynamics updates compared to the other algorithms is due to the stout smearing and the stout force recursion required for the metadynamics force calculation.

Update scheme	Relative time	
	Wilson action	DBW2 action
1HB+4OR	1	1
1HMC	2.90	4.21
MetaD-HMC (4stout)	61.91	24.27
MetaD-HMC (5stout)	73.64	28.11

the involved scales are different from those relevant for the gauge force, it seems natural to split the integration into multiple timescales in a similar fashion to the Sexton-Weingarten scheme [99]: The force contributions from the bias potential are correlated to the topological charge, which is an IR observable, whereas the gauge force is usually dominated by short-range UV fluctuations. Therefore, it is conceivable that integrating the metadynamics force contribution on a coarser timescale than the gauge force could significantly decrease the required computational effort, while still being sufficiently accurate to lead to reasonable acceptance rates.

We have attempted to use combinations of both the leapfrog and the Omelyan-Mryglod-Folk second-order integrator with the Omelyan-Mryglod-Folk fourth-order minimum norm integrator in a multiple timescale integration scheme. Unfortunately, we were unable to achieve a meaningful reduction of metadynamics force evaluations without encountering integrator instabilities and deteriorating acceptance rates. However, this approach might still be helpful for simulations with dynamical fermions, where it is already common to split the forces into more than two levels.

Even if such a multiple timescale approach should turn out to be unsuccessful in reducing the number of metadynamics force evaluations, we expect the relative overhead of metadynamics to be much smaller for simulations including dynamical fermions. In a previous study [42] it was found that compared to conventional HMC simulations, simulations with metadynamics using 20 steps of stout smearing were about 3 times slower in terms of real time.

## B. Scaling of the reweighting factor and improvements to the bias potential

Due to the addition of the bias potential, unweighted averages do not lead to expectation values with respect to the original physical probability density. If corrected with a reweighting procedure, then the overlap between the sampled distribution and the distribution of physical

interest needs to be sufficiently large for the method to work properly. A common measure to quantify the efficiency of the reweighting procedure is the effective sample size (ESS), defined as

$$\text{ESS} = \frac{(\sum_i w_i)^2}{\sum_i w_i^2}, \quad (25)$$

where  $w_i$  is the respective weight associated with each individual configuration. In the case of a static bias potential, this is simply  $e^{V(Q_{\text{meta},i})}$ . We found the normalized ESS, i.e., the ESS divided by the total number of configurations, to generally be of order  $\mathcal{O}(10^{-2})$  or lower when simulating in regions of parameter space where conventional algorithms fail to sample different topological sectors.

Although the low ESS ultimately results from the fact that the bias potential is constructed in such a way as to have a flat marginal distribution over the CV, we can nonetheless distinguish between two contributions towards this effect. On the one hand, there is the inevitable flattening of the intersector barriers by the bias potential, which is necessary to facilitate tunneling between adjacent topological sectors. On the other hand, however, the different weights of the individual topological sectors are also canceled by the bias potential. While it is necessary to faithfully reproduce the intersector barriers, the leveling of the weights of the different topological sectors is often unwanted. It increases the time that the simulation spends at large values of  $|Q|$ , so that these sectors are over-represented compared to their true statistical weight. It is therefore conceivable that by retaining only the intersector barrier part of the bias potential, the relative weights of the different topological sectors will be closer to their physical values, and the ESS will increase. The resulting marginal distribution over the topological charge is then expected to no longer be constant, but rather resemble a parabola. In cases where this modification to the bias potential is used, we will either explicitly mention it or include the abbreviation “mod.” to make a clear distinction between it and the original potential.

Here and in Sec. VI of this work, we perform scaling tests of the proposed improvements in two-dimensional U(1) gauge theory, where high statistics can be generated more easily than in four-dimensional SU(3) gauge theory. The action is given by the standard Wilson plaquette action,

$$S_g = \beta \sum_{n \in \Lambda} (1 - \text{Re}[P_{t,x}(n)]), \quad (26)$$

and updates are performed with a single-hit Metropolis algorithm. The topological charge is defined using a geometric integer-valued definition:

$$Q = \frac{1}{2\pi} \text{Im} \left[ \sum_{n \in \Lambda} \log P_{t,x}(n) \right]. \quad (27)$$

For all metadynamics updates, we use a field-theoretic definition of the topological charge that is generally not integer valued:

$$Q_{\text{meta}} = \frac{1}{2\pi} \text{Im} \left[ \sum_{n \in \Lambda} P_{l,x}(n) \right]. \quad (28)$$

Since the charge distributions obtained from the two definitions already show reasonable agreement without any smearing for the parameters considered here, we can use local update algorithms and directly include the metadynamics contribution in the staple. A similar idea that encourages tunneling in the Schwinger model by adding a small modification to the action was proposed in [100].

In previous tests in two-dimensional U(1) gauge theory, we found that the bias potentials could be described by a sum of a quadratic and multiple oscillating terms [101]:

$$V(Q) = A Q^2 + \sum_{i=1}^N B_i \sin^2(\pi f_i Q). \quad (29)$$

Here, we fit our bias potentials that are obtained from the two-dimensional U(1) simulations to this form. We then obtain a modified bias potential by subtracting the resulting quadratic term from the data.

Table VI contains the normalized ESS and integrated autocorrelation times for different lattice spacings on the same line of constant physics in two-dimensional U(1) theory. We compare metadynamics runs using bias potentials obtained directly from previous simulations with metadynamics runs using potentials that were modified to retain the relative weights of the topological sectors as

TABLE VI. Normalized effective sample sizes for simulations carried out on different lattices on the same line of constant physics in two-dimensional U(1) gauge theory. For each set of parameters,  $10^7$  measurements were performed with a separation of ten update sweeps between every measurement. More details on the simulation setup can be found in Sec. VB.

$L/a$	$\beta$	ESS/ $n_{\text{meas}}$	$\tau_{\text{int}}(Q^2)$
Regular bias potential			
16	3.2	33.088(74)%	52
20	5.0	21.81(11)%	207
24	7.2	12.70(11)%	567
28	9.8	8.805(81)%	676
32	12.8	8.261(85)%	1167
Modified bias potential (see Sec. VB)			
16	3.2	99.50627(11)%	5
20	5.0	64.3084(87)%	33
24	7.2	28.572(12)%	123
28	9.8	27.291(16)%	227
32	12.8	18.751(25)%	247

described above. We see large improvements for both the ESS and  $\tau_{\text{int}}$  in the modified case, even for the finest lattices considered.

We expect that the quadratic term is mostly relevant for small volumes and high temperatures. With larger volumes and lower temperatures, the slope should decrease, and with it the importance of correctly capturing this term. On the other hand, the oscillating term is expected to grow more important with finer lattice spacings, as the barriers between the different sectors grow steeper. Thus, the oscillating term needs to be described more and more accurately towards the continuum.

A standard technique to decrease, but not completely eliminate, action barriers is well-tempered metadynamics [102]. In this approach, the height of the added Gaussians  $w$  decays with increasing potential. In our tests, we found that this method increases the ESS at the cost of higher autocorrelation times. Whether the gains of the ESS outweigh the loss from the higher autocorrelation times depends on the choice of parameters. Although this technique might yield moderate improvements in the overall sampling efficiency, we decided not to attempt any fine-tuning of the parameters at this point.

### C. Accelerating the equilibration/buildup of the bias potential

Another avenue of improvement is accelerating the buildup of the bias potential. This aspect becomes especially relevant when considering large-scale simulations, where current simulations are typically limited to  $\mathcal{O}(10^4)$  update sweeps, and a lengthy buildup phase of the bias potential would render the method infeasible. Additionally, the range of relevant  $Q$  values will also increase with larger physical volumes.

The first idea explored here is to exploit the aforementioned well-tempered variant of metadynamics, by choosing a larger starting value of the Gaussian height  $w$  and letting it decay slowly so as to minimize the change in the potential that arises from the decay. While this approach introduces the decay rate as another fine-tunable parameter, we found that this did indeed reduce the number of update iterations required to thermalize the potential. A small caveat is that an optimal choice of the decay rate requires prior knowledge on the approximate height of the action barriers.

A second way of reducing the buildup time is to use an enhancement of metadynamics which is most commonly referred to as multiple walkers metadynamics [103], where the potential is simultaneously built up by several independent streams in a trivially parallelizable way. In our case, it is convenient to make each stream start in a distinct topological sector. In two-dimensional U(1) gauge theory, this can be achieved by seeding each stream with an instanton configuration of charge  $Q$ , which can be constructed according to [104],

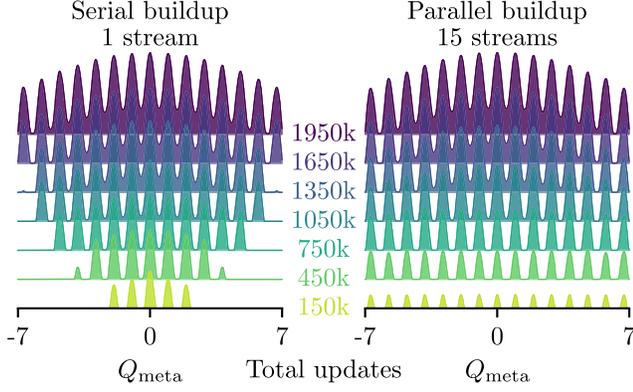


FIG. 7. Comparison of serial and parallel buildup of the bias potential in two-dimensional U(1) gauge theory for  $32^2$  lattices. The ratio of update iterations per stream was held fixed at 15:1, so that both methods would use the same number of total update steps during each snapshot.

$$U_t^l(Q; t, x) = \exp\left(-2\pi i x \frac{Q}{N_x N_t}\right),$$

$$U_x^l(Q; t, x) = \exp\left(2\pi i t \frac{Q}{N_t} \delta_{x, N_x}\right). \quad (30)$$

The serial and parallel buildup are compared in Fig. 7 where the potential parameters for each stream are given by:  $Q_{\max/\min} = \pm 7$ ,  $n_{\text{bins}} = 1400$ , and  $w = 0.002$ . In the case of four-dimensional SU(3) the direct construction of instantons with higher charge is not quite as simple as in two-dimensional U(1) gauge theory. The construction of lattice instantons with even charge is described in [104], and lattice instantons with odd charge can be constructed by combining multiple instantons with charge  $|Q| = 1$  [88,105]. Regardless, starting with instantons is not required, since we only need each stream to fall into the specified sector. The time until the streams start to tunnel is a first indicator of the thermalization timescale of the potential.

Independent of the possible improvements mentioned here, a fine-tuning of the standard metadynamics parameters could also prove to be worthwhile in regard to accelerating the buildup and improving the quality of the bias potential.

## VI. COMBINING METADYNAMICS WITH PARALLEL TEMPERING

In order to eliminate the problem of small effective sample sizes observed in our metadynamics simulations due to the required reweighting, we propose to combine metadynamics with parallel tempering [77]. This is done in a spirit similar to the parallel tempering on a line defect proposed by Hasenbusch [30]. We introduce two simulation streams: One with a bias potential, and the other without it, while the physical actions  $S(U)$  are the same for

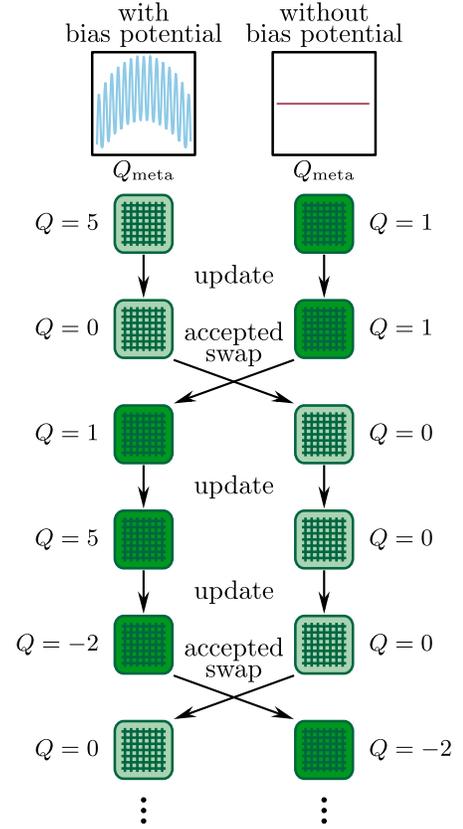


FIG. 8. Illustration of the PT-MetaD algorithm, with the two columns representing two simulation streams. The bias potentials are illustrated at the top, while subsequent rows represent the time series, with different shades of the configurations for better visibility.  $Q$  values are indicative and do not represent results from an actual simulation.

both streams, as illustrated in Fig. 8. Since we are working in pure gauge theory, this means the second stream without bias potential can be simulated with local update algorithms. After a fixed number of updates have been performed on the two streams, a swap of the configurations is proposed and subject to a standard Metropolis accept-reject step. The action difference is given by

$$\Delta S_t^M = [S_t^M(U_1) + S(U_2)] - [S_t^M(U_2) + S(U_1)]$$

$$= V_t(Q_{\text{meta},1}) - V_t(Q_{\text{meta},2}), \quad (31)$$

where the indices of the quantities denote the number of the stream, and  $V_t$  is the bias potential in the first stream. It is apparent and important to note that the action difference is simple to compute regardless of what the physical action looks like. Even in simulations where dynamical fermions are present, the contributions from the physical action are always canceled out by virtue of the two streams having the same action parameters; only the contribution from the bias potential remains. Moreover, the action differences,

and thus the swap rates, should be largely independent of the volume.

Since the second/measurement stream samples configurations according to the physical distribution no reweighting is needed, and thus the effective sample size is not reduced. Additionally, if the swaps are effective, the measurement stream will inherit the topological sampling from the stream with bias potential and thus also sample topological sectors well. Effectively, the accept-reject step for swap proposals serves as a filter for configurations with vanishing weight, thereby decreasing the statistical uncertainties on all observables weakly correlated to the topological charge. What remains to be seen is whether the efficiency of the sampling of the topological sectors carries over from the bias potential stream to the measurement stream. In this section, we address this question via both scaling tests in two-dimensional U(1) and exploratory runs in four-dimensional SU(3) in a region where conventional update algorithms are effectively frozen.

### A. Scaling tests in two-dimensional U(1)

We carried out a number of simulations in two-dimensional U(1) gauge theory using the same parameters and simulation setup as described in Sec. VB. We use bias potentials already built for these metadynamics runs and keep them static in a number of parallel tempered metadynamics runs. For each set of parameters, we carried out one run with the respective unmodified potential and one run with a potential modified as described in Sec. VB. In these runs, swaps between the two streams were proposed after each had completed a single update sweep over all lattice sites. The run parameters, as well as the resulting autocorrelation times of the topological charge  $Q$ , can be found in Table VII.

Since the relevant configuration space is now the product of configuration spaces of the individual streams, autocorrelation timescales of observables that are defined on the product space will now contain important information about the dynamics of the system. Therefore, we additionally monitor the sum of the squared topological charges on both streams. This observable allows us to distinguish the fluctuations in  $Q$  originating from true tunneling events from repeated swaps between the two streams without tunneling.

Figure 9 shows the scaling of the total amount of independent configurations, which is given by the quotient of the effective sample size Eq. (25) and the integrated autocorrelation time of the topological susceptibility. The performance of the standard Metropolis algorithm is compared to parallel tempered and standard metadynamics, with both modified (see Sec. VB) and nonmodified bias potentials.

PT-MetaD performs well throughout the entire range of lattice spacings, consistently outperforming standard metadynamics by more than an order of magnitude. Most

TABLE VII. Integrated autocorrelation times for different lattices on the same line of constant physics in two-dimensional U(1) gauge theory, using both Metropolis and PT-MetaD updates. Observables indexed with 1 are taken from the stream with bias potential, whereas those indexed with 2 are taken from the regular stream. The modification of the bias potential is discussed in Sec. VB. Overall,  $10^7$  measurements were performed with a separation of ten update sweeps between every measurement.

$L/a$	$\beta$	$\tau_{\text{int}}(Q_2^2)$	$\tau_{\text{int}}(Q_1^2 + Q_2^2)$
Single-hit Metropolis			
16	3.2	3	N/A
20	5.0	71	N/A
24	7.2	3939	N/A
28	9.8	462472	N/A
32	12.8	> 10000000	N/A
PT-MetaD (regular bias potential)			
16	3.2	5	20
20	5.0	59	142
24	7.2	939	534
28	9.8	1730	818
32	12.8	1926	1319
PT-MetaD (modified bias potential)			
16	3.2	4	5
20	5.0	47	47
24	7.2	184	208
28	9.8	316	403
32	12.8	312	465

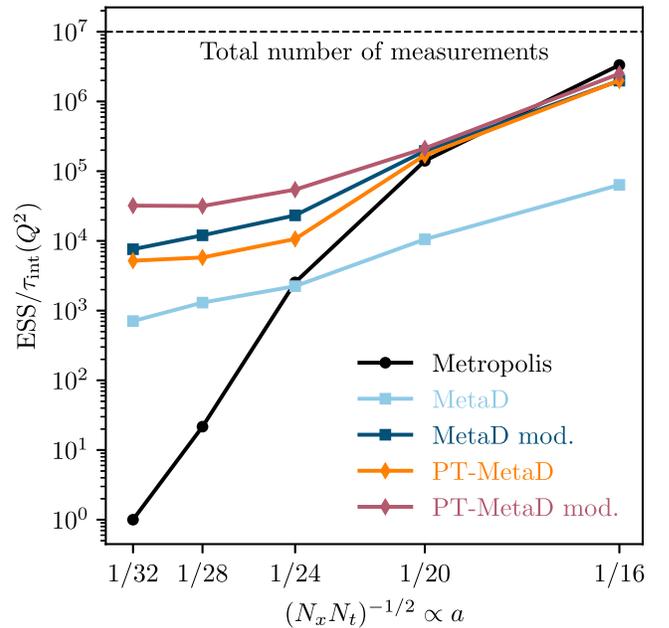


FIG. 9. Continuum scaling of the total sample size for the standard Metropolis algorithm and variations of MetaD-Metropolis in two-dimensional U(1), including the use of a modified bias potential (see Sec. VB). The corresponding data can be found in Table VI and Table VII. Lines are drawn to guide the eyes.

TABLE VIII. Normalized effective sample sizes and swap rates for PT-MetaD simulations with regular and modified bias potential in two-dimensional U(1) gauge theory, where the normalized effective sample sizes were measured on the bias potential streams of the PT-MetaD runs. The nonmonotonous behavior of the quantities for the PT-MetaD runs may be due to a suboptimal choice of bias potentials.

$L/a$	$\beta$	PT-MetaD		PT-MetaD mod.	
		ESS/ $n_{\text{meas}}$	Swap rate	ESS/ $n_{\text{meas}}$	Swap rate
16	3.2	21.0%	0.198	99.6%	0.960
20	5.0	13.6%	0.151	64.3%	0.585
24	7.2	7.3%	0.086	28.6%	0.271
28	9.8	5.0%	0.057	27.3%	0.249
32	12.8	8.3%	0.081	18.8%	0.169

importantly, the ratio of independent configurations seems to reach a plateau for finer lattice spacings, indicating an improved scaling behavior compared to conventional metadynamics, which in itself already scales significantly better than the single-hit Metropolis algorithm. It is also worth noting that the modified bias potential provides better results than the nonmodified one, as evidenced by the normalized effective sample sizes and swap rates presented in Table VIII. This is consistent with our expectation that large excursions in the topological charge, which produce irrelevant configurations, are curbed by the modification of the bias potential. A more detailed look at the effectiveness of the new algorithm is provided by Fig. 10.

It compares the results of PT-MetaD and standard metadynamics at our finest lattice spacing, with and without modification of the bias potential. For reference, exact values from analytical solutions [40,106–108] are also provided (see Appendix E for more details). First, we note that there is no significant difference in the performance between standard and parallel tempered metadynamics in the topology related observables  $Q$  and  $Q^2$  in the case of a modified bias potential. This indicates that the swaps are effective in carrying over the topological sampling of the bias potential stream to the measurement stream. On the other hand, the inclusion of the irrelevant higher sectors with the unmodified bias potential does increase the error bars, and there is some indication that not all of the topological sector sampling is carried over into the measurement run of PT-MetaD. Furthermore, Fig. 10 reveals that for observables not related to the topology, PT-MetaD with a modified bias potential is superior to standard metadynamics. This is clearly the effect of the higher effective sample size and number of independent configurations.

In summary, our scaling tests in two-dimensional U(1) suggest that parallel tempered metadynamics with a modified bias potential has much improved topological sampling, equivalent to standard metadynamics, while at the same time not suffering from a reduced effective sample size. There is some indication that the ratio of statistically independent to total configurations does reach a stable plateau in the continuum limit.

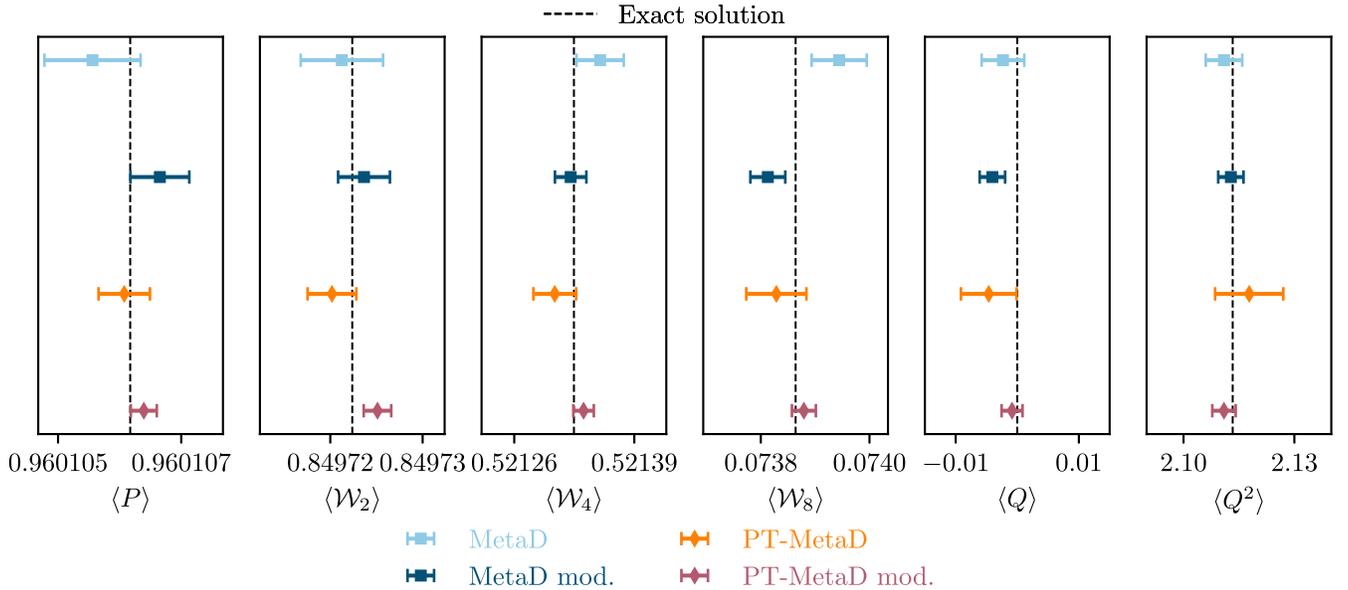


FIG. 10. Comparison of expectation values and uncertainties for the plaquette  $P$ , larger Wilson loops  $\mathcal{W}_n$ , the topological charge  $Q$ , and the topological susceptibility  $Q^2$  for variations of MetaD-Metropolis in two-dimensional U(1) for  $32^2$  lattices and  $\beta = 12.8$ . The modification of the bias potential is discussed in Sec. VB. Dashed lines correspond to the exact solutions [40,106–108]. Results obtained from a run with standard Metropolis updates are off the scale everywhere except for  $Q$ , where the expectation value is exactly equal to zero due to complete topological freezing.

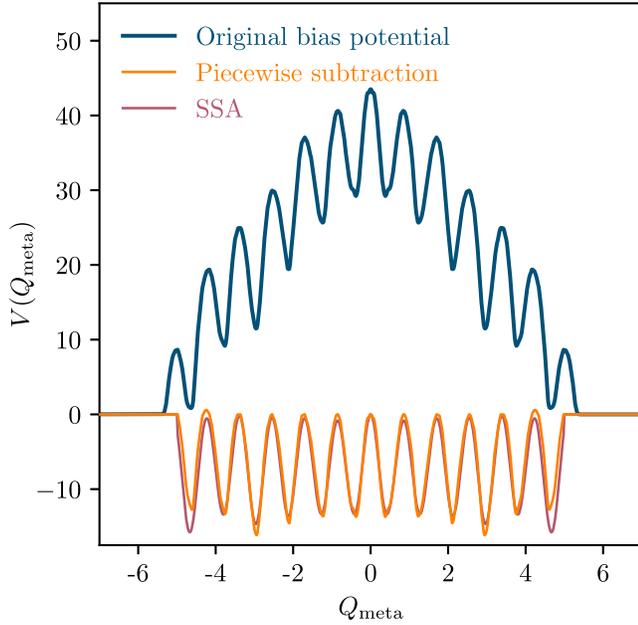


FIG. 11. Comparison between the original bias potential and its trend subtracted modifications from singular spectrum analysis and piecewise subtraction of the  $Q^2$  term.

### B. First results in four-dimensional SU(3)

For our exploratory study in four-dimensional SU(3), we turn to the DBW2 gauge action at  $\beta = 1.25$  on a  $V = 16^4$  lattice, which we have already used in Sec. V. For our first run, which is depicted in the left panels of Fig. 12, we have combined a local 1HB+4OR measurement stream with a 4stout MetaD-HMC stream that dynamically generates the bias potential. Between swap proposals, updates for the two streams are performed at a ratio of 10 1HB+4OR update sweeps to a single unit length MetaD-HMC trajectory, which roughly reflects the relative wall clock times between the algorithms. One can see that the measurement run starts exploring other topological sectors almost as soon as the parallel run with active bias potential has gained access to them. In the later stages of the run, when the bias potential is sufficiently built up to allow the metadynamics run to enter higher topological sectors, one can see that the swap rate is lowered by the action difference between the topological sectors, leading to an overall swap rate of  $\sim 8\%$ . This effect mirrors the reduction of the effective sample size in pure metadynamics updates and may be ameliorated by removing the quadratic term in the bias potential, as

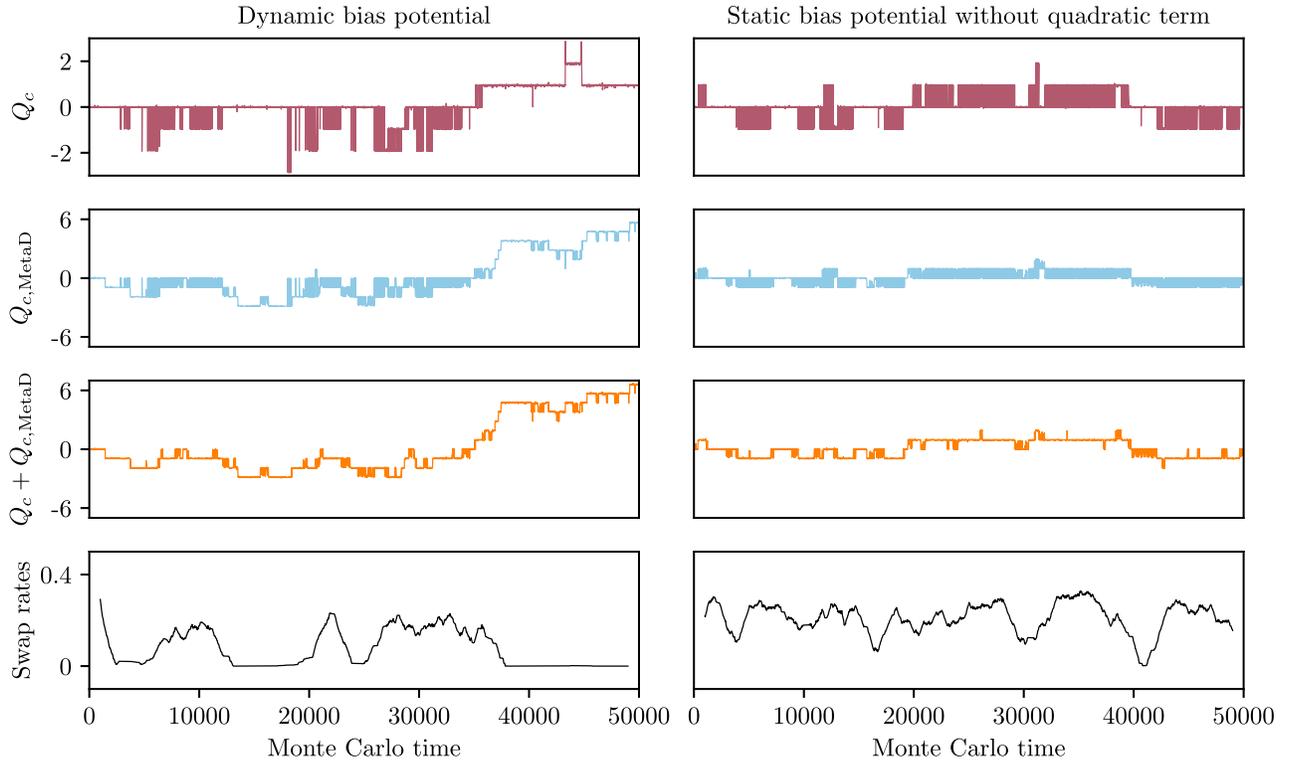


FIG. 12. Topological charge time series for our parallel tempered metadynamics runs on a  $V = 16^4$  lattice at  $\beta = 1.25$  with the DBW2 gauge action and four steps of stout smearing in the definition of  $Q_{\text{meta}}$ . The left panels show results of our first run with a dynamically built bias potential, while the right panels show our second run with a modified static potential. The topmost row shows the time series of the topological charges in the respective measurement runs, while the second row is for the metadynamics part (note the different y scales). The third row displays the sum of the topological charges on the measurement stream and the stream including a bias potential, serving as an indicator for genuine transitions of the entire system into new topological sectors. In the bottom row, the running average of the swap acceptance rate with a window size of 2000 is displayed.

discussed in Sec. V B. In fact, the relevant point is that the action difference between the maxima of the bias potential for different topological sectors reflects the relative weight of these sectors in the path integral and should not be flattened out. Ideally, we want the bias potential to only reproduce the barriers between the sectors, not their relative weights. For a second exploratory PT-MetaD run, we therefore opted for a static bias potential of this sort. Lacking data that are precise enough to model the bias potential in detail, as we did in two-dimensional U(1), we started from the bias potential of a previous metadynamics run and extracted the high frequency (in  $Q_{\text{meta}}$ ) part corresponding to the topological barriers, while eliminating the long range part corresponding to the relative weights of the topological sectors. For this purpose, we chose to perform a singular spectrum analysis [109] and cross-checked the result with a simple piecewise subtraction of the  $Q^2$  term between consecutive local maxima. As displayed in Fig. 11, both methods result in a similar modified bias potential that seems to reproduce the intersector barriers rather well.

The right panels of Fig. 12 display the results of the corresponding PT-MetaD run. Notably, excursions to large absolute values of the topological charge in the stream with bias potential are now curbed, and the swap acceptance rate has increased to  $\sim 21\%$ . In addition, the acceptance rate is approximately constant over the entire run, as it should be expected for a static bias potential. The resulting autocorrelation times are  $\tau_{\text{int}}(Q_c^2) = 53(3)$  and  $\tau_{\text{int}}(Q_c^2 + Q_{c,\text{MetaD}}^2) = 333(35)$ . We would like to emphasize that the bias potential we extracted is a rather rough guess. With a larger amount of data, it might be possible to extract a better bias potential, possibly leading to even higher acceptance rates. Considering the rather simple forms used to model the bias potential, it might also be possible to describe it with sufficient accuracy for good initial guesses at other run parameters. We plan to address these points in the future.

In any case, these first results clearly show that the parallel tempered metadynamics algorithm is able to achieve enhanced topological sampling in four-dimensional SU(3) without the reduction of the effective sample size that is typical for algorithms with a bias potential.

## VII. CONCLUSION AND OUTLOOK

In this paper, we have proposed a new update algorithm, PT-MetaD and applied it to four-dimensional SU(3) gauge theory. In its simplest form, which we have investigated here, it consists of two parallel streams simulating the same physical system. One of the streams includes a fixed bias potential, which facilitates transitions between topological sectors. This bias potential can, for example, be extracted from a metadynamics simulation of the system. The second

stream can utilize any efficient conventional update algorithm that, by itself, may exhibit topological freezing. At regular intervals, swaps between the two streams are proposed and subjected to a Metropolis accept-reject step. We have demonstrated that in this way, the good topological sampling of the stream with bias potential carries over to the conventional update stream without compromising its effective sample size. When performing measurements exclusively on the second conventional update stream, one therefore obtains topological unfreezing without any reweighting or additional modifications. The price to pay is the additional update stream including a bias potential, which requires minimal communication with the measurement stream and thus is embarrassingly parallel. The proposed algorithm may be helpful in overcoming potential barriers in more general cases without compromising the effective sample size.

We have demonstrated that PT-MetaD can unfreeze four-dimensional SU(3) gauge theory at parameter values where conventional algorithms are frozen. Furthermore, scaling tests in two-dimensional U(1) gauge theory indicate gains of more than an order of magnitude compared to standard metadynamics in the total sample size and an improved scaling of autocorrelation times with the lattice spacing compared to standard update algorithms. We have also demonstrated that the buildup of the metadynamics bias potential may be accelerated by running multiple metadynamics simulations in parallel.

We believe these results are promising and plan to study the scaling behavior of the methods tested here in more detail for four-dimensional SU(3) gauge theory, and eventually in full QCD. Conceptually, there seem to be no obstacles for implementing parallel tempered metadynamics in full QCD. We also plan to explore possible optimizations for parallel tempered metadynamics. These include optimizing the bias potential via enhanced buildup and extraction and, possibly, describing it parametrically. Furthermore, it would be interesting to investigate whether adding additional streams with the same or modified bias potentials could increase performance, despite the additional computational cost. This might be especially interesting for large scale simulations, where additional streams may offer a reduction of autocorrelation times in a trivially parallelizable manner.

## ACKNOWLEDGMENTS

We thank Philip Rouenhoff for collaboration in the early stages of this work. We gratefully acknowledge helpful discussions with Szabolcs Borsanyi, Stephan Dürr, Fabian Frech, Jana Günther, Ruben Kara, Andrey Kotov, and Kalman Szabo. Calculations were performed on a local PC cluster at the University of Wuppertal.

**APPENDIX A: CONVENTIONS**

We consider the Lie group  $SU(3)$  along with the associated Lie algebra  $\mathfrak{su}(3)$ . The generators  $T^a$  may be represented as traceless anti-Hermitian  $3 \times 3$  matrices with the group index  $a \in \{1, \dots, 8\}$ . In terms of the Gell-Mann matrices, the generators may be chosen to be

$$T^a = \frac{i}{2} \lambda^a, \quad (\text{A1})$$

but more generally, the generators are normalized in such a way that

$$\text{tr}[T^a T^b] = -\frac{1}{2} \delta_{ab}. \quad (\text{A2})$$

The scalar product on the algebra is defined as

$$\langle A, B \rangle = -2\text{tr}[AB] = A^a B^a, \quad (\text{A3})$$

where summation over the group indices  $a$  is implied. Note that the definition of the scalar product affects the trajectory length in the HMC and the MetaD-HMC. The convention we use throughout this paper agrees with, e.g., [18].

We also introduce the following abbreviation for the projector that maps any  $3 \times 3$  matrix onto  $\mathfrak{su}(3)$ :

$$\begin{aligned} \mathcal{P}_{\mathfrak{su}(3)}(A) &= \frac{1}{2}(A - A^\dagger) - \frac{1}{6}\text{tr}[A - A^\dagger] \\ &= -2T^a \text{Re tr}[T^a A]. \end{aligned} \quad (\text{A4})$$

The derivative of a scalar function of the gauge field  $f(U)$  with respect to a single link  $U_\alpha(n)$  is defined as a linear combination of generators  $T^a$  and partial derivative operators  $\partial_{n,\alpha}^a$ :

$$\partial_{n,\alpha} f(U) = T^a \partial_{n,\alpha}^a f(U). \quad (\text{A5})$$

Individually, the action of the  $a$ th partial derivative operator on  $f(U)$  is defined as

$$\partial_{n,\alpha}^a f(U) = \left. \frac{d}{ds} f(U^{[s]}) \right|_{s=0}, \quad (\text{A6})$$

where

$$U_\nu^{[s]}(m) = \begin{cases} e^{sT^a} U_\alpha(n) & \text{if } (m, \nu) = (n, \alpha), \\ U_\nu(m) & \text{else.} \end{cases} \quad (\text{A7})$$

While the result of the partial derivatives acting on a function is basis dependent, the linear combination in Eq. (A5) is not.

**APPENDIX B: CLOVER CHARGE DERIVATIVE**

In order to obtain an expression for the force contribution from the topological bias potential in Eq. (19), the algebra-valued derivative of the collective variable  $Q_{\text{meta}}$  with respect to the group-valued fully smeared links has to be calculated first. Evidently, this part of the force calculation does not depend on the smearing procedure used. The subsequent stout force recursion [19,85] required to relate this term to the derivative with respect to unsmeared links is discussed in Appendix C.

Recall that the clover-based definition of the topological charge is given by

$$Q_c = \frac{1}{32\pi^2} \sum_{n \in \Lambda} \epsilon_{\mu\nu\rho\sigma} \text{tr}[\hat{F}_{\mu\nu}^{\text{clov}}(n) \hat{F}_{\rho\sigma}^{\text{clov}}(n)], \quad (\text{B1})$$

where the field strength tensor is defined as

$$\hat{F}_{\mu\nu}^{\text{clov}}(n) = -\frac{i}{8} (C_{\mu\nu}(n) - C_{\nu\mu}(n)), \quad (\text{B2})$$

and the clover term is given by

$$C_{\mu\nu}(n) = P_{\mu,\nu}(n) + P_{\nu,-\mu}(n) + P_{-\mu,-\nu}(n) + P_{-\nu,\mu}(n). \quad (\text{B3})$$

For notational convenience, we introduce the auxiliary variables  $R_{\mu\nu}(n) = C_{\mu\nu}(n) - C_{\nu\mu}(n)$  and drop the specification of the lattice site  $n$  unless pertinent to the formula.

What we need for the force is the sum of the partial derivatives in all eight group directions. By using the cyclicity of the trace and the symmetry relation

$$R_{\mu\nu}^\dagger = -R_{\mu\nu} = R_{\nu\mu}, \quad (\text{B4})$$

we can rewrite the sum of partial derivatives as

$$T^a \partial_{n,\alpha}^a \epsilon_{\mu\nu\rho\sigma} \text{tr}[R_{\mu\nu} R_{\rho\sigma}] = 4T^a \sum_{\nu\rho\sigma} \partial_{n,\alpha}^a \epsilon_{\alpha\nu\rho\sigma} \text{tr}[R_{\alpha\nu} R_{\rho\sigma}]. \quad (\text{B5})$$

Note that in the second line of Eq. (B5) and the subsequent equations in this section, the index  $\alpha$  is not summed over.

Applying the partial derivative operators, we get two nonvanishing contributions from every plaquette term. Using the same symmetry arguments as before, we can further simplify the expression to obtain the following result for the derivative:

$$\begin{aligned}
4T^a \sum_{\nu\rho\sigma} \partial_{n,\alpha}^a \epsilon_{\alpha\nu\rho\sigma} \text{tr}[R_{\alpha\nu} R_{\rho\sigma}] &= 8T^a \sum_{\nu\rho\sigma} \epsilon_{\alpha\nu\rho\sigma} \text{Re tr} [T^a U_\alpha(n) U_\nu(n+\alpha) U_\alpha^\dagger(n+\nu) U_\nu^\dagger(n) R_{\rho\sigma}(n) \\
&\quad - T^a U_\alpha(n) U_\nu^\dagger(n+\alpha-\nu) U_\alpha^\dagger(n-\nu) R_{\rho\sigma}(n-\nu) U_\nu(n-\nu) \\
&\quad - T^a U_\alpha(n) U_\nu^\dagger(n+\alpha-\nu) R_{\rho\sigma}(n+\alpha-\nu) U_\alpha^\dagger(n-\nu) U_\nu(n-\nu) \\
&\quad + T^a U_\alpha(n) R_{\rho\sigma}(n+\alpha) U_\nu(n+\alpha) U_\alpha^\dagger(n+\nu) U_\nu^\dagger(n) \\
&\quad - T^a U_\alpha(n) U_\nu^\dagger(n+\alpha-\nu) U_\alpha^\dagger(n-\nu) U_\nu(n-\nu) R_{\rho\sigma}(n) \\
&\quad + T^a U_\alpha(n) U_\nu(n+\alpha) U_\alpha^\dagger(n+\nu) R_{\rho\sigma}(n+\nu) U_\nu^\dagger(n) \\
&\quad - T^a U_\alpha(n) R_{\rho\sigma}(n+\alpha) U_\nu^\dagger(n+\alpha-\nu) U_\alpha^\dagger(n-\nu) U_\nu(n-\nu) \\
&\quad + T^a U_\alpha(n) U_\nu(n+\alpha) R_{\rho\sigma}(n+\alpha+\nu) U_\alpha^\dagger(n+\nu) U_\nu^\dagger(n)] \\
&= 8T^a \sum_{\nu\rho\sigma} \epsilon_{\alpha\nu\rho\sigma} \text{Re tr}[T^a A_{\alpha\nu\rho\sigma}] \\
&= 8T^a \text{Re tr}[T^a A_\alpha]. \tag{B6}
\end{aligned}$$

An expression of the above form can be rewritten using the projector induced by the scalar product of the algebra defined in Eq. (A4):

$$8T^a \text{Re tr}[T^a A_\alpha] = -4\mathcal{P}_{\mathfrak{su}(3)}(A_\alpha). \tag{B7}$$

Including the missing prefactors from Eq. (B1) and Eq. (B2), we obtain the following final result for the derivative of the clover-based topological charge with respect to the gauge link  $U_\alpha(n)$ :

$$\begin{aligned}
T^a \partial_{n,\alpha}^a Q_c &= \frac{1}{32\pi^2} T^a \partial_{n,\alpha}^a \epsilon_{\mu\nu\rho\sigma} \text{tr}[\hat{F}_{\mu\nu}^{\text{clov}} \hat{F}_{\rho\sigma}^{\text{clov}}] \\
&= -\frac{1}{2048\pi^2} T^a \partial_{n,\alpha}^a \epsilon_{\mu\nu\rho\sigma} \text{tr}[R_{\mu\nu} R_{\rho\sigma}] \\
&= \frac{1}{512\pi^2} \mathcal{P}_{\mathfrak{su}(3)}(A_\alpha). \tag{B8}
\end{aligned}$$

In a similar way, expressions for the derivatives of improved definitions of the topological charge involving larger loops can be obtained.

### APPENDIX C: STOUT FORCE RECURSION

The derivative of the topological charge with respect to an unsmeared link can be computed by propagating the initial derivative with respect to a fully smeared link through the different smearing levels [19,85]. This part of the force calculation is completely independent of the topological charge operator and only depends on the smearing procedure used. In fact, the recursion is exactly the same as the recursion used during the force calculation with stout smeared fermions. Note that starting the calculation at the highest smearing level is vastly more efficient,

despite having to track intermediate smearing levels, since the function's domain is high dimensional, whereas the codomain is one dimensional.

We begin by establishing the required definitions before describing the actual force recursion. For notational convenience, we only consider a single smearing step without loss of generality and denote smeared quantities with a prime symbol. For the most part, we will use the same notation as in [85]. Starting from a gauge field  $U$ , the stout smeared gauge field  $U'$  is defined as

$$U'_\mu(n) = e^{iQ_\mu(n)} U_\mu(n), \tag{C1}$$

$$Q_\mu(n) = -i\mathcal{P}_{\mathfrak{su}(3)}(\Omega_\mu(n)), \tag{C2}$$

$$\Omega_\mu(n) = C_\mu(n) U_\mu^\dagger(n). \tag{C3}$$

Note that no summation is performed over  $\mu$  in Eq. (C3).  $C_\mu(n)$  is a weighted sum of staples:

$$\begin{aligned}
C_\mu(n) &= \sum_{\nu \neq \mu} \rho_{\mu\nu} (U_\nu(n) U_\mu(n+\hat{\nu}) U_\nu^\dagger(n+\hat{\mu}) \\
&\quad + U_\nu^\dagger(n-\hat{\nu}) U_\mu(n-\hat{\nu}) U_\nu(n+\hat{\mu}-\hat{\nu})). \tag{C4}
\end{aligned}$$

If the smeared force  $F'$  is known, the unsmeared force  $F$  is given by

$$\begin{aligned}
F_\mu(n) &= \mathcal{P}_{\mathfrak{su}(3)}(U_\mu(n) \Sigma'_\mu(n) e^{iQ_\mu(n)} \\
&\quad + iU_\mu(n) C_\mu^\dagger(n) \Lambda_\mu(n) - iU_\mu(n) \Upsilon_\mu(n)), \tag{C5}
\end{aligned}$$

where

$$\begin{aligned} \Upsilon_\mu(n) = & \sum_{\nu \neq \mu} (\rho_{\nu\mu} U_\nu(n + \hat{\mu}) U_\mu^\dagger(n + \hat{\nu}) U_\nu^\dagger(n) \Lambda_\nu(n) + \rho_{\mu\nu} U_\nu^\dagger(n + \hat{\mu} - \hat{\nu}) U_\mu^\dagger(n - \hat{\nu}) \Lambda_\mu(n - \hat{\nu}) U_\nu(n - \hat{\nu}) \\ & + \rho_{\nu\mu} U_\nu^\dagger(n + \hat{\mu} - \hat{\nu}) \Lambda_\nu(n + \hat{\mu} - \hat{\nu}) U_\mu^\dagger(n - \hat{\nu}) U_\nu(n - \hat{\nu}) - \rho_{\nu\mu} U_\nu^\dagger(n + \hat{\mu} - \hat{\nu}) U_\mu^\dagger(n - \hat{\nu}) \Lambda_\nu(n - \hat{\nu}) U_\nu(n - \hat{\nu}) \\ & - \rho_{\nu\mu} \Lambda_\nu(n + \hat{\mu}) U_\nu(n + \hat{\mu}) U_\mu^\dagger(n + \hat{\nu}) U_\nu^\dagger(n) + \rho_{\mu\nu} U_\nu(n + \hat{\mu}) U_\mu^\dagger(n + \hat{\nu}) \Lambda_\mu(n + \hat{\nu}) U_\nu^\dagger(n)), \end{aligned} \quad (\text{C6})$$

$$\Lambda = \frac{1}{2}(\Gamma + \Gamma^\dagger) - \frac{1}{6}\text{tr}[\Gamma + \Gamma^\dagger], \quad (\text{C7})$$

$$\begin{aligned} \Gamma = & \text{tr}[B_1 U \Sigma'] Q + \text{tr}[B_2 U \Sigma'] Q^2 \\ & + f_1 U \Sigma' + f_2 Q U \Sigma' + f_2 U \Sigma' Q, \end{aligned} \quad (\text{C8})$$

$$\Sigma' = U'^\dagger F. \quad (\text{C9})$$

The exact form of the  $f_j$  and  $B_i$  is given in Appendix D. For multiple smearing levels, the procedure in Eq. (C5) can simply be repeated.

#### APPENDIX D: COMPUTATION OF THE MATRIX EXPONENTIAL USING THE CAYLEY-HAMILTON THEOREM

The computation of the matrix exponential is done along the lines of [85] using the Cayley-Hamilton theorem, which also affects the details of the stout force recursion. The characteristic polynomial of a  $3 \times 3$  matrix  $A$  reads as follows:

$$p_A(\lambda) = \lambda^3 - \text{tr}[A]\lambda^2 + \frac{1}{2}(\text{tr}[A]^2 - \text{tr}[A^2])\lambda - \det(A). \quad (\text{D1})$$

For a traceless Hermitian matrix  $Q$ , this reduces to

$$p_Q(\lambda) = \lambda^3 - \frac{1}{2}\text{tr}[Q^2]\lambda - \det(Q). \quad (\text{D2})$$

According to the Cayley-Hamilton theorem,  $Q$  satisfies its own characteristic equation:

$$Q^3 - c_1 Q - c_0 I = 0, \quad (\text{D3})$$

where we again follow the notation in [85]:

$$c_0 = \det(Q) = \frac{1}{3}\text{tr}[Q^3], \quad (\text{D4})$$

$$c_1 = \frac{1}{2}\text{tr}[Q^2] \geq 0. \quad (\text{D5})$$

The structure of  $Q$  restricts the range of  $c_0$ :

$$c_0^{\max} = -c_0^{\min} = 2\left(\frac{c_1}{3}\right)^{3/2}. \quad (\text{D6})$$

For analytic functions, Eq. (D3) may therefore be used to reduce the series representation of the function to a polynomial of degree 2 in the matrix itself:

$$f(Q) = f_2 Q^2 + f_1 Q + f_0. \quad (\text{D7})$$

More generally, analytic functions of  $n \times n$  matrices can be reduced to polynomials of degree  $n - 1$  using the same strategy. The complex coefficients  $f_j$  may be expressed in terms of complex auxiliary functions  $h_j$  and real variables  $u$  and  $w$  that are closely related to the eigenvalues of  $Q$ :

$$f_j = \frac{h_j}{9u^2 - w^2}, \quad (\text{D8})$$

where

$$u = \sqrt{\frac{c_1}{3}} \cos\left(\frac{\theta}{3}\right), \quad (\text{D9})$$

$$w = \sqrt{c_1} \sin\left(\frac{\theta}{3}\right), \quad (\text{D10})$$

$$\theta = \arccos\left(\frac{c_0}{c_0^{\max}}\right). \quad (\text{D11})$$

The  $h_j$  in turn are defined as

$$\begin{aligned} h_0 = & (u^2 - w^2)e^{2iu} + e^{-iu}(8u^2 \cos(w) \\ & + 2iu(3u^2 + w^2)\xi_0(w)), \end{aligned} \quad (\text{D12})$$

$$\begin{aligned} h_1 = & 2ue^{2iu} - e^{-iu}(2u \cos(w) \\ & - i(3u^2 - w^2)\xi_0(w)), \end{aligned} \quad (\text{D13})$$

$$h_2 = e^{2iu} - e^{-iu}(\cos(w) + 3iu\xi_0(w)), \quad (\text{D14})$$

with

$$\xi_0(w) = \text{sinc}(w) = \frac{\sin(w)}{w}, \quad (\text{D15})$$

$$\xi_1(w) = \frac{\text{sinc}'(w)}{w} = \frac{\cos(w)}{w^2} - \frac{\sin(w)}{w^3}. \quad (\text{D16})$$

For small values of  $w$ , the direct evaluation of the expressions in Eq. (D15) and Eq. (D16) may suffer from

numerical inaccuracies. To ensure sufficient precision, we use sixth-order Taylor expansions below certain thresholds:

$$\xi_0(w) = \begin{cases} 1 - \frac{w^2}{6} \left(1 - \frac{w^2}{20} \left(1 - \frac{w^2}{42}\right)\right) & |w| \leq \xi_{0,\text{thr}}, \\ \frac{\sin(w)}{w} & |w| > \xi_{0,\text{thr}}. \end{cases} \quad (\text{D17})$$

$$\xi_1(w) = \begin{cases} -\frac{1}{3} + \frac{w^2}{30} \left(1 - \frac{w^2}{28} \left(1 - \frac{w^2}{54}\right)\right) & |w| \leq \xi_{1,\text{thr}}, \\ \frac{\cos(w)}{w^2} - \frac{\sin(w)}{w^3} & |w| > \xi_{1,\text{thr}}. \end{cases} \quad (\text{D18})$$

The thresholds used in our simulations were determined empirically:

$$\xi_{0,\text{thr}} = \begin{cases} 0.56 & \text{in single precision,} \\ 0.05 & \text{in double precision.} \end{cases} \quad (\text{D19})$$

$$\xi_{1,\text{thr}} = \begin{cases} 0.75 & \text{in single precision,} \\ 0.115 & \text{in double precision.} \end{cases} \quad (\text{D20})$$

The stout force recursion also involves terms related to the derivative of the exponential function that appear in Eq. (C8). Specifically, the matrix polynomials  $B_i$  are defined as

$$B_i = b_{i2}Q^2 + b_{i1}Q + b_{i0}. \quad (\text{D21})$$

The coefficients  $b_{ij}$  are given by

$$b_{1j} = \frac{2ur_j^{(1)} + (3u^2 - w^2)r_j^{(2)} - 2(15u^2 + w^2)f_j}{2(9u^2 - w^2)^2}, \quad (\text{D22})$$

$$b_{2j} = \frac{r_j^{(1)} - 3ur_j^{(2)} - 24uf_j}{2(9u^2 - w^2)^2}, \quad (\text{D23})$$

and the auxiliary quantities  $r_j^{(i)}$  are defined as

$$r_0^{(1)} = 2(u + i(u^2 - w^2))e^{2iu} + 2e^{-iu}(4u(2 - iu)\cos(w) + i(9u^2 + w^2 - iu(3u^2 + w^2))\xi_0(w)), \quad (\text{D24})$$

$$r_1^{(1)} = 2(1 + 2iu)e^{2iu} + e^{-iu}(-2(1 - iu)\cos(w) + i(6u + i(w^2 - 3u^2))\xi_0(w)), \quad (\text{D25})$$

$$r_2^{(1)} = 2ie^{2iu} + ie^{-iu}(\cos(w) - 3(1 - iu)\xi_0(w)), \quad (\text{D26})$$

$$r_0^{(2)} = -2e^{2iu} + 2iue^{-iu}(\cos(w) + (1 + 4iu)\xi_0(w) + 3u^2\xi_1(w)), \quad (\text{D27})$$

$$r_1^{(2)} = -ie^{-iu}(\cos(w) + (1 + 2iu)\xi_0(w) - 3u^2\xi_1(w)), \quad (\text{D28})$$

$$r_2^{(2)} = e^{-iu}(\xi_0(w) - 3iu\xi_1(w)). \quad (\text{D29})$$

In addition to the previously discussed functions  $\xi_0(w)$  and  $\xi_1(w)$ , numerical instabilities may also be caused by a small denominator appearing in Eqs. (D8), (D22), and (D23). In particular, the expressions become problematic when  $w \rightarrow 3u$ , as either  $c_0 \rightarrow c_0^{\min}$  or  $c_1 \rightarrow 0$ . The former case can be circumvented by using the following symmetry relations of  $f_j$  and  $b_{ij}$  with respect to the sign of  $c_0$ :

$$f_j(-c_0, c_1) = (-1)^j f_j^*(c_0, c_1), \quad (\text{D30})$$

$$b_{ij}(-c_0, c_1) = (-1)^{i+j+1} b_{ij}^*(c_0, c_1). \quad (\text{D31})$$

By only working with positive  $c_0$ , the range of  $\theta$  is restricted to  $\theta \in [0, \pi/2]$ , which in turn implies that the expression  $9u^2 - w^2$  will lie in the interval  $[2, 3]$ .

The latter case  $c_1 \rightarrow 0$  implies that  $c_0^{\max} \rightarrow 0$ , which can cause problems during the division in the argument of the arccos in Eq. (D11). In practice, this can be handled by explicitly setting

$$u = w = 0, \quad (\text{D32})$$

$$f_j = h_j = \begin{cases} 1 & j = 0, \\ 0 & \text{else} \end{cases} \quad (\text{D33})$$

for very small values of  $c_0^{\max}$  (namely, values smaller than the smallest representable positive normalized floating point number). In particular, this also covers the case of the trivial gauge configuration where  $Q = 0$  and all three eigenvalues of the matrix will be degenerate and exactly equal to 0.

Finally, problems appear if the argument of the arccos in Eq. (D11) exceeds 1 due to numerical inaccuracies. Since the argument is guaranteed to be bounded by 1 in exact arithmetic, this problem can be solved by rounding the argument down to 1.

## APPENDIX E: EXACT SOLUTION OF TWO-DIMENSIONAL U(1) GAUGE THEORY

In two-dimensional U(1) gauge theory, exact results for expectation values of many observables are known analytically [40,106–108]. A detailed derivation for Wilson loops in the aforementioned theory with both open and periodic boundary conditions can be found in Appendix A.3 of [110], of which we summarize the main results. For convenience, we introduce the notation

$$R_{n,m}(\beta) = \frac{I_n(\beta)}{I_m(\beta)}, \quad (\text{E1})$$

where  $I_n(\beta)$  is the  $n$ th modified Bessel function of the first kind.

For open boundary conditions in two dimensions, all plaquettes decouple and can be integrated independently. In particular, a consequence of this is the complete absence of finite volume effects, independent of the gauge group. Expectation values of Wilson loops with area  $A$  (the exact shape does not matter due to the Abelian nature of the gauge group) are given by

$$\langle \mathcal{W}_A \rangle_{\text{open}} = \left( \frac{I_1(\beta)}{I_0(\beta)} \right)^A = (R_{1,0}(\beta))^A. \quad (\text{E2})$$

Evidently, the plaquette expectation value is simply given by

$$\langle P \rangle_{\text{open}} = \frac{I_1(\beta)}{I_0(\beta)} = R_{1,0}(\beta). \quad (\text{E3})$$

When periodic boundary conditions are imposed, the theory is not completely trivial, as additional constraints are imposed by the periodicity. In that case, the expectation values of Wilson loops of area  $A$  on a square lattice of size  $L \times L$  can be expressed in terms of sums over irreducible characters of the gauge group:

$$\langle \mathcal{W}_A \rangle_{\text{periodic}} = \frac{\sum_r (I_r(\beta))^{L^2-A} (I_{r+1}(\beta))^A}{\sum_r (I_r(\beta))^{L^2}}, \quad (\text{E4})$$

where the index  $r \in \mathbb{Z}$  runs over the irreducible representations of  $U(1)$ . The direct evaluation of this expression is problematic for fine lattice spacings and larger volumes, since the power of the Bessel functions quickly exceeds the representable range for floating point numbers. This problem can be circumvented by slightly rewriting the expression and only working with ratios of Bessel functions:

$$\langle \mathcal{W}_A \rangle_{\text{periodic}} = \frac{\sum_r (R_{r,0}(\beta))^{L^2} (R_{r+1,r}(\beta))^A}{\sum_r (R_{r,0}(\beta))^{L^2}}. \quad (\text{E5})$$

For the plaquette, this simplifies to

$$\langle P \rangle_{\text{periodic}} = \frac{\sum_r (R_{r,0}(\beta))^{L^2} R_{r+1,r}(\beta)}{\sum_r (R_{r,0}(\beta))^{L^2}}. \quad (\text{E6})$$

Finally, the (unnormalized) topological susceptibility is given by [40]

$$\begin{aligned} \langle Q^2 \rangle_{\text{periodic}} &= -V \frac{\sum_r A_r(\beta) (I_r(\beta))^{V-1}}{\sum_r (I_r(\beta))^V} \\ &\quad - (V^2 - V) \frac{\sum_r (B_r(\beta))^2 (I_r(\beta))^{V-2}}{\sum_r (I_r(\beta))^V}, \end{aligned} \quad (\text{E7})$$

where

$$A_r(\beta) = -\frac{1}{2\pi} \int_{-\pi}^{\pi} d\phi \left( \frac{\phi}{2\pi} \right)^2 e^{ir\phi + \beta \cos \phi}, \quad (\text{E8})$$

$$B_r(\beta) = \frac{i}{2\pi} \int_{-\pi}^{\pi} d\phi \frac{\phi}{2\pi} e^{ir\phi + \beta \cos \phi}. \quad (\text{E9})$$

Again, it is advantageous from a numerical standpoint to work with ratios of Bessel functions:

$$\begin{aligned} \langle Q^2 \rangle_{\text{periodic}} &= -V \frac{\sum_r A_r(\beta) (R_{r,0}(\beta))^{V-1}}{I_0(\beta) \sum_r (R_{r,0}(\beta))^V} \\ &\quad - (V^2 - V) \frac{\sum_r (B_r(\beta))^2 (R_{r,0}(\beta))^{V-2}}{(I_0(\beta))^2 \sum_r (R_{r,0}(\beta))^V}. \end{aligned} \quad (\text{E10})$$

All of the sums in the previous expressions converge rapidly; in fact, for our parameters ( $V = 32^2$ ,  $\beta = 12.8$ ), only considering the leading terms already gives the correct results up to machine precision, which are indistinguishable from the results obtained for open boundary conditions for the Wilson loops.

- 
- [1] Y. Aoki *et al.* (Flavour Lattice Averaging Group (FLAG) Collaboration), FLAG review 2021, *Eur. Phys. J. C* **82**, 869 (2022).
  - [2] M. Creutz, Confinement and the critical dimensionality of space-time, *Phys. Rev. Lett.* **43**, 553 (1979).
  - [3] M. Creutz, Monte Carlo study of quantized SU(2) gauge theory, *Phys. Rev. D* **21**, 2308 (1980).
  - [4] K. Fabricius and O. Haan, Heat bath method for the twisted Eguchi-Kawai model, *Phys. Lett.* **143B**, 459 (1984).
  - [5] A. D. Kennedy and B. J. Pendleton, Improved heat bath method for Monte Carlo calculations in lattice gauge theories, *Phys. Lett.* **156B**, 393 (1985).
  - [6] N. Cabibbo and E. Marinari, A new method for updating SU(N) matrices in computer simulations of gauge theories, *Phys. Lett.* **119B**, 387 (1982).
  - [7] S. L. Adler, An overrelaxation method for the Monte Carlo evaluation of the partition function for multiquadratic actions, *Phys. Rev. D* **23**, 2901 (1981).
  - [8] M. Creutz, Overrelaxation and Monte Carlo simulation, *Phys. Rev. D* **36**, 515 (1987).
  - [9] F. R. Brown and T. J. Woch, Overrelaxed heat bath and metropolis algorithms for accelerating pure gauge Monte Carlo calculations, *Phys. Rev. Lett.* **58**, 2394 (1987).

- [10] S. Duane, A. D. Kennedy, B. J. Pendleton, and D. Roweth, Hybrid Monte Carlo, *Phys. Lett. B* **195**, 216 (1987).
- [11] 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**, 107 (1996).
- [12] K. Orginos (RBC Collaboration), Chiral properties of domain wall fermions with improved gauge actions, *Nucl. Phys. B, Proc. Suppl.* **106**, 721 (2002).
- [13] L. Del Debbio, H. Panagopoulos, and E. Vicari, theta dependence of SU(N) gauge theories, *J. High Energy Phys.* **08** (2002) 044.
- [14] T. A. DeGrand, A. Hasenfratz, and T. G. Kovacs, Improving the chiral properties of lattice fermions, *Phys. Rev. D* **67**, 054501 (2003).
- [15] J. Noaki (RBC Collaboration), Calculation of weak matrix elements in domain wall QCD with the DBW2 gauge action, *Nucl. Phys. B, Proc. Suppl.* **119**, 362 (2003).
- [16] L. Del Debbio, G. M. Manca, and E. Vicari, Critical slowing down of topological modes, *Phys. Lett. B* **594**, 315 (2004).
- [17] Y. Aoki *et al.*, The Kaon B-parameter from quenched domain-wall QCD, *Phys. Rev. D* **73**, 094507 (2006).
- [18] S. Schaefer, R. Sommer, and F. Viotto (ALPHA Collaboration), Critical slowing down and error analysis in lattice QCD simulations, *Nucl. Phys.* **B845**, 93 (2011).
- [19] S. Durr, Z. Fodor, C. Hoelbling, S. D. Katz, S. Krieg, T. Kurth, L. Lellouch, T. Lippert, K. K. Szabo, and G. Vulvert (BMW Collaboration), Lattice QCD at the physical point: Simulation and analysis details, *J. High Energy Phys.* **08** (2011) 148.
- [20] C. Bernard and D. Toussaint (MILC Collaboration), Effects of nonequilibrated topological charge distributions on pseudoscalar meson masses and decay constants, *Phys. Rev. D* **97**, 074502 (2018).
- [21] S. Borsanyi *et al.*, Leading hadronic contribution to the muon magnetic moment from lattice QCD, *Nature (London)* **593**, 51 (2021).
- [22] R. Brower, S. Chandrasekharan, J. W. Negele, and U. J. Wiese, QCD at fixed topology, *Phys. Lett. B* **560**, 64 (2003).
- [23] S. Aoki, H. Fukaya, S. Hashimoto, and T. Onogi, Finite volume QCD at fixed topological charge, *Phys. Rev. D* **76**, 054508 (2007).
- [24] M. Lüscher, Stochastic locality and master-field simulations of very large lattices, *EPJ Web Conf.* **175**, 01002 (2018).
- [25] M. Luscher and S. Schaefer, Lattice QCD without topology barriers, *J. High Energy Phys.* **07** (2011) 036.
- [26] B. Joo, B. Pendleton, S. M. Pickles, Z. Sroczynski, A. C. Irving, and J. C. Sexton (UKQCD Collaboration), Parallel tempering in lattice QCD with O(a)-improved Wilson fermions, *Phys. Rev. D* **59**, 114501 (1999).
- [27] E.-M. Ilgenfritz, W. Kerler, M. Muller-Preussker, and H. Stuben, Parallel tempering in full QCD with Wilson fermions, *Phys. Rev. D* **65**, 094506 (2002).
- [28] S. Borsanyi, R. Kara, Z. Fodor, D. A. Godzieba, P. Parotto, and D. Sexty, Precision study of the continuum SU(3) Yang-Mills theory: How to use parallel tempering to improve on supercritical slowing down for first order phase transitions, *Phys. Rev. D* **105**, 074513 (2022).
- [29] S. Mages, B. C. Toth, S. Borsanyi, Z. Fodor, S. D. Katz, and K. K. Szabo, Lattice QCD on nonorientable manifolds, *Phys. Rev. D* **95**, 094512 (2017).
- [30] M. Hasenbusch, Fighting topological freezing in the two-dimensional CPN-1 model, *Phys. Rev. D* **96**, 054504 (2017).
- [31] C. Bonanno, C. Bonati, and M. D’Elia, Large- $N$  SU( $N$ ) Yang-Mills theories with milder topological freezing, *J. High Energy Phys.* **03** (2021) 111.
- [32] M. G. Endres, R. C. Brower, W. Detmold, K. Orginos, and A. V. Pochinsky, Multiscale Monte Carlo equilibration: Pure Yang-Mills theory, *Phys. Rev. D* **92**, 114516 (2015).
- [33] W. Detmold and M. G. Endres, Multiscale Monte Carlo equilibration: Two-color QCD with two fermion flavors, *Phys. Rev. D* **94**, 114502 (2016).
- [34] W. Detmold and M. G. Endres, Scaling properties of multiscale equilibration, *Phys. Rev. D* **97**, 074507 (2018).
- [35] F. Fucito and S. Solomon, Does standard Monte Carlo give justice to instantons?, *Phys. Lett.* **134B**, 230 (1984).
- [36] J. Smit and J. C. Vink, Topological charge and fermions in the two-dimensional lattice U(1) model. 1. Staggered fermions, *Nucl. Phys.* **B303**, 36 (1988).
- [37] H. Dilger, Screening in the lattice Schwinger model, *Phys. Lett. B* **294**, 263 (1992).
- [38] H. Dilger, Topological zero modes in Monte Carlo simulations, *Int. J. Mod. Phys. C* **06**, 123 (1995).
- [39] S. Durr, Physics of  $\eta'$  with rooted staggered quarks, *Phys. Rev. D* **85**, 114503 (2012).
- [40] D. Albandea, P. Hernández, A. Ramos, and F. Romero-López, Topological sampling through windings, *Eur. Phys. J. C* **81**, 873 (2021).
- [41] A. Laio, G. Martinelli, and F. Sanfilippo, Metadynamics surfing on topology barriers: The  $CP^{N-1}$  case, *J. High Energy Phys.* **07** (2016) 089.
- [42] C. Bonati, The topological properties of QCD at high temperature: problems and perspectives, *EPJ Web Conf.* **175**, 01011 (2018).
- [43] B. A. Berg and T. Neuhaus, Multicanonical ensemble: A new approach to simulate first-order phase transitions, *Phys. Rev. Lett.* **68**, 9 (1992).
- [44] C. Bonati, M. D’Elia, G. Martinelli, F. Negro, F. Sanfilippo, and A. Todaro, Topology in full QCD at high temperature: A multicanonical approach, *J. High Energy Phys.* **11** (2018) 170.
- [45] G. Cossu, D. Lancastera, B. Lucini, R. Pellegrini, and A. Rago, Ergodic sampling of the topological charge using the density of states, *Eur. Phys. J. C* **81**, 375 (2021).
- [46] S. Duane, R. Kenway, B. J. Pendleton, and D. Roweth, Acceleration of gauge field dynamics, *Phys. Lett. B* **176**, 143 (1986).
- [47] S. Duane and B. J. Pendleton, Gauge invariant fourier acceleration, *Phys. Lett. B* **206**, 101 (1988).
- [48] C. T. H. Davies, G. G. Batrouni, G. R. Katz, A. S. Kronfeld, G. P. Lepage, P. Rossi, B. Svetitsky, and K. G. Wilson, Fourier acceleration in lattice gauge theories. 3. Updating field configurations, *Phys. Rev. D* **41**, 1953 (1990).
- [49] G. Cossu, P. Boyle, N. Christ, C. Jung, A. Jüttner, and F. Sanfilippo, Testing algorithms for critical slowing down, *EPJ Web Conf.* **175**, 02008 (2018).

- [50] T. Nguyen, P. Boyle, N. H. Christ, Y.-C. Jang, and C. Jung, Riemannian manifold hybrid Monte Carlo in lattice QCD, *Proc. Sci. LATTICE2021* (2022) 582.
- [51] M. Luscher, Trivializing maps, the Wilson flow and the HMC algorithm, *Commun. Math. Phys.* **293**, 899 (2010).
- [52] G. P. Engel and S. Schaefer, Testing trivializing maps in the Hybrid Monte Carlo algorithm, *Comput. Phys. Commun.* **182**, 2107 (2011).
- [53] P. Boyle, T. Izubuchi, L. Jin, C. Jung, C. Lehner, N. Matsumoto, and A. Tomiya, Use of Schwinger-Dyson equation in constructing an approximate trivializing map, *Proc. Sci. LATTICE2022* (2023) 229.
- [54] 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**, 034515 (2019).
- [55] G. Kanwar, M. S. Albergo, D. Boyda, K. Cranmer, D. C. Hackett, S. Racanière, D. J. Rezende, and P. E. Shanahan, Equivariant flow-based sampling for lattice gauge theory, *Phys. Rev. Lett.* **125**, 121601 (2020).
- [56] K. A. Nicoli, C. J. Anders, L. Funcke, T. Hartung, K. Jansen, P. Kessel, S. Nakajima, and P. Stornati, Estimation of thermodynamic observables in lattice field theories with deep generative models, *Phys. Rev. Lett.* **126**, 032001 (2021).
- [57] D. Boyda, G. Kanwar, S. Racanière, D. J. Rezende, M. S. Albergo, K. Cranmer, D. C. Hackett, and P. E. Shanahan, Sampling using  $SU(N)$  gauge equivariant flows, *Phys. Rev. D* **103**, 074504 (2021).
- [58] L. Del Debbio, J. Marsh Rossney, and M. Wilson, Efficient modeling of trivializing maps for lattice  $\phi^4$  theory using normalizing flows: A first look at scalability, *Phys. Rev. D* **104**, 094507 (2021).
- [59] M. S. Albergo, G. Kanwar, S. Racanière, D. J. Rezende, J. M. Urban, D. Boyda, K. Cranmer, D. C. Hackett, and P. E. Shanahan, Flow-based sampling for fermionic lattice field theories, *Phys. Rev. D* **104**, 114507 (2021).
- [60] D. C. Hackett, C.-C. Hsieh, M. S. Albergo, D. Boyda, J.-W. Chen, K.-F. Chen, K. Cranmer, G. Kanwar, and P. E. Shanahan, Flow-based sampling for multimodal distributions in lattice field theory, [arXiv:2107.00734](https://arxiv.org/abs/2107.00734).
- [61] K. A. Nicoli, C. J. Anders, L. Funcke, T. Hartung, K. Jansen, P. Kessel, S. Nakajima, and P. Stornati, Machine learning of thermodynamic observables in the presence of mode collapse, *Proc. Sci., LATTICE2021* (2022) 338.
- [62] S. Foreman, T. Izubuchi, L. Jin, X.-Y. Jin, J. C. Osborn, and A. Tomiya, HMC with normalizing flows, *Proc. Sci., LATTICE2021* (2022) 073.
- [63] J. Finkenrath, Tackling critical slowing down using global correction steps with equivariant flows: The case of the Schwinger model, [arXiv:2201.02216](https://arxiv.org/abs/2201.02216).
- [64] M. S. Albergo, D. Boyda, K. Cranmer, D. C. Hackett, G. Kanwar, S. Racanière, D. J. Rezende, F. Romero-López, P. E. Shanahan, and J. M. Urban, Flow-based sampling in the lattice Schwinger model at criticality, *Phys. Rev. D* **106**, 014514 (2022).
- [65] J. M. Pawłowski and J. M. Urban, Flow-based density of states for complex actions, *Phys. Rev. D* **108**, 054511 (2023).
- [66] 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**, 238 (2023).
- [67] A. Singha, D. Chakrabarti, and V. Arora, Conditional normalizing flow for Monte Carlo sampling in lattice scalar field theory, *Phys. Rev. D* **107**, 014512 (2023).
- [68] R. Abbott *et al.*, Gauge-equivariant flow models for sampling in lattice field theories with pseudofermions, *Phys. Rev. D* **106**, 074506 (2022).
- [69] R. Abbott *et al.*, Sampling QCD field configurations with gauge-equivariant flow models, *Proc. Sci. LATTICE2022* (2023) 036.
- [70] R. Abbott *et al.*, Aspects of scaling and scalability for flow-based sampling of lattice QCD, *Eur. Phys. J. A* **59**, 257 (2023).
- [71] S. Bacchio, P. Kessel, S. Schaefer, and L. Vaitl, Learning trivializing gradient flows for lattice gauge theories, *Phys. Rev. D* **107**, L051504 (2023).
- [72] J. Komijani and M. K. Marinkovic, Generative models for scalar field theories: How to deal with poor scaling?, *Proc. Sci., LATTICE2022* (2023) 019.
- [73] D. Albandea, L. Del Debbio, P. Hernández, R. Kenway, J. M. Rossney, and A. Ramos, Learning trivializing flows, *Eur. Phys. J. C* **83**, 676 (2023).
- [74] 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**, 114501 (2023).
- [75] R. Abbott *et al.*, Normalizing flows for lattice gauge theory in arbitrary space-time dimension, [arXiv:2305.02402](https://arxiv.org/abs/2305.02402).
- [76] J. Finkenrath, Review on algorithms for dynamical fermions, *Proc. Sci., LATTICE2022* (2023) 227.
- [77] R. H. Swendsen and J.-S. Wang, Replica Monte Carlo simulation of spin-glasses, *Phys. Rev. Lett.* **57**, 2607 (1986).
- [78] A. Laio and M. Parrinello, Escaping free-energy minima, *Proc. Natl. Acad. Sci. U.S.A.* **99**, 12562 (2002).
- [79] G. Bussi, A. Laio, and M. Parrinello, Equilibrium free energies from nonequilibrium metadynamics, *Phys. Rev. Lett.* **96**, 090601 (2006).
- [80] Y. Crespo, F. Marinelli, F. Pietrucci, and A. Laio, Metadynamics convergence law in a multidimensional system, *Phys. Rev. E* **81**, 055701(R) (2010).
- [81] T. M. Schäfer and G. Settanni, Data reweighting in metadynamics simulations, *J. Chem. Theory Comput.* **16**, 2042 (2020).
- [82] K. G. Wilson, Confinement of quarks, *Phys. Rev. D* **10**, 2445 (1974).
- [83] T. Takaishi, Heavy quark potential and effective actions on blocked configurations, *Phys. Rev. D* **54**, 1050 (1996).
- [84] M. Luscher and P. Weisz, On-shell improved lattice gauge theories, *Commun. Math. Phys.* **97**, 59 (1985); **98**, 433(E) (1985).
- [85] C. Morningstar and M. J. Peardon, Analytic smearing of  $SU(3)$  link variables in lattice QCD, *Phys. Rev. D* **69**, 054501 (2004).
- [86] I. Omelyan, I. Mryglod, and R. Folk, Symplectic analytically integrable decomposition algorithms: Classification, derivation, and application to molecular dynamics, quantum and celestial mechanics simulations, *Comput. Phys. Commun.* **151**, 272 (2003).

- [87] S. Capitani, S. Durr, and C. Hoelbling, Rationale for UV-filtered clover fermions, *J. High Energy Phys.* **11** (2006) 028.
- [88] T. Eichhorn, C. Hoelbling, P. Rouenhoff, and L. Varnhorst, Topology changing update algorithms for SU(3) gauge theory, *Proc. Sci., LATTICE2022* (2023) 009.
- [89] Y. Nagai and A. Tomiya, Gauge covariant neural network for 4 dimensional non-abelian gauge theory, [arXiv:2103.11965](https://arxiv.org/abs/2103.11965).
- [90] S. Durr, Z. Fodor, C. Hoelbling, and T. Kurth, Precision study of the SU(3) topological susceptibility in the continuum, *J. High Energy Phys.* **04** (2007) 055.
- [91] S. Necco and R. Sommer, The  $N(f) = 0$  heavy quark potential from short to intermediate distances, *Nucl. Phys.* **B622**, 328 (2002).
- [92] U. Wolff (ALPHA Collaboration), Monte Carlo errors with less errors, *Comput. Phys. Commun.* **156**, 143 (2004); **176**, 383(E) (2007).
- [93] P. Marenzoni, L. Pugnetti, and P. Rossi, Measure of autocorrelation times of local hybrid Monte Carlo algorithm for lattice QCD, *Phys. Lett. B* **315**, 152 (1993).
- [94] K. Hashimoto and T. Izubuchi (RBC Collaboration), Static anti-Q—Q potential from  $N(f) = 2$  dynamical domain-wall QCD, *Nucl. Phys. B, Proc. Suppl.* **140**, 341 (2005).
- [95] S. Necco, Universality and scaling behavior of RG gauge actions, *Nucl. Phys.* **B683**, 137 (2004).
- [96] M. Lüscher, Properties and uses of the Wilson flow in lattice QCD, *J. High Energy Phys.* **08** (2010) 071; **03** (2014) 092(E).
- [97] S. Borsanyi *et al.* (BMW Collaboration), High-precision scale setting in lattice QCD, *J. High Energy Phys.* **09** (2012) 010.
- [98] R. Sommer, Scale setting in lattice QCD, *Proc. Sci., LATTICE2013* (2014) 015.
- [99] J. C. Sexton and D. H. Weingarten, Hamiltonian evolution for the hybrid Monte Carlo algorithm, *Nucl. Phys.* **B380**, 665 (1992).
- [100] P. de Forcrand, J. E. Hetrick, T. Takaishi, and A. J. van der Sijs, Three topics in the Schwinger model, *Nucl. Phys. B, Proc. Suppl.* **63**, 679 (1998).
- [101] P. Rouenhoff, T. Eichhorn, C. Hoelbling, and L. Varnhorst, Metadynamics surfing on topology barriers in the Schwinger model, *Proc. Sci., LATTICE2022* (2022) 253.
- [102] A. Barducci, G. Bussi, and M. Parrinello, Well-tempered metadynamics: A smoothly converging and tunable free-energy method, *Phys. Rev. Lett.* **100**, 020603 (2008).
- [103] P. Raiteri, A. Laio, F. L. Gervasio, C. Micheletti, and M. Parrinello, Efficient reconstruction of complex free energy landscapes by multiple walkers metadynamics, *J. Phys. Chem. B* **110**, 3533 (2006).
- [104] J. Smit and J. C. Vink, Remnants of the index theorem on the lattice, *Nucl. Phys.* **B286**, 485 (1987).
- [105] P. T. Jahn, The Topological Susceptibility of QCD at High Temperatures, Ph.D. thesis, Darmstadt, Tech. Hochsch., 2019.
- [106] T. G. Kovacs, E. T. Tomboulis, and Z. Schram, Topology on the lattice: 2-d Yang-Mills theories with a theta term, *Nucl. Phys.* **B454**, 45 (1995).
- [107] S. Elser, The Local bosonic algorithm applied to the massive Schwinger model, Ph.D. thesis, Humboldt University of Berlin, 2001, [arXiv:hep-lat/0103035](https://arxiv.org/abs/hep-lat/0103035).
- [108] C. Bonati and P. Rossi, Topological susceptibility of two-dimensional  $U(N)$  gauge theories, *Phys. Rev. D* **99**, 054503 (2019).
- [109] R. Vautard and M. Ghil, Singular spectrum analysis in nonlinear dynamics, with applications to paleoclimatic time series, *Physica (Amsterdam)* **35D**, 395 (1989).
- [110] G. Kanwar, Machine learning and variational algorithms for lattice field theory, Ph.D. thesis, MIT, 2021, [arXiv:2106.01975](https://arxiv.org/abs/2106.01975).