Multicanonical reweighting for the QCD topological susceptibility

P. Thomas Jahn[1,*] Parikshit M. Junnarkar[1,†] Guy D. Moore[1,‡] and Daniel Robaina[2,§]

[1]*Institut für Kernphysik (Theoriezentrum), Technische Universität Darmstadt,*
Schlossgartenstraße 2, D-64289 Darmstadt, Germany
[2]*Max-Planck-Institut für Quantenoptik, Hans-Kopfermann-Straße 1, D-85748 Garching, Germany*

We introduce a reweighting technique which allows for a continuous sampling of temperatures in a single simulation and employ it to compute the temperature dependence of the QCD topological susceptibility χ_{top} at high temperatures. The method determines the ratio of χ_{top} between any two temperatures within the explored temperature range. We find that the results from the method agree with our previous determination and that it is competitive with but not better than existing methods of determining $d\chi_{\text{top}}/dT$. The method may also be useful in exploring the temperature dependence of other thermodynamical observables in QCD in a continuous way.

I. INTRODUCTION

The axion solution to the strong CP problem, proposed more than four decades ago [1–3], solves the fine tuning problem of the smallness of the θ parameter in QCD by introducing a new light ($<$meV) pseudo Goldstone boson in the Standard Model. Soon after it was proposed, it was also realized that long-wavelength axions would be produced abundantly early in the big bang and can serve as a candidate for dark matter [4,5]. These proposals have heightened interest both in searching experimentally for the axion and in better understanding its phenomenology and cosmological history; for a review of these topics see for instance Ref. [6]. An important input for the cosmological production of axions and for their contemporary properties is the QCD topological susceptibility $\chi_{\text{top}}(T)$, which determines the axion mass:

$$m_{\text{A}}^2(T) = \frac{\chi_{\text{top}}(T)}{f_{\text{A}}^2}. \tag{1}$$

Here $m_{\text{A}}(T)$ is the temperature dependent axion mass and f_{A} is the axion decay constant. The low-temperature value of χ_{top} is well determined [7,8], but the high-temperature

*tjahn@theorie.ikp.physik.tu-darmstadt.de
†parikshit@theorie.ikp.physik.tu-darmstadt.de
‡guymoore@theorie.ikp.physik.tu-darmstadt.de
§daniel.robaina@mpq.mpg.de

regime determines axion production efficiency; it is particularly important to determine $\chi_{\text{top}}(T)$ in the temperature regime from 500 to 1200 MeV [9]. In this temperature range, this can only be achieved by nonperturbative lattice investigations [10]. The most straightforward lattice methods, based on brute-force sampling of the gauge field configuration space, face difficulties in this temperature range because topologically nontrivial gauge field configurations become very rare, leading to a loss of statistical power. New methodologies are needed to overcome this problem. In recent years one such methodology has been developed [11,12] which provides access to high temperatures. In these calculations, it was shown that the difference of the expectation of the QCD action in two topological sectors can provide a determination of $d(\ln\chi_{\text{top}})/d(\ln T)$, which can be integrated to provide the temperature dependence of χ_{top}.

Alternatively, one can approach the problem at a fixed, high temperature by reweighting between topological sectors. A first attempt, based on a fixed guess for the reweighting function [13], explored temperatures around 500 MeV. Another method, in which the reweighting function is determined dynamically via an iterative self-consistent technique, was introduced in [14] and further improved in [15], where it was shown to be effective in the pure-glue theory up to at least $7T_{\text{c}}$.

Our main motivation for this work is to explore whether a new technique might improve existing methods for determining χ_{top}. The method introduced here is related to but distinct from the technique of Refs [11,12]. We train a single Markov chain Monte Carlo simulation to explore a wide range of temperatures in a detailed-balance respecting way by replacing the weighting function $\exp(-\beta S)$ with $\exp(-W(S))$, where $W(S)$ is established by an iterative

procedure such that the resulting ensemble can be reweighted to describe any temperature in a relatively wide range. We do this separately in the $Q = 0$ and $|Q| = 1$ (nontopological and instanton-number ± 1) sectors, and combine with a determination of $\chi_{\text{top}}(T)$ at one temperature to determine χ_{top} across the full accessible temperature range. We show that the approach has a similar efficiency to [11], with both (slight) advantages and disadvantages relative to their approach. The technique can be extended to include fermions if the line of constant physics [lattice $m(\beta)$] is known. However we find that our approach developed in Ref. [15] seems to afford better numerical efficiency.

In the next section we review the definition of topological susceptibility and lay the groundwork for our technique. Section III introduces our technique to carry out a detailed-balance preserving Markov chain over a range of temperatures. Then Sec. IV presents our results and Sec. V closes with a discussion.

II. TOPOLOGICAL SUSCEPTIBILITY

The topological susceptibility is defined as

$$\chi_{\text{top}}(T) = \frac{T}{V}\langle Q^2 \rangle = \frac{T}{V}\int d^4x d^4y \langle q(x)q(y)\rangle, \quad (2)$$

where $Q = \int d^4x q(x)$ and q is the topological charge density, V the spatial volume and T the temperature. In the continuum Q always takes an integer value while on the lattice one requires a refined definition of Q. Considering the continuum integer topological charge Q one can write the partition function as a sum over topological sectors:

$$Z = \int \mathcal{D}A e^{-\beta S} = \sum_{N \in \mathcal{Z}} \int \mathcal{D}A e^{-\beta S}\delta(Q - N) \equiv \sum_N Z_N. \quad (3)$$

The susceptibility is then given by

$$\chi_{\text{top}}(T) = \frac{T}{V}\frac{\int \mathcal{D}A e^{-\beta S}Q^2}{\int \mathcal{D}A e^{-\beta S}},$$

$$= \frac{T}{V}\frac{\sum_N N^2 Z_N}{\sum_N Z_N}. \quad (4)$$

At low temperatures and/or large volumes, this sum will have important contributions from many N values. But at a sufficiently high temperature, such that $V/T\chi_{\text{top}}(T) \ll 1$, the numerator is dominated by $N = 1$ and $N = -1$ and the denominator is dominated by $N = 0$. Renaming $Z_1 + Z_{-1} \rightarrow Z_1$ (the part of the partition function where $Q^2 = 1$, that is, $Q = \pm 1$) and considering a finite volume we find

$$\chi_{\text{top}}(T) \simeq \frac{T}{V}\frac{Z_1}{Z_0} = \frac{1}{a^4 N_x N_y N_z N_\tau}\frac{Z_1}{Z_0},$$

$$\chi_{\text{top}}(\beta)a^4(\beta) = \frac{1}{V_L}\frac{Z_1}{Z_0}. \quad (5)$$

Here we have reexpressed the temperature and volume as they would appear in a lattice calculation, with lattice spacing a and lattice extents (N_τ, N_x, N_y, N_z) in the temporal and the three space directions, and $V_L = N_\tau N_x N_y N_z$ the number of lattice sites. The last line emphasizes that the lattice spacing is a function of the lattice gauge coupling $\beta = 6/g_{\text{latt}}^2$. The nontrivial relation between these two quantities is determined by a scale setting measurement.

In this work, we will present a method which determines the ratio of susceptibilities at two specified temperatures. With a fixed lattice extent in the spatial and temporal directions, the two temperatures will correspond to two gauge couplings β_h and β_c (h for hot and c for cold) which will in turn correspond to two different lattice spacings $a(\beta_h)$ and $a(\beta_c)$. The ratio of susceptibilities at two temperatures will then be given as

$$\frac{\chi_{\text{top}}(\beta_h)a^4(\beta_h)}{\chi_{\text{top}}(\beta_c)a^4(\beta_c)} = \frac{Z_1(\beta_h)}{Z_1(\beta_c)}\frac{Z_0(\beta_c)}{Z_0(\beta_h)}. \quad (6)$$

Our method will also allow for the determination of the above ratio for *any* pair of temperatures within the prespecified range. We will do so by computing the ratio of partition functions using two Monte Carlo simulations: one that works within the $Q = 1$ topological sector and one which works within the $Q = 0$ sector, but with each simulation exploring the full range of β values, as we describe in the next section.

III. TEMPERATURE REWEIGHTING

In a given Monte Carlo simulation, we intend to sample a range of temperatures continuously over a prespecified range in pure glue QCD regularized on a finite space-time lattice. We will accomplish this using a reweighting method very similar to the old proposal of Berg and Neuhaus [16]. In this section we first show how a reweighted Monte Carlo simulation can be used to determine the susceptibility; then we show how the reweighted Monte Carlo calculation can be carried out; and finally we show how to determine the reweighting function itself. In practice, the numerical work proceeds in exactly the opposite order.

A. Susceptibility using reweighting

In the standard Monte Carlo simulation one evaluates the partition function $\mathcal{Z}(\beta)$ shown below at a particular gauge

coupling β by generating gauge configurations with the probability distribution $dP[U]$ as[1]

$$\mathcal{Z}(\beta) = \int \mathcal{D}U e^{-\beta S}, \qquad dP[U] = \frac{dU e^{-\beta S}}{\mathcal{Z}(\beta)}. \qquad (7)$$

In such a Monte Carlo calculation, a single temperature is simulated which is controlled by the gauge coupling β since the lattice dimensions are fixed. In order to simulate a range of β values and therefore a range of temperatures, the sampling weight in the usual Monte Carlo simulation in Eq. (7) is replaced with

$$\int \mathcal{D}U e^{-\beta S} \to \int \mathcal{D}U e^{-W(S)}. \qquad (8)$$

The weight function $W(S)$ is a general function of the action S, which should be chosen such that the above integral samples a wide range of S values rather uniformly. Assuming that such a $W(S)$ function can be found, a Monte Carlo simulation with $W(S)$ generates an ensemble which can be reweighted to determine expectation values at a given β value β_0 via

$$\mathcal{Z}(\beta_0) = \int \mathcal{D}U e^{-W(S)} e^{+W(S)-\beta_0 S}$$
$$\propto \sum_i e^{W(S_i)-\beta_0 S_i}, \qquad (9)$$

where i indexes the sampled configurations. In practice this partition function is used to determine expectation values for operators via

$$\langle \mathcal{O} \rangle = \mathcal{Z}^{-1} \int \mathcal{D}U e^{-\beta_0 S} \mathcal{O} \simeq \frac{\sum_i e^{W(S_i)-\beta_0 S_i} \mathcal{O}_i}{\sum_i e^{W(S_i)-\beta_0 S_i}}. \qquad (10)$$

The condition that $W(S)$ will work well is that it samples S rather uniformly over a range of action values. (If there is an S-range where $W(S)$ strongly undersamples, a Markov chain will have trouble moving from one side of this range to the other; and if there is a range where it strongly oversamples, that will reduce the statistical power afforded to all other ranges.) Call the smallest and largest action values over which we want uniform sampling $S_{\min}$ and $S_{\max}$. Also define the density of states

$$\rho(S_0) \equiv \int \mathcal{D}U \delta(S[U] - S_0). \qquad (11)$$

Then our goal is that, between $S_{\min}$ and $S_{\max}$, the function $W(S)$ should be close to $\ln(\rho(S))$. Then the probability

---

[1]We use square brackets to emphasize functionals, e.g., functions of fields; $W(S)$ because S takes a real value, but $S[U]$ because U is a field.
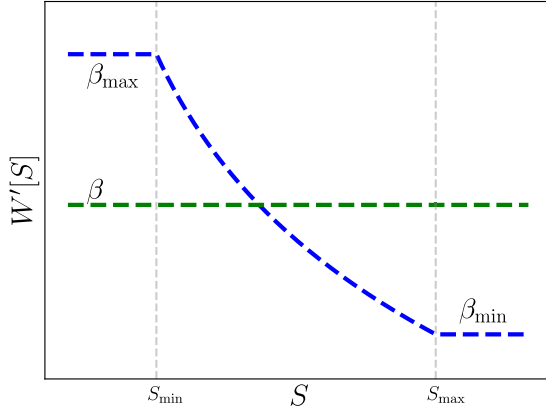
(density) to obtain a particular S value S_0 under the W-weighted path integral, $P(S_0) = \int \mathcal{D}U e^{-W(S[U])} \delta(S[U]-S_0) = \rho(S_0) e^{-W(S_0)}$, will be nearly uniform.

It is also informative to think about the function $W(S)$ in terms of its derivative

$$W'(S) = \frac{dW(S)}{dS} \equiv \beta(S). \qquad (12)$$

We name $W'(S) = \beta(S)$ for two reasons. First, if we choose $W(S) = \beta S$ with a fixed β value, then we get precisely the standard ensemble, and W' is precisely the β value of the ensemble. And if standard sampling with $\beta = \beta_0$ returns an expectation value for the action of $\langle S \rangle_{\beta_0} = S_0$, then we will find that under sampling using $W(S) = \ln(\rho(S))$, that $W'(S_0) \simeq \beta_0$. To see this, expand $\ln \rho(S)$ in a Taylor series about S_0:

$$\ln \rho(S) = \ln \rho(S_0) + \beta(S_0)(S - S_0)$$
$$+ \beta'(S_0)(S - S_0)^2/2 + \cdots. \qquad (13)$$

The range of S values of interest is set by $1/\sqrt{\beta'(S_0)}$ and is small compared to S in the thermodynamical limit; this controls the convergence of the Taylor series. Using the series to evaluate $\langle S \rangle$ for the choice $\beta = \beta(S_0)$, we find

$$\langle S \rangle_{\beta_0} = \frac{1}{Z} \int dS S \rho(S) e^{-\beta(S_0)S},$$
$$= \frac{1}{Z} \int dS S e^{-\beta(S_0)S + \ln \rho(S_0) + \beta(S_0)(S-S_0)}$$
$$\times e^{\beta'(S_0)(S-S_0)^2/2 + \cdots},$$
$$\simeq \frac{1}{Z} e^{-\beta(S_0)S_0 + \ln \rho(S_0)} \int dS S e^{\beta'(S_0)(S-S_0)^2/2}, \qquad (14)$$

where Z is the value of the integral without the explicit S factor. Since $\beta' < 0$, this is a Gaussian integral centered at $S = S_0$, and the expectation value is $S = S_0$ up to corrections which vanish in the thermodynamical limit. Therefore, up to corrections which vanish in the thermodynamical limit, a good choice for $W(S)$ will have, as a first derivative, $W'(S) = \beta(S)$ the β value which returns the given S as the expectation value for the action. Hence, $W'(S)$ will look approximately as shown in Fig. 1.

Now let us return to the application of this approach to the study of the susceptibility χ_{top}. By establishing two reweighting functions $W(S)$ and $W_Q(S)$ for the $Q = 0$ and $|Q| = 1$ ensembles, respectively, we can generate multi-temperature ensembles, labeled by i and iQ, which, respectively, sample the $Q = 0$ and the $|Q| = 1$ ensembles across temperatures. The ratio needed in Eq. (6) is then given by

FIG. 1. A cartoon of the Monte Carlo sampling of $W(S)$. The green line represents a W choice which will sample one β value, while the blue line will sample all β values between $\beta_{\max}$ and $\beta_{\min}$.

$$\frac{\chi_{\text{top}}(\beta_{\text{h}})a^4(\beta_{\text{h}})}{\chi_{\text{top}}(\beta_{\text{c}})a^4(\beta_{\text{c}})}$$
$$= \left(\frac{(\sum_{iQ} e^{W_Q(S_{iQ})} e^{-\beta_{\text{h}} S_{iQ}})(\sum_i e^{W(S_i)} e^{-\beta_{\text{c}} S_i})}{(\sum_{iQ} e^{W_Q(S_{iQ})} e^{-\beta_{\text{c}} S_{iQ}})(\sum_i e^{W(S_i)} e^{-\beta_{\text{h}} S_i})}\right). \quad (15)$$

There are two subtleties associated with this expression. The first is that Eq. (9) only determines the partition function up to an overall multiplicative factor. However, this multiplicative factor cancels between the numerator and denominator expressions computed from the same sample. The second subtlety is that the partition function $\mathcal{Z}$ also has severe lattice-spacing dependent renormalizations, which we cannot easily compute. Fortunately, these cancel because each lattice spacing occurs once in the numerator and once in the denominator in Eq. (15).

B. Update algorithm with $W(S)$

In this section we explain the algorithm to perform a Monte Carlo simulation with weight function $e^{-W(S)}$. Our aim is to generate a sampling with probability distribution

$$dP[U] = \frac{e^{-W(S[U])}}{\int \mathcal{D}U e^{-W(S[U])}}, \quad (16)$$

where $S[U]$ is the standard lattice gauge action and we assume that $W(S[U])$ is a known differentiable function of the action S. Simulating such a weight requires a slight modification of the standard hybrid Monte Carlo (HMC) algorithm [17].[2]

As in the standard HMC algorithm, we introduce canonical momenta π_μ for the link variables U_μ, and define a Hamiltonian for this system as

---

[2]It is also straightforward to use a mixture of heatbath and overrelaxation steps. However this approach does not generalize to the unquenched theory, so we concentrate on the HMC approach.

$$\mathcal{H}(\pi, U) \equiv \sum_{\mu,x} \frac{1}{2}(\pi_\mu(x))^2 + W(S[U]), \quad (17)$$

$$S[U] = \sum_\square \left(1 - \frac{1}{3}\text{Tr}\square\right), \quad (18)$$

where $S[U]$ in Eq. (18) is the standard Wilson gauge action written *without* the gauge coupling β prefactor. The standard HMC algorithm would use the same Hamiltonian but with $W(S[U])$ replaced by $\beta S[U]$, that is, it would use a strictly linear function for $W(S[U])$.

A single HMC update trajectory consists of the standard steps:
(1) Picking a random canonical momentum $\pi_\mu(x)$ from a Gaußian ensemble independently for each of the elements of the Lie algebra.
(2) Solving the following Hamilton equations of motion (shown here schematically):

$$\frac{dU}{dt} = -i\pi U, \quad (19)$$

$$\frac{d\pi}{dt} = iU^\dagger \frac{\partial W(S[U])}{\partial U} = iU^\dagger \frac{dW(S[U])}{dS[U]} \frac{\partial S[U]}{\partial U} \quad (20)$$

for a total time t_0. The derivative with respect to U_μ is a Lie derivative and in this sense these equations are a schematic representation. Here the time t is a fictitious variable under which the Hamiltonian $\mathcal{H}$ is conserved.

These Hamiltonian equations are discretized using a time-symmetric solver such as the leapfrog or Omelyan algorithms. Under these algorithms, one iteratively solves Eq. (19) for all link variables U at fixed π, and then solves Eq. (20) for all π variables at fixed U. Before applying Eq. (20), we must compute $S[U]$ and use the (instantaneous) value of $W'(S[U])$ in place of the usual factor of β at each time step during the update.

The use of a time-symmetric algorithm is essential, since it ensures the property that, if the pair (U_i, π_i) (i for initial) is carried to (U_f, π_f) under the update algorithm, then the pair $(U_f, -\pi_f)$ is carried to $(U_i, -\pi_i)$ up to roundoff error effects. This is sufficient to ensure that the algorithm converges, in a Fokker-Planck sense, to the probability distribution $\exp(-\mathcal{H})$ provided that we also include a Metropolis accept/reject step.
(3) In adding the Metropolis step, the change in the Hamiltonian $\Delta\mathcal{H} = \mathcal{H}[U_f, P_f] - \mathcal{H}[U_i, P_i]$ is compared to a random number drawn uniformly from the interval [0, 1]: $\mathcal{R}[0, 1]$. Whenever $e^{\Delta\mathcal{H}} < \mathcal{R}[0, 1]$, we accept the change, and proceed with U_f as our new configuration. Otherwise we revert to U_i, that is, we reject the update.

The only differences with respect to the standard HMC algorithm [17] are the use of $W'(S[U])$ the "instantaneous β value" in place of β in Eq. (20) and the use of $W(S[U])$ in place of $\beta S[U]$ in the Metropolis accept-reject step. Both modifications are compatible with the time-symmetry of the update algorithm and therefore preserve detailed balance. With these modifications, the HMC algorithm now generates the desired probability distribution.

For the case of a $Q = 1$ simulation, an additional accept-reject step is needed, in which we check to see whether the configuration has fallen down into the $Q = 0$ sector and reject the update if this is the case. In practice we can buffer every Nth configuration and only perform this step after every N HMC update steps, reverting to the last buffered configuration when the check fails. We define Q as the lattice sum of an a^2-improved topological density definition after $\tau_F = 2.4a^2$ units of Wilson flow, as in [15]. We find that values of $N = 5$ or $N = 10$ are adequate to preserve a good acceptance rate.

Lastly we remark on the optimal length of the individual trajectories. The figure of merit for trajectory length is the mean-squared change in $S[U]$ per unit numerical effort. The numerical effort is approximately linear in trajectory length t_0. For a short trajectory, $\Delta S[U]$ is linear, and $(\Delta S[U])^2$ quadratic, in trajectory length; but beyond a certain (fairly short) trajectory length the action change saturates. Therefore we started with a study of how $(\Delta S[U])^2$ varies with trajectory length and chose the value which maximizes $(\Delta S[U])^2 / t_0$; $t_0 \simeq 0.75a$. A single trajectory leads to a change of $(\Delta S[U]/S[U])^2 \simeq 3/N_{\text{dof}}$ where $N_{\text{dof}} = 24V_L$ is the number of lattice degrees of freedom (3 polarizations and 8 colors per site). Therefore, for a well-chosen $W(S[U])$ function, since changes to $S[U]$ accumulate in a Brownian fashion, the number of updates needed to explore the full β range is of order

$$N_{\text{updates}} \sim N_{\text{dof}} \ln^2(\beta_{\max}/\beta_{\min}). \tag{21}$$

C. Choice and determination of $W(S)$

Now we return to the question of how to determine the weight function $W(S)$. We start by choosing the range of β values we want to explore, $\beta \in [\beta_{\min}, \beta_{\max}]$. This choice is purely determined by the temperature range we want to study. Since W is a function of S, not β, we must convert these limiting values into limiting S values. We do this by performing short fixed-β Markov chains to establish $S_{\max} = \langle S[U] \rangle_{\beta_{\min}}$ and $S_{\min} = \langle S[U] \rangle_{\beta_{\max}}$. Our goal is to find a good description of $W(S)$ in the range between these two limiting values.

Our procedure will be almost the same as in our previous work [14,15]. First, we choose a discrete set of values S_i, $i = (0, \ldots, N_i)$, with $S_0 = S_{\min}$ and $S_{N_i} = S_{\max}$. (N_i is the total number of intervals between $S_{\min}$ and $S_{\max}$.) The

function $W(S)$ will be determined based on the values at these S points, $W_i = W(S_i)$. In our previous work, we used weight functions which were interpolated in a piecewise-linear fashion between such control points. But because the HMC algorithm described above works best when both W and W' are continuous, we instead define $W(S)$ to be the cubic spline interpolation between the $W(S_i)$. The spline function's definition also requires boundary conditions; we take these to be that $W'(S_{\min}) = \beta_{\max}$ and $W'(S_{\max}) = \beta_{\min}$. We define W outside the range of the spline function by fixing $W'(S > S_{\max}) = \beta_{\min}$ and $W'(S < S_{\min}) = \beta_{\max}$. This is also illustrated in Fig. 1.

It remains to determine the values for $W(S_i)$. We make an initial guess as follows. We choose an intermediate β value $\beta_{\text{mid}} = (\beta_{\max} + \beta_{\min})/2$ and we perform a short fixed-β Monte Carlo simulation to find the corresponding S value S_{mid}. We take $W'(S)$ to be a quadratic fit based on the three values $(\beta_{\max}, \beta_{\text{mid}}, \beta_{\min})$ at $(S_{\max}, S_{\text{mid}}, S_{\min})$ respectively, and integrate to get a starting guess for $W(S)$, which we evaluate at the S_i to get $W(S_i)$. Using more intermediate points could give a better starting guess, but it would still be necessary to apply the following automated iterative improvement scheme.

We iteratively improve the values $W(S_i)$ by using an automated scheme which we introduced in Ref. [14] (see also [18] where a similar scheme was developed). Consider the difference in the way $W(S)$ samples and the way $W_{\text{ideal}}(S) = \ln \rho(S)$ samples. Wherever $W(S) < W_{\text{ideal}}(S)$, $e^{-W(S)}$ is too large, and too many samples are taken. Similarly, where $W(S) > W_{\text{ideal}}(S)$, $e^{-W(S)}$ is too small, and we fail to sample. Therefore, it is natural to assume that, during a Markov chain, we are mostly sampling in places where $W(S)$ is too small. So the philosophy of our update algorithm it to run a Markov chain, and after every update step, to assume that the current S value has a too-small W value, and to slightly increase $W(S)$ at the current S value of the chain. [This updating philosophy was originally proposed by Berg and Neuhaus [16], and it is also used in "metadynamics" updates [19–21]; but our definition and update of $W(S_i)$ are different.]

There are two issues with applying this in practice. First, $W(S)$ is determined by values at a discrete set of points S_i; which $W(S_i)$ should be adjusted, and by how much? We address this by adjusting the two S_i values *closest to* the current S value, in proportion to how close they are. Second, how large should the adjustments to $W(S_i)$ be? At the beginning, our guess for $W(S_i)$ may be off by quite a bit, and we should use larger adjustments. But as our determination improves, we want to make the algorithm more stable so that it converges towards well determined values. We address this by defining a parameter s_r which tells how strongly we change the W function, which is reduced over the course of the Markov chain based on a criterion which shows whether the $W(S)$ function is giving good performance. In detail, our procedure after each HMC step is the following:

(1) If $S[U] < S_0$ or $S[U] > S_{N_i}$, then S is out of range and $W(S)$ is not updated.

(2) If $S[U]$ is in range, then we determine S_i, S_{i+1} such that the current S value lies between them, $S_i < S[U] < S_{i+1}$.

(3) We increase $W(S_i)$ and $W(S_{i+1})$ by $\Delta W(S_i) = \frac{s_r(S_{i+1} - S[U])}{(S_{i+1} - S_i)}$ and $\Delta W(S_{i+1}) = \frac{s_r(S[U] - S_i)}{(S_{i+1} - S_i)}$. Here s_r is an update strength, whose evolution we discuss next.

(4) If $i = 0$ or $i + 1 = N_i$ so we are in the first or last interval, then the boundary value $W(S_0)$ or $W(S_{N_i})$ is updated with double strength. This is because every $W(S_i)$ value is updated when S lies in the two intervals on either side of S_i, except these boundary values which are only updated when S lies in the single interval they bound. Therefore, to be updated the same amount as the other $W(S_i)$ values requires that these updates occur with double strength.

The value s_r determines how strongly the algorithm adjusts $W(S)$. Since the initial guess is rather crude, we must initially adjust $W(S)$ rather aggressively. Therefore we choose s_r initially such that, in the estimated time it takes for the Monte Carlo simulation to go from the top to the bottom and back, the $W(S_i)$ will change by of order 100: $s_r = 100 N_i / N_{\text{updates}}$, with N_{updates} given in Eq. (21). We know that the update algorithm is successfully modifying $W(S_i)$ because the algorithm is capable of moving from $S_{\min}$ to $S_{\max}$ and back successfully. Therefore, each time the S-value makes its way from the first interval $(S < S_1)$ to the last interval $(S > S_{N_i-1})$ and back (which we call a "sweep"), we reduce s_r; at first we reduce it by a factor of 2 after each sweep, but after the average $W(S_i)$ changes by less than 1 per sweep, we change it by a reduced amount. Specifically, if a sweep is performed in N_s HMC updates and $N_s s_r / N_i > 1$, then s_r is multiplied by $1/2$; but if $N_s s_r / N_i < 1$, then s_r is multiplied by $1 - N_s s_r / (2 N_i)$. The update ends when five sweeps change the average $W(S_i)$ by a total of less than 1, that is, the total N_s value summed over five consecutive sweeps has $(\sum N_s s_r)/N_i < 1$. Since the changes to $W(S_i)$ and s_r and the stopping criterion are all simple and predefined, the algorithm runs in an automated fashion without the need for by-hand intervention.

We show the difference between the initial guess for $W(S)$ and the finally determined $W(S)$ function, for the lattice and β ranges specified in the next section, in Fig. 2. The figure shows that our starting guess differed from the finally determined $W(S)$ by of order 60, which is enough that the initial guess would completely fail to sample all S values; the iterative improvement is necessary to provide a quality $W(S)$ determination.[3] The complete built W' is shown in Fig. 3.

---

[3]We could further improve the initial guess by using more intermediate β values; but using enough values, with precise enough $\langle S \rangle$ determinations, to determine $W(S)$ to better than ± 1 costs the same as the W refinement algorithm described here.
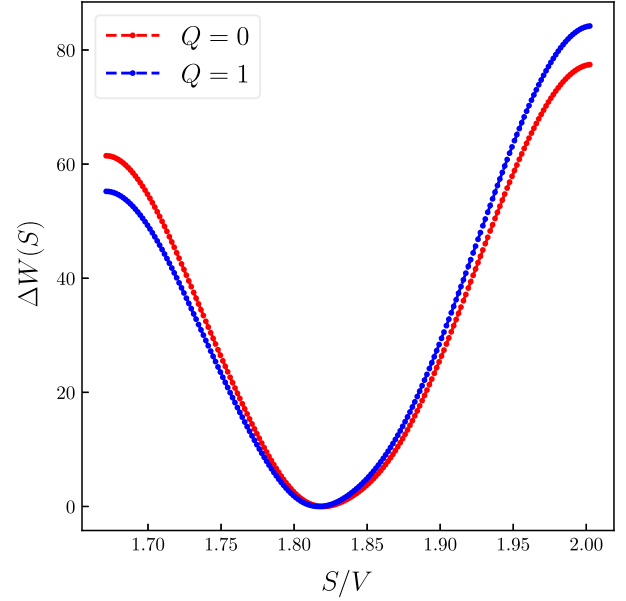

FIG. 2. Difference between the final determined $W(S)$ and the initial guess.

Finally, one must determine $W(S[U])$ for the $Q = 1$ sector. Here we can take as an initial guess the $W(S[U])$ value determined in the $Q = 0$ sector. To further refine this guess, we shift it by

$$W_{Q=1}(S) = W_{Q=0}(S) - 11 \ln(T/T_0), \qquad (22)$$

where T is the temperature associated with the β value described by the slope W' using a scale-setting relation between lattice coupling β and temperature T, and T_0 is a reference temperature which could for instance be the temperature at $\beta_{\min}$. The factor 11 is the expected temperature dependence of $\chi_{\text{top}} a^4$ when the lattice spacing a varies as $1/T$, at leading perturbative order. Again, Eq. (22) is only used to refine the initial guess for $W_{Q=1}(S)$; we again perform an automated W changing Markov chain to improve this guess; however the initial s_r value can be chosen to be smaller, $s_r \sim 1 N_i / N_{\text{updates}}$.

After the $W(S[U])$-setting Markov chains are completed, we freeze the values of $W_{Q=0,Q=1}(S)$ and use them in detailed-balance respecting Markov chains which we will use to determine the susceptibility as described in Sec. III A.

IV. RESULTS

In this section we present results of a simulation with the aforementioned $W(S)$ sampling. Our goal is only to test the method; we will not study multiple lattice spacings to attempt a continuum limit. Instead, we choose one lattice geometry from among the geometries studied in Refs. [14,15], and which we can therefore use to compare our results for susceptibility ratios to those obtained via a competing
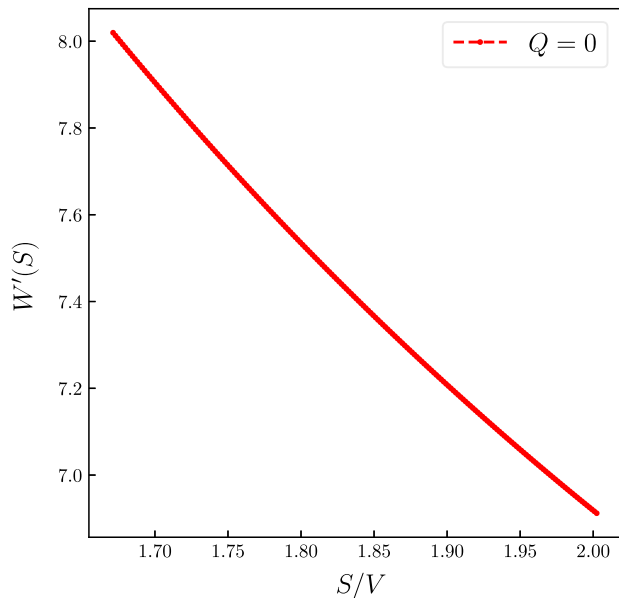
FIG. 3.   Final choice of $W'(S)$ for the lattice study described in Sec. IV.

TABLE I.   Numerical cost of the simulation on a lattice four-volume $V = 10 \times 32^2 \times 36$ with $\beta_{\min} = 6.9076$ and $\beta_{\max} = 8.01951$.

Procedure	Q	HMC Trajectories	Sweeps
Building	0	4×10^6	22
Building	1	8×10^6	17
Measurements	0	6.8×10^6	45
Measurements	1	4.9×10^6	35

Our main results are presented in Fig. 5, which shows how $\chi_{\mathrm{top}}a^4$ changes as a function of β across the range we study, evaluated from our Markov chains using Eq. (15). We have chosen to use a value near the beginning of the β range ($T = 2.8T_{\mathrm{c}}$) as the low temperature and to express all other temperatures in relation to this one. The figure, and further data presented in Table II, show that the error bars are smallest between nearby β values and grow to around ± 0.35 in $\ln(\chi_{\mathrm{top}})$ for the widest-separated temperatures. Our results agree within error bars with the results in

technique. Our goal is to establish which approach, the one described here or the one described in Ref. [15], is more efficient at establishing the topological susceptibility at high temperature.

We choose to investigate a lattice with temporal extent $N_\tau = 10$ and spatial extent $32^2 \times 36$, with $(\beta_{\min}, \beta_{\max}) = (6.9076, 8.01951)$, which corresponds, according to the scale setting calculation of Ref. [22], which we will use throughout, to $T_{\beta_{\min}} = 2.5T_{\mathrm{c}}$ and $T_{\beta_{\max}} = 9.4T_{\mathrm{c}}$. This scale-setting relation involves applying a fit to scale-setting data beyond the range where the reference has performed simulations, that is, an extrapolation, which means it may not be absolutely trustworthy. However, by expressing our results in terms of $\chi_{\mathrm{top}}(\beta)a^4(\beta)$, we can remain agnostic about the relation between T and β and just determine how accurately we can determine $\chi_{\mathrm{top}}a^4$ as a function of β for *this specific* lattice geometry. Getting continuum results over a range of temperatures will then of course require multiple lattice N_τ values and a reliable scale setting. However, for the time being our goal is just to evaluate the precision-to-numerical-cost ratio of the technique, so we leave this problem for later. The number of updates used in each procedure are listed in Table I. Note that an unfortunate choice of s_r made the $Q = 1$ building unnecessarily inefficient; a more careful choice should have required a fraction as many trajectories, so that the trajectory count would be dominated by the measurements.

We begin with a check that our $W(S)$ function correctly generates a rather uniform sample of configurations across the desired β range. We investigate this by plotting a histogram of the S values measured during the sampling Markov chain, both for the $Q = 0$ and the $Q = 1$ ensembles, shown in Fig. 4. The sample is adequately uniform.
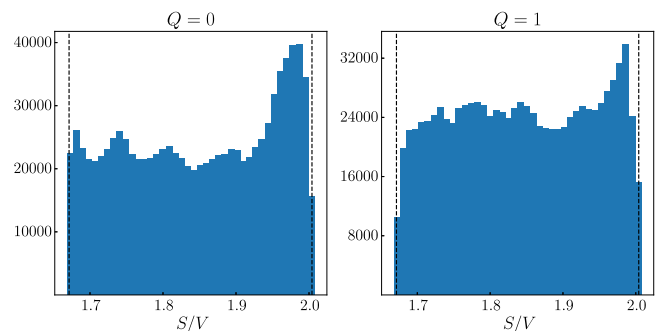

FIG. 4.   A histogram of the S distribution of our Markov chain samples in the $Q = 0$ sector (left) and the $Q = 1$ sector (right).
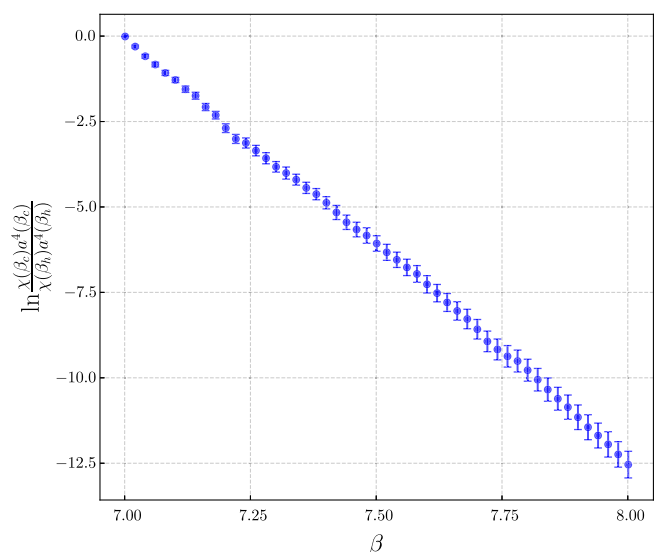

FIG. 5.   Results of the ratio in Eq. (15).

TABLE II. Several temperatures, the corresponding β values using the scale setting of Ref. [22], and our results for the log susceptibility ratio and its 1-sigma statistical error.

T/T_c	β	$\ln(\chi_{\text{top}}a^4(T)/\chi_{\text{top}}a^4(2.8T_c))$	1-σ statistical error
3.5	7.1771	-2.25	0.17
4.0	7.2885	-3.68	0.20
5.0	7.4764	-5.79	0.24
7.0	7.7629	-9.39	0.30
9.0	7.9788	-12.20	0.36

Ref. [15] for those values where they are directly comparable.

V. DISCUSSION

We have shown that the method we propose can successfully find the β dependence of $\chi_{\text{top}}a^4$, and therefore the temperature dependence of the susceptibility if the line of constant physics [that is, $a(\beta)$] is known. This determines χ_{top} over a range of temperatures if it is known at the lowest temperature, which is where it is most easily determined by other approaches.

There are two key questions. Is it more or less effective than the rather similar approach of Refs [11,12]? And how does it compare with the approach of Ref. [15]?

The approach of Refs. [11,12] computes $d \ln(\chi_{\text{top}}a^4)/d\beta$ at several β values, which it integrates to determine $\chi_{\text{top}}a^4(\beta)$. We review this approach and compare it to our own in an appendix. To summarize, in a high-statistics determination, the approaches have essentially the same numerical precision. However, if lower precision is desired, the numerical cost associated with building the $W(S)$ functions in our approach is essentially "dead weight" which does not contribute to the statistical power. The other approach does not suffer from this problem and so it is more efficient for a low-statistics determination. Our approach has the advantage that it automatically includes all intermediate temperatures, while the alternative bases the determined $d \ln(\chi_{\text{top}}a^4)/d\beta$ on a finite set of values which may leave discrete integration errors. But it is not difficult to use enough values to render this a minor concern.

Finally, we want to compare the numerical efficiency to the method of Ref. [15]. Fortunately, the single lattice we investigated in this work was also used in that reference, and we can directly compare the error on $\chi_{\text{top}}a^4(\beta_h)/\chi_{\text{top}}a^4(\beta_c)$ found here with the error on the same quantity found there, along with the number of HMC trajectories needed in each case. In that reference $\chi_{\text{top}}(T)$ was determined at $\beta = (6.90097, 7.30916, 7.76294)$, corresponding to $T = (2.5, 4.1, 7.0)T_c$, a little narrower than the range considered here. The three determinations required a total of 9.2×10^6 trajectories, about half the

number which should have been needed here. The average trajectory length used was also shorter in that reference than what we used here. The final errors on $\ln(\chi_{\text{top}})$ in that study, for this lattice, were (0.09,0.09,0.08) at the three temperatures. In comparison, in comparing $2.8T_c$ to $7.0T_c$ we find statistical errors of 0.30. To reduce these errors to the level of the other study would therefore require about 10 times more statistics in our measurement runs, indicating that the present method is of order 10 times less numerically efficient. Moreover, Ref. [15] finds that the number of trajectories needed for a given statistical error barely changes as one increases the volume (larger aspect ratio) or makes the lattice finer (larger N_τ at fixed aspect ratio), whereas we know that the number of updates needed for the method described in this paper should scale with the number of lattice sites, see Eq. (21).

We conclude that our method is at least 10 times less efficient than the single-temperature reweighting approach of [15], and will become still less efficient as one goes closer to the large-volume and continuum limits. As we understand it, this also implies that it should be easier in principle to achieve small statistical errors with the approach of [15] than with the approach of [11,12].

APPENDIX: COMPARISON WITH THE SLOPE METHOD

Our approach is closely related to the approach of Refs. [11,12], and as we understand it, the errors per numerical effort are nearly the same. To explain this conclusion, we start with a quick review of their approach, and then look at the issue of statistical power in each approach.

Their approach also seeks to compute Eq. (6) and then use a determination of $\chi_{\text{top}}(\beta_c)$ to determine χ_{top} at other temperatures. Taking the log of Eq. (6) we find

$$\ln \frac{\chi_{\text{top}}(\beta_h)a^4(\beta_h)}{\chi_{\text{top}}(\beta_c)a^4(\beta_c)} = (\ln(Z_1(\beta_h)) - \ln(Z_1(\beta_c)))$$

$$- (\ln(Z_0(\beta_h)) - \ln(Z_0(\beta_c))). \quad \text{(A1)}$$

Now note that

$$Z(\beta) = \int \mathcal{D}A_\mu e^{-\beta S} \Rightarrow$$

$$-\frac{\partial \ln Z}{\partial \ln \beta} = \frac{1}{Z}\int \mathcal{D}A_\mu e^{-\beta S}\beta S = \langle \beta S\rangle \qquad (A2)$$

the β dependence of $\ln Z$ is set by the expectation value of the action. Therefore

$$\ln Z_1(\beta_h) - \ln Z_1(\beta_c) = -\int_{\ln \beta_c}^{\ln \beta_h}\langle \beta S\rangle_1 d(\ln\beta), \qquad (A3)$$

and the ratio we want is

$$\ln\frac{\chi_{\text{top}}(\beta_h)a^4(\beta_h)}{\chi_{\text{top}}(\beta_c)a^4(\beta_c)} = -\int_{\ln\beta_c}^{\ln\beta_h}(\langle\beta S\rangle_1 - \langle\beta S\rangle_0)d(\ln\beta). \quad (A4)$$

References [11,12] evaluate $\langle\beta S\rangle$ in both $Q = 0$ and $Q = 1$ ensembles at a number of β values, which are then used to estimate this integral by, e.g., the trapezoid rule.

To compute Eq. (A4) using our approach, first write

$$\int_{\ln\beta_c}^{\ln\beta_h} -\langle\beta S\rangle d(\ln\beta) = \int_{\ln\beta_c}^{\ln\beta_h}\frac{-\sum_i e^{W(S_i)-\beta S_i}\beta S_i}{\sum_i e^{W(S_i)-\beta S_i}}d(\ln\beta),$$

$$= \int_{\ln\beta_c}^{\ln\beta_h}\frac{d}{d\ln\beta}\ln\left(\sum_i e^{W(S_i)-\beta S_i}\right)d(\ln\beta),$$

$$= \ln\frac{\sum_i e^{W(S_i)-\beta_h S_i}}{\sum_i e^{W(S_i)-\beta_c S_i}}, \qquad (A5)$$

so applying Eq. (10) to Eq. (A4) leads directly to Eq. (15). Therefore the real difference between the approaches is whether Eq. (A4) is estimated based on interpolating results for several temperatures, or using a single Markov chain which spans all temperatures.

Now consider the statistical power of each approach. The accuracy of a Monte Carlo evaluation of $\langle\beta S\rangle$ is set by the variance of βS and the number of independent configurations used. The variance should be reasonably approximated as that for N_{dof} Gaussian random variables: $\sigma_{\beta S}^2 \simeq N_{\text{dof}}/2$. Therefore order-1 errors in $\langle\beta S\rangle$ require

$N_{\text{dof}}/2$ evaluations. Since the expectation value determines an integrand, this is multiplied by the integration range, so $N_{\text{dof}}/2$ evaluations return an error in the ratio of partition functions which is of order $\ln(\beta_{\max}) - \ln(\beta_{\min})$. Evaluating $\langle\beta S\rangle$ at multiple β values leads to a larger error at each evaluation, but because each is responsible for a narrower $\Delta\beta$ range and the errors are uncorrelated, the final statistical uncertainty is independent of the number of β values used in the evaluation and depends only on the total number of Markov steps and the width of the β range considered. The final error estimate is $\Delta\ln\chi = \ln(\beta_{\max}/\beta_{\min})\sqrt{N_{\text{dof}}/2N_{\text{updates}}}$. The error rises by $\sqrt{2}$ and N_{updates} is doubled when we recall that separate simulations are needed in the $Q = 0$ and $Q = 1$ sectors.

In comparison, we see in Eq. (21) that our approach can explore the full β range, leading to order-1 errors in $\ln\chi$, in $N_{\text{updates}} \sim N_{\text{dof}}\ln^2(\beta_{\max}/\beta_{\min})$. Therefore the two approaches produce errors per unit numerical effort which are the same up to an order-1 factor. In a numerical experiment on a toy problem (N independent Gaussian random variables x with action $S = \sum x^2/2$) we find that the order-1 factor is in fact 1, so the two approaches have the same statistical power per compute time, provided that $W(S)$ is well determined and neglecting the computational effort expended in evaluating it.

We should also remark on how each approach is extended to full (unquenched) QCD. In each case the main challenge is dealing with the way quark masses must be varied with the lattice spacing and therefore with β: $m = m(\beta)$ (which must also be determined as part of the scale setting). This added β dependence changes Eq. (A2), replacing $\langle\beta S\rangle \rightarrow \langle\beta S + \beta dm/d\beta\bar{\psi}\psi\rangle$. In our approach one must replace $W(S) \rightarrow W(S) + \bar{\psi}(\slashed{D} + m(W'(S)))\psi$ where we use W' in place of β for the scale dependence of m. This amends Eq. (20) by the addition of the standard fermionic force term and by the replacement $dW/dS[U] \rightarrow (dW/dS[U]) + (dm/d\beta)(dW'/dS[U])\langle\bar{\psi}\psi\rangle$ where $\langle\bar{\psi}\psi\rangle$ is the sum of the $\bar{\psi}\psi$ value over all sites in the current configuration. Finally, in Eq. (15), the $W - \beta S$ reweighting must be complemented by a determinant-ratio from the S-dependent mass to the physical mass for the desired β value.

[1] R. D. Peccei and H. R. Quinn, Phys. Rev. Lett. **38,** 1440 (1977).

[2] S. Weinberg, Phys. Rev. Lett. **40,** 223 (1978).

[3] F. Wilczek, Phys. Rev. Lett. **40,** 279 (1978).

[4] L. F. Abbott and P. Sikivie, Phys. Lett. **120B,** 133 (1983).

[5] J. Preskill, M. B. Wise, and F. Wilczek, Phys. Lett. **120B,** 127 (1983).

[6] I. G. Irastorza and J. Redondo, Prog. Part. Nucl. Phys. **102,** 89 (2018).

[7] G. Grilli di Cortona, E. Hardy, J. P. Vega, and G. Villadoro, J. High Energy Phys. 01 (2016) 034.

[8] M. Gorghetto and G. Villadoro, J. High Energy Phys. 03 (2019) 033.

[9] G. D. Moore, EPJ Web Conf. **175,** 01009 (2018).

[10] S. Borsanyi, M. Dierigl, Z. Fodor, S. D. Katz, S. W. Mages, D. Nogradi, J. Redondo, A. Ringwald, and K. K. Szabo, Phys. Lett. B **752**, 175 (2016).

[11] S. Borsanyi *et al.*, Nature (London) **539**, 69 (2016).

[12] J. Frison, R. Kitano, H. Matsufuru, S. Mori, and N. Yamada, J. High Energy Phys. 09 (2016) 021.

[13] C. Bonati, M. D'Elia, G. Martinelli, F. Negro, F. Sanfilippo, and A. Todaro, J. High Energy Phys. 11 (2018) 170.

[14] P. T. Jahn, G. D. Moore, and D. Robaina, Phys. Rev. D **98**, 054512 (2018).

[15] P. T. Jahn, G. D. Moore, and D. Robaina, arXiv:2002.01153.

[16] B. A. Berg and T. Neuhaus, Phys. Lett. B **267**, 249 (1991).

[17] S. Duane, A. D. Kennedy, B. J. Pendleton, and D. Roweth, Phys. Lett. B **195**, 216 (1987).

[18] M. Laine and K. Rummukainen, Nucl. Phys. **B535**, 423 (1998).

[19] A. Barducci, G. Bussi, and M. Parrinello, Phys. Rev. Lett. **100**, 020603 (2008).

[20] A. Laio, G. Martinelli, and F. Sanfilippo, J. High Energy Phys. 07 (2016) 089.

[21] F. Sanfilippo, A. Laio, and G. Martinelli, Proc. Sci., LATTICE2016 (**2017**) 274.

[22] Y. Burnier, H. T. Ding, O. Kaczmarek, A. L. Kruse, M. Laine, H. Ohno, and H. Sandmeyer, J. High Energy Phys. 11 (2017) 206.

[23] http://luscher.web.cern.ch/luscher/openQCD/index.html.