

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

S. Borsányi¹ and R. Kara²

University of Wuppertal, Department of Physics, Wuppertal D-42119, Germany

Z. Fodor

*Pennsylvania State University, Department of Physics, University Park, Pennsylvania 16802, USA,
University of Wuppertal, Department of Physics, Wuppertal D-42119, Germany,
Institute for Theoretical Physics, ELTE Eötvös Loránd University, Pázmány P. sétány 1/A,
H-1117 Budapest, Hungary
and Jülich Supercomputing Centre, Forschungszentrum Jülich, D-52425 Jülich, Germany*

D. A. Godzieba and P. Parotto³

Pennsylvania State University, Department of Physics, University Park, Pennsylvania 16802, USA

D. Sexty⁴

University of Graz, Institute for Physics, A-8010 Graz, Austria



(Received 2 March 2022; accepted 5 April 2022; published 29 April 2022)

We perform large scale simulations to characterize the transition in quenched QCD. It is shown by a rigorous finite size scaling that the transition is of first order. After this qualitative feature, quantitative results are obtained with unprecedented precision: We calculate the transition temperature $w_0 T_c = 0.25384(23)$ —which is the first per-mill accurate result in QCD thermodynamics—and the latent heat $\Delta E/T_c^4 = 1.025(21)(27)$, in both cases carrying out controlled continuum and infinite volume extrapolations. As it is well known, the cost of lattice simulations explodes in the vicinity of phase transitions, a phenomenon called critical slowing down for second order phase transitions and supercritical slowing down for first order phase transitions. We show that a generalization of the parallel tempering algorithm of Marinari and Parisi [Europhys. Lett. **19**, 451 (1992)] originally for spin systems can efficiently overcome these difficulties even if the transition is of first order, like in the case of QCD without quarks, or with very heavy quarks. We also report on our investigations on the autocorrelation times and other details.

DOI: [10.1103/PhysRevD.105.074513](https://doi.org/10.1103/PhysRevD.105.074513)

I. INTRODUCTION

There is no doubt in the theoretical physics community that the SU(3) Yang-Mills theory, or QCD without quarks, features a first order phase transition. The transition takes place at a temperature around 290 MeV, the precise value strongly depending on the choice of the variable to set the scale. It separates the cold confining phase from a hot phase with spontaneously broken center symmetry.

In the context of full QCD, this transition was expected to happen as a consequence of Hagedorn's exponential

spectrum [1], and a first phase diagram was drawn in Ref. [2]. Later, the necessity of a transition was derived from the underlying quantum field theory and linked to the center of the gauge group [3]. In contrast to the well known crossover transition in full QCD [4], the order of the transition between the confining and weak-coupling phases was argued to be first order in the quarkless SU(3) theory [5,6]. The main argument relies on the Z(3) symmetry of the three-dimensional effective theory and that no stable renormalization group fixed point is known with that symmetry. Such an effective theory has already been studied extensively in the context of the Potts model [7].

Our main motivation to study the transition of the SU(3) Yang-Mills theory is the exploration of the so-called “Columbia plot,” a diagram showing the nature of the QCD transition as a function of the light and strange quark masses $m_u = m_d = m_l, m_s$. Because of the absence of an exact Z(3) symmetry for finite quark mass, the transition

*borsanyi@uni-wuppertal.de

Published by the American Physical Society under the terms of the [Creative Commons Attribution 4.0 International license](https://creativecommons.org/licenses/by/4.0/). Further distribution of this work must maintain attribution to the author(s) and the published article's title, journal citation, and DOI. Funded by SCOAP³.

will turn into crossover below a certain critical mass [6]. This mass was estimated using Dyson-Schwinger equations [8] and in effective models [9]. In the opposing corner of the Columbia plot, referred to as the “chiral limit,” one again expects a first order transition for three or more massless flavors, following from the ϵ expansion of a linear σ model [10]. This diagram was first studied in lattice QCD by the Columbia group [11] and has been a subject of intense research ever since. In terms of simulation costs, the study of the upper right corner is considerably simpler than working close to the chiral limit, partly because of the slow convergence of the iterative solvers and partly because of the more elaborate fermion formulations that the chiral limit requires. However, simulations are in progress even near the chiral corner, and recent results challenge the theoretical assumptions on the phase diagram [12].

However, even the upper right corner is nontrivial to simulate. It has been addressed using reweighting from quenched QCD [13,14], reweighting from a leading order hopping-term expansion [15] or a three-dimensional effective theory for the Polyakov loop [16], or with direct simulations [17,18]. While these results are promising and appear to be consistent with matrix models [9], a continuum extrapolation could not be achieved, yet.

The main difficulty in pinpointing the critical mass is not simply coming from the presence of quarks, as the iterative solvers converge very quickly in these studies. The problem is more general: The critical slowing down becomes severe as the critical mass is approached. Even at a safe distance from the critical mass, simulations on the first order side of the phase transition line can be difficult [17], and the exploration of both phases with the proper statistical weight turns out to be extremely challenging.

Even in the quarkless SU(3) Yang-Mills theory, the clear evidence of the expected first order nature of the transition was quite challenging to obtain. Followed by the pioneering papers [19,20] to observe the transition in a gauge theory, Kajantie *et al.* observed a hysteresis in the Polyakov loop, the first numerical hint for a first order transition [21]. The initial two-state signal [22,23] was followed by first estimates of the latent heat [24,25]. To determine the nature of a transition from finite-box simulations, a finite size scaling is necessary. This was done for fixed (coarse) lattice spacing in Refs. [26,27]. However, not all works agreed with this: Refs. [28,29] proposed a second order transition. In fact, the SU(3) transition is weak. In comparison to other SU(N) theories [30], the latent heat in SU(3) falls short of the linear $1/N^2$ scaling [31].

While spin model simulations underwent great progress thanks to cluster algorithms [32], the gauge theory studies cited here mostly relied on local update algorithms. In such cases, the only tool at hand is the reweighting method of Ferrenberg and Swendsen [33,34], which allows combination and interpolation between simulation results. Reference [35] introduced a tempering algorithm to circumvent “freezing

simulations” in spin models. In a nutshell, the algorithm treats temperature (or any other control parameter) as a dynamical variable that walks around in the region of interest. The strong dependence of the order parameter on the dynamical control parameter fosters a quick decorrelation. Parallel tempering is a variant of this idea where a simulation is started for each possible parameter value. This has been applied successfully to, e.g., various spin models [36,37], adjoint SU(2) Wilson action modified by a Z_2 monopole suppression term [38,39], and to lattice QCD [40–43], though in the latter case, not for the study of phase transitions.

In this work, we advocate the use of parallel tempering in lattice QCD thermodynamics and demonstrate its virtues by presenting a precise description of the phase transition in the SU(3) gauge theory, calculating the transition temperature and the latent heat. We do this to lay the foundations of a longer term work with dynamical fermions [18]. Although there have been many studies on this transition, a complete infinite volume limit of continuum extrapolated observables has not been calculated, yet. In Sec. II, we describe the parallel tempering algorithm and show the resulting improvement in the autocorrelation times and simulation precision. In Sec. III, we present our determination of the critical couplings for the tree-level Symanzik improved gauge action used in this study, the continuum extrapolated transition temperature, and its infinite volume limit. We confirm the first order nature of the transition by performing a finite volume analysis of the continuum extrapolated susceptibilities in Sec. IV. Finally, we calculate the latent heat and compare this to the latest result in the literature [44] in Sec. IV B, then conclude.

II. TEMPERED SIMULATIONS OF THE TRANSITION

A. The parallel tempering algorithm

We first describe the parallel tempering algorithm [36,41,45,46].

Suppose we want to simulate some theory at various parameter sets p_i (which can include the gauge coupling β , fermionic masses, etc.). Connected to each parameter, set we have a subensemble Γ_i containing configurations (link variables, pseudofermion fields, etc.) and an action S_i dependent on the parameters p_i and on configurations from Γ_i . The parallel tempering simulation considers a Markov chain built on the direct product of the subensembles’ configuration spaces: $\Gamma_{\text{PT}} = \prod_{i=1}^N \Gamma_i$. Our intent is to build a Markov chain that equilibrates to the distribution,

$$P_{\text{PT}}^{\text{eq}}[\{a_i\}] = \prod_i P_i^{\text{eq}}(a_i) = \prod_i \frac{1}{Z_i} e^{-S_i(a_i)}, \quad (1)$$

where a_i are configurations from Γ_i , and Z_i is the partition function of subensemble Γ_i . The partition function of the whole ensemble is then calculated as

$$Z_{\text{PT}} = \prod_i Z_i. \quad (2)$$

We define two kinds of transitions on Γ_{PT} :

- (1) *Transitions within a subensemble:* We can use any Markovian updating procedure (HMC, local Metropolis update, etc.) on the subensemble Γ_i using the action S_i .
- (2) *Swapping update of two subensembles:* This step mixes the phase space of Γ_i and Γ_j . We propose to swap configurations $a \in \Gamma_i$ and $b \in \Gamma_j$ with probability $P_s(i, j)$.

To satisfy the detailed balance of the swapping updates, we must have

$$P_s(i, j)e^{-S_i(a)}e^{-S_j(b)} = P_s(j, i)e^{-S_i(b)}e^{-S_j(a)}. \quad (3)$$

This is easily satisfied with the usual Metropolis acceptance: $P_s(i, j) = \min(1, e^{-\Delta S})$, with

$$\Delta S = S_i(b) + S_j(a) - S_i(a) - S_j(b). \quad (4)$$

To reach the desired equilibrium distribution, updates within a subensemble are essential; they would suffice without any swapping updates (that would be equivalent to running independent calculations for each parameter set). As we show below, swapping updates reduce the autocorrelation times of the simulations immensely. The intuitive understanding behind this observation is that the swapping updates couple parameter sets with different autocorrelation times τ . As configurations do a “random walk” in parameter space, they decorrelate quickly at the low τ region and wander back to the high τ region to contribute an independent configuration. This is a more efficient way to produce independent configurations than running the Markov chain at a fixed parameter set.

For the parallel tempering to be effective, the acceptance rate of swapping updates must be carefully controlled, such that the action distributions of neighboring ensembles have a substantial overlap. This can be achieved through the control of the distance of the parameter sets p_i , which, thus, have to move closer to each other as the physical volume increases. This is easily satisfied if the number of streams in the transition region is kept constant as the width of the transition region scales with inverse volume.

To locate the phase transition point for a given spatial and temporal lattice size, we perform a β scan to search for the peak of the Polyakov-loop susceptibility or the zero of the third Binder cumulant (see Sec. III for details). Typically, we have simulations at 16–256 β values, and we use the parallel tempering algorithm. We introduce swapping updates at predefined points in the Markov chain (typically after five sweeps in each subensemble). This algorithm can be especially efficiently parallelized in our case where the β dependence of the action is simply given

as an overall factor of some function of the link variables. We set up a number of streams updating a configuration at certain β values (typically we use equidistant points). After five sweeps on their configuration, all the streams need to communicate one number (their action) to a master node, which proposes (and accepts or rejects) several swapping updates for each stream, and afterward, informs each stream which β value they ended up at. This means that the network bandwidth and computational requirements for the swapping updates are negligible, although some efficiency is lost as synchronization between the streams is required. Note that this synchronization loss is independent of the time that is needed for the calculation of one update on the configurations (e.g., it is $< 10\%$ if a stream waits on average less than half of an update time after every five updates).

B. Explorations with the tempering algorithm

Below we show results of the parallel tempering method, comparing it with the brute force (independent simulations at each β value) and brute force with the multihistogram method.

To set up the simulations, we need to choose the set of β values we want to carry out simulations at. The distance between neighboring ensembles, $\Delta\beta$, needs to be small enough such that the acceptance of the swapping updates is $O(1)$ (see below for more details).

In Fig. 1, we show the β history of two streams (out of 128) for a simulation on a $45^3 \times 10$ lattice. As one observes, the streams follow a trajectory similar to a

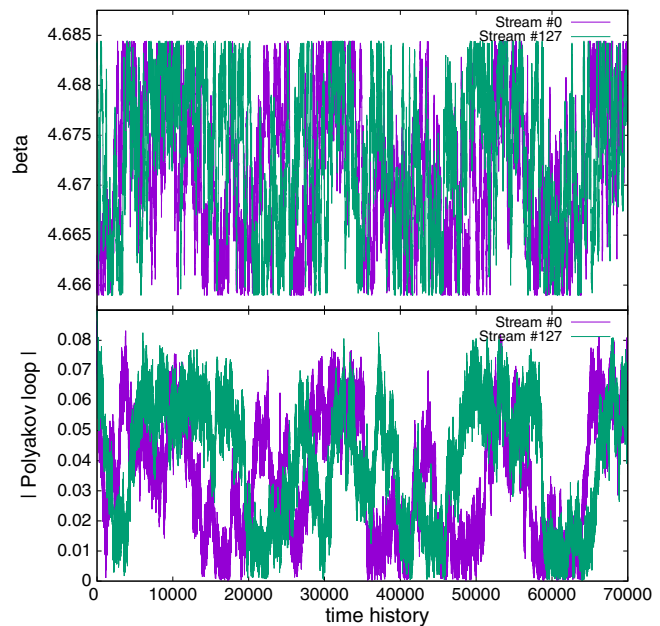


FIG. 1. The β history (above) and Polyakov loop average (below) of two streams as a function of the Monte Carlo sweeps in a simulation on a $45^3 \times 10$ lattice with 128 streams on the range $4.659 \leq \beta \leq 4.6844$ divided equidistantly.

random walk on the allowed β range, and they visit all points for a long enough simulation. This introduces correlations between ensembles at different β values, which we will take into account later. As also visible on Fig. 1, the value of the Polyakov loop changes relatively slowly during the history of the stream. If we look at the history of the Polyakov loop for a given β , the largest autocorrelation comes from the instances when the same stream contributes (even if the stream visited other β values in the

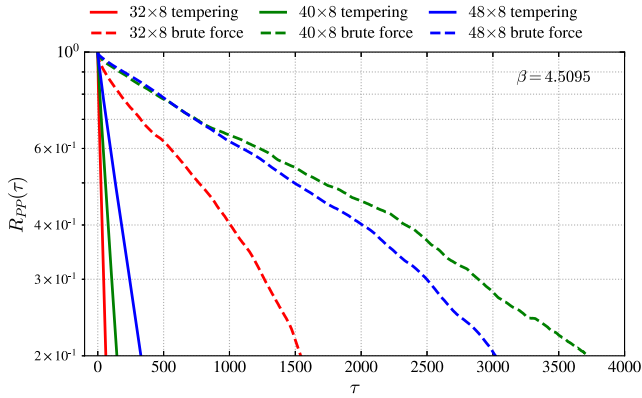


FIG. 2. The autocorrelation function of the absolute value of the Polyakov loop at $\beta = 4.5095$ for the lattice sizes $32^3 \times 8$, $40^3 \times 8$, $48^3 \times 8$, in a parallel tempering simulation (solid lines), using 64β values with spacings $\Delta\beta = 0.001$ and in single simulations (dashed lines).

meantime). We therefore reorder the Monte Carlo chain using the stream ID number of the contributions (and keep the chronological order among contributions from each stream). This ensures that the most correlated contributions will be close to each other (i.e., this ordering results in the largest autocorrelation for the given Monte-Carlo chain), which helps to minimize the correlations between blocks in the jackknife analysis. In Fig. 2, we show the resulting autocorrelation function for several lattice sizes at $\beta = 4.5095$, compared with autocorrelation functions of brute force simulations. In the left panel of Fig. 3, we show the measured autocorrelation times for various simulation setups. As one observes, the autocorrelation is substantially smaller for the parallel tempering simulations, although at the increased cost of having to simulate at all β values. Therefore, a fairer comparison would consider several brute force simulations at all β values. In the right panel of Fig. 3, we show the number of statistically independent configurations created in brute force and in parallel tempering simulations at similar parameters, using the same amount of computational resources. We observe that τ_{PP} , the autocorrelation time of the Polyakov loop, also improves greatly as the number of β values increases in a given β range and as $\Delta\beta$ decreases (increasing the acceptance rate of swap updates). From the red and black points in Fig. 3, we see that the number of independent configurations created by the tempering algorithm increases substantially by simply decreasing $\Delta\beta$. Increasing the density of β

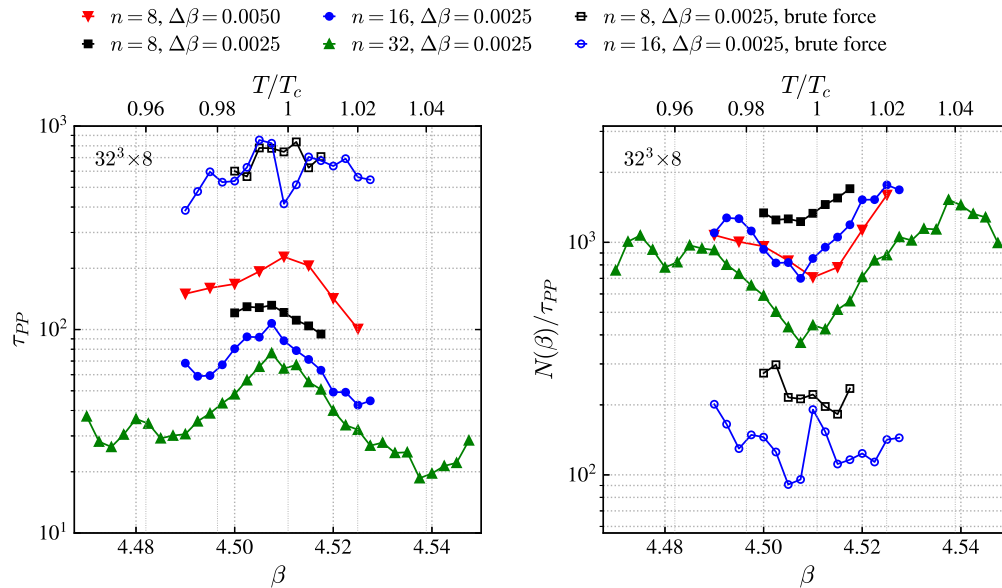


FIG. 3. (Left) The autocorrelation time of the Polyakov loop as a function of β for simulations on a $32^3 \times 8$ lattice computed using the tempering algorithm (solid points) and using the brute force method (unfilled points) around the transition temperature for $N_t = 8$ with equal amounts of computer time. The number of β values and their spacing, $\Delta\beta$, are given for each simulation. τ_{PP} is improved by about an order of magnitude due to the stream swapping of the tempering algorithm compared to that of the brute force approach. τ_{PP} also improves greatly as the number of β values increases (increasing the number of participating streams at any given β) and as $\Delta\beta$ decreases (increasing the probability of any one stream being swapped). (Right) The efficiency of each simulation in terms of the number of computed updates divided by the autocorrelation time.

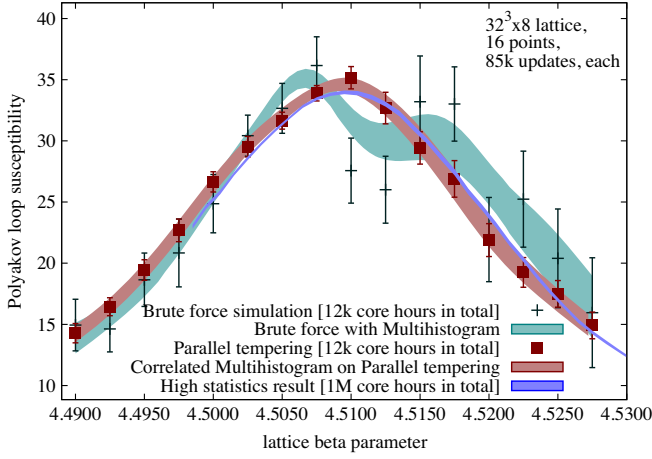


FIG. 4. The susceptibility of the Polyakov loop average Eq. (5) measured on a $32^3 \times 8$ lattice with several different algorithms and analyses as indicated, using 12k core hours of computer time. One local update step consists of three complete sweeps of overrelaxation and one heat bath step at each site and each of the three SU(2) subgroups. On the $32^3 \times 8$ lattice, the full local update completes in 0.64s on a 68-core KNL card. A high statistics result is also shown for comparison.

values, as between the red and blue points, results in nearly the same efficiency, indicating that the decrease in the number of updates (due to having twice as many streams for the same computer time) scales roughly with the decrease in the autocorrelation time.

Finally, in Fig. 4, we show one of the observables of interest: the Polyakov loop susceptibility. We show results from parallel tempering and brute force simulations using the same amount of computational resources. For comparison, we also show the “correct” result we calculated using much larger statistics. We observe that the parallel tempered simulations give much smaller statistical errors. Although the brute force simulations are normally analyzed with the multihistogram method [34], we see that the parallel tempered simulations are superior to that as well. Moreover, one can also use the multihistogram method for the results of the parallel tempering algorithm, for a slight reduction of its errors, as most of the benefit of information of nearby ensembles is gained already using the swapping updates. As the parallel tempering introduces correlations between ensembles at different β s, the multihistogram analysis becomes slightly more complicated (see Appendix A). In fact, the multihistogram algorithm is based on a linear combination of the reweighted results of all the available ensembles to minimize statistical errors. Therefore, if we use the uncorrelated multihistogram procedure on the parallel tempered ensembles, we might get slightly larger error bars, as we use a suboptimal linear combination. In practice, the correlations between ensembles are quite small; neglecting these correlations in the multihistogram procedure causes a negligible change in the results (much smaller than the statistical errors).

To summarize, we find the usage of the parallel tempering algorithm is very beneficial in these investigations. A multihistogram analysis helps to further reduce the statistical errors only slightly, as most of the effect of the swap updates is present already in the single β ensembles.

III. TRANSITION TEMPERATURE

A well-defined transition temperature can be established using a variety of observables that are moments of the order parameter, which for pure SU(3) theory, is the Polyakov loop P . In the analysis of this section, we extract the transition temperature using two of these observables: the susceptibility of the Polyakov loop,

$$\chi = N_s^3 (\langle |P|^2 \rangle - \langle |P| \rangle^2), \quad (5)$$

and the third order Binder cumulant of the Polyakov loop,

$$b_3 = \frac{\langle |P|^3 \rangle - 3\langle |P| \rangle \langle |P|^2 \rangle + 2\langle |P| \rangle^3}{(\langle |P|^2 \rangle - \langle |P| \rangle^2)^{3/2}}. \quad (6)$$

At finite volume, the susceptibility as a function of the coupling β shows a peak around the transition temperature. The value of the coupling at which the peak occurs, which we call β_χ , approaches a value β_c in the infinite volume limit as the height of peak diverges linearly with the volume, and the width of the peak goes to zero linearly with the inverse volume; this value corresponds to the transition temperature. b_3 is used in an analogous fashion: As an odd central moment of the order parameter, instead of a peak, b_3 has a zero crossing at a particular coupling value, which we call β_b . The slope of b_3 around the zero crossing increases linearly with the volume. In the infinite volume limit, the zero crossing of b_3 becomes a jump discontinuity at the phase transition, and β_b approaches the same value of the critical couplings as does β_χ .

If one wants to compare results at different lattice spacings, the susceptibility defined in Eq. (5) should be renormalized when one extracts β_χ [47]. The zero crossing of b_3 is clearly unaffected by scaling, so b_3 does not need to be renormalized when extracting β_b . For every lattice, we renormalize the bare susceptibility as

$$\chi_R(\beta, N_s, N_t) = Z^{N_t}(\beta) \chi(\beta, N_s, N_t), \quad (7)$$

where N_t is the temporal extension of that lattice, with the renormalization condition that

$$\chi_R(\beta_b, 4N_t, N_t) = Z^{N_t}(\beta_b) \chi(\beta_b, 4N_t, N_t) = 1; \quad (8)$$

i.e., we fix $\chi_R(\beta_b) = 1$ for all lattices with aspect ratio $LT = N_s/N_t = 4$. β_b is used in the renormalization condition because, as just mentioned, it is a scale-invariant quantity. Because the normalization function $Z(\beta)$ is itself β -dependent, the location of the peak of $\chi_R(\beta)$ differs

slightly from that of $\chi(\beta)$; hence, β_χ is dependent on the renormalization. Obviously, it is not strictly necessary to renormalize the susceptibility in order to extract β_c . As the width of the bare peak decreases linearly with the inverse volume, $Z(\beta)$ becomes approximately constant across that width, and so the difference between β_χ of $\chi(\beta)$ and β_χ of $\chi_R(\beta)$ should go to zero in the infinite volume limit. Thus, the value of β_c extracted from $\chi(\beta)$ should agree with the value of β_c extracted from $\chi_R(\beta)$. However, that our results may be formally correct and to avoid any potential ambiguities, we use the renormalized susceptibility in this analysis.

We extract β_χ and β_b by fitting $\chi_R(\beta)$ and $b_3(\beta)$ in the vicinity of the peak and of the zero crossing, respectively. We employ basic polynomial fits of degrees 3, 4, 5, and 6. As a consequence of the stream swapping in the tempering algorithm, correlations exist in the Polyakov loop data computed at different values of β for a given lattice (see Sec. II for details); thus, correlations also exist among each moment of the Polyakov loop computed at different β for each lattice. Therefore, we use the well-established practice of fitting with correlations using a subspace of the covariance matrix. The essential details are as follows. When fitting either $\chi_R(\beta)$ or $b_3(\beta)$ for a particular lattice, the covariance matrix of the fitted quantity is computed and diagonalized. In the diagonalized matrix, the smallest eigenvalues are averaged over and replaced with that average. After the averaging, the covariance matrix is converted back to its original basis. The cutoff for what counts as a small eigenvalue is freely chosen. We consider three choices for the cutoff: 10%, 5%, and 1% of the largest eigenvalue. For completeness and comparison, we also consider an uncorrelated fit. Ultimately, whether one uses a correlated fit or no correlation at all, we find that it makes very little difference on the final result for the transition temperature (see the error budget). In order to be able to interpret the χ^2 of the fit to $\chi_R(\beta)$ or $b_3(\beta)$, one can use a correlated fit. The final result for the critical couplings including combined systematic and statistical errors are included in Appendix B.

To extrapolate β_χ and β_b to the continuum and infinite volume limits to get β_c , we use the following set of lattice parameters:

N_t	N_s
5	15, 20, 22*, 23*, 25, 30, 40
6	18, 21, 24, 27, 30, 36, 48
7	28
8	24, 28, 32, 36, 40, 48, 64
10	30, 35, 40, 45, 50, 60, 80
12	48

The volumes are chosen such that there are lattices with aspect ratios $LT = 3, 4, 4.5, 5, 6$, and 8 , for $N_t = 5, 6, 8$, and 10 (to obtain data at $LT = 4.5$ for $N_t = 5$, we find β_χ and β_b for the lattices $22^3 \times 5$ and $23^3 \times 5$ and then

interpolate linearly in $1/N_s^3$ to get β_χ and β_b for a hypothetical “ $22.5^3 \times 5$ ” lattice) and that there are lattices with $LT = 3.5$ for $N_t = 6, 8$, and 10 . Two additional lattices of $LT = 4$ are also chosen for $N_t = 7$ and 12 .

Now, the value of the coupling β is specific to the choice of action on the lattice; hence, while β is useful for comparing results computed using the same action, it must be translated into a physically meaningful quantity, the lattice temperature T , to yield results that are independent of the choice of action. This requires the use of a scale setting. We use the w_0 scale based on the Wilson flow, implicitly defined by

$$t \frac{d}{dt} [t^2 E(t)]_{t=w_0^2} = 0.3, \quad (9)$$

where $E(t)$ is the expectation value of the gauge action of lattice configurations evolved via the Wilson flow [48,49]. We compute the w_0 scale in lattice units, w_0/a , for many values of β for two different discretisations of the flow (WSC and SSC in the notation of [50]). This gives us another systematic choice as to which version of the scale setting to use. We then interpolate these results to get $w_0/a(\beta)$ by fitting with a polynomial of order 6 and 7 in the β range [4.0,4.95]. It is critical to ensure that finite-volume effects on the w_0 scale remain small, as these effects increase with the flow time [48]. In Fig. 5, we present a volume study of the w_0 scale (WSC) computed in this analysis. Fluctuations in w_0/a are comparable in size to the statistical error at each volume; consequently, no significant volume dependence can be seen.

Once $w_0/a(\beta)$ is found for a particular β on a lattice with a temporal extension of N_t , it can be converted into the lattice temperature $T = (N_t a)^{-1}$ by dividing by N_t :

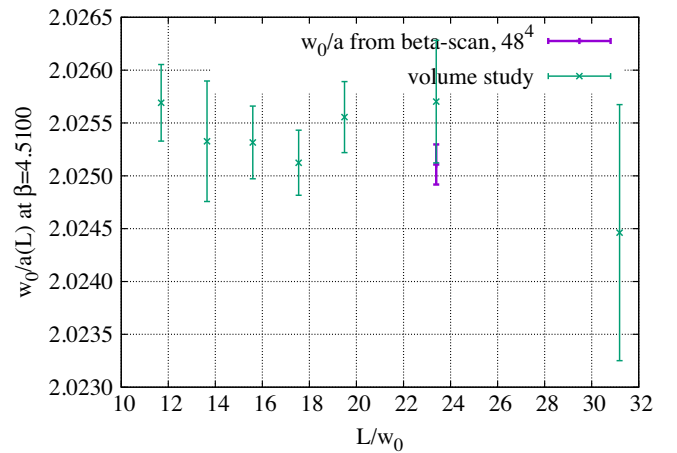


FIG. 5. Volume dependence of the w_0 scale. w_0/a is computed using fixed bare quark masses that correspond to the physical point and is shown as a function of the spatial extent of the box in lattice units L/w_0 . Fluctuations are comparable to the statistical error at each volume in the study; hence, finite-volume effects remain small.

$$\frac{w_0}{aN_t} = w_0 T. \quad (10)$$

In this way, β_χ and β_b are converted into the dimensionless quantity $w_0 T_c$ for each lattice.

To compute the continuum value of the transition temperature $w_0 T_c$ at infinite volume, we first find the continuum value of $w_0 T_c$ at finite volume and fixed lattice aspect ratio $LT = N_s/N_t$ and then extrapolate the results of the continuum theory at different LT to the limit where $(LT)^3 \rightarrow \infty$. That a continuum limit for $w_0 T_c$ exists for pure SU(3) at finite volume is not immediately obvious; however, as shown in Fig. 6, a well-defined continuum value of $w_0 T_c$ can in fact be established at fixed aspect ratio LT using both definitions for β_c . We use large lattices with aspect ratios $LT = 4, 4.5, 5, 6,$ and 8 for four different temporal extensions: $N_t = 5, 6, 8,$ and 10 . To check the consistency of the continuum extrapolation at other temporal extensions, we also include two lattices with $N_t = 7$ and 12 in the $LT = 4$ extrapolation. In the left panel of Fig. 6, one can see that $w_0 T_c$ goes roughly linearly with N_t^{-2} for each LT . We perform two types of fit to extrapolate to the continuum: a linear fit to all lattices with $N_t \geq 6$, and a quadratic fit to all lattices with $N_t \geq 5$. The fits are shown in Fig. 6; the error bands give the combined statistical and systematic errors from all the similar fits performed in the analysis.

The continuum results from χ_R and b_3 are shown in Fig. 7 as functions of $(LT)^{-3}$, which behaves as the inverse volume. A quadratic fit to $LT \geq 4$ is used to extrapolate to the infinite volume limit. This fit is shown in Fig. 7 for both χ_R and b_3 . The error bands give the combined statistical and systematic errors from all the similar fits that were performed. We find that the infinite volume value of the transition temperature $w_0 T_c$ computed using χ_R agrees with the value computed using b_3 , as expected.

Following the analysis method introduced in [51] to estimate the statistical and systematic uncertainties of the results, in total, we have performed 256 different analyses; these are characterized by the choice of the moment of the Polyakov loop (χ_R or b_3), the degree of the fit to the moment (3, 4, 5, or 6), whether and how the fit is correlated (uncorrelated or correlated with an eigenvalue cutoff of 10%, 5%, 1%), the choice of the w_0 scale calculation (WSC, SSC), the degree of fit to the w_0 scale data (6 or 7), and the degree and range of the continuum extrapolation (linear or quadratic). The cumulative distribution function (CDF) of the infinite volume results of these analyses is shown in Fig. 8 in red, where the central value and statistical error of each result are given. The statistical and systematic uncertainties are synthesized into a smooth CDF defined by the equation,

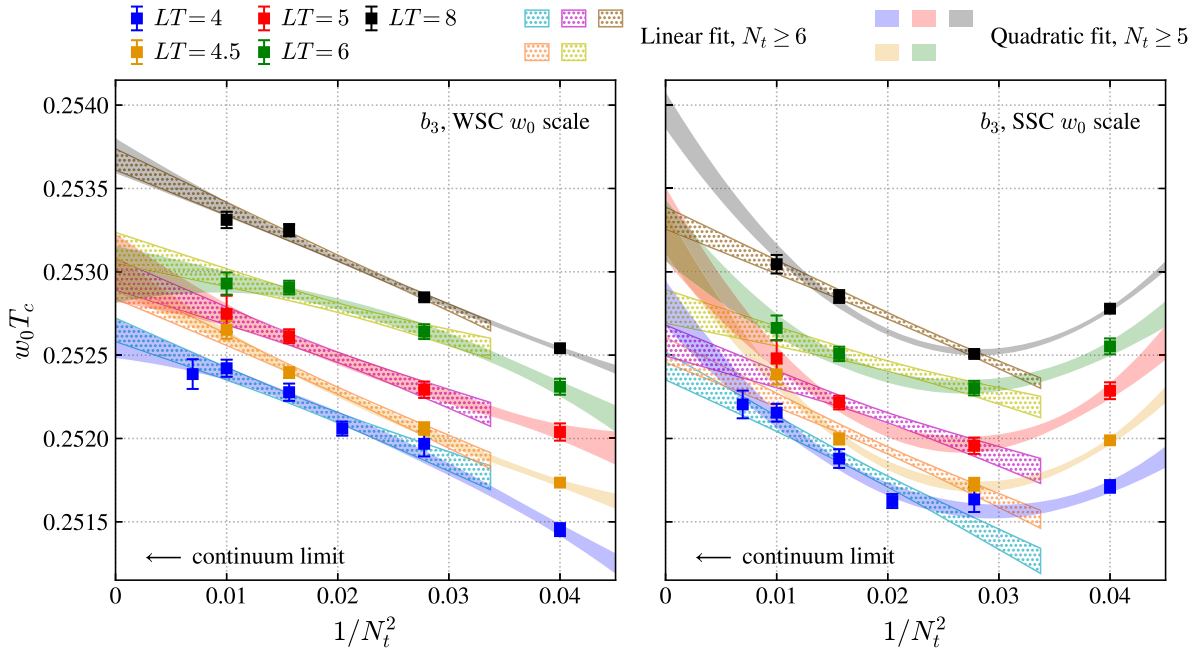


FIG. 6. Two example plots that summarize the continuum extrapolation of the transition temperature $w_0 T_c$ computed from the zero crossing of b_3 at five fixed lattice aspect ratios LT ; the zero crossing is converted into the quantity $w_0 T_c$ using (left) the WSC w_0 scale and (right) the SSC w_0 scale. Each data point is the median value of $w_0 T_c$ computed for that lattice from the systematic analysis, and the error bars give the combined statistical and systematic error. Error bands are shown for the two kinds of extrapolating fit that are used: a linear fit to lattices with $N_t \geq 6$ and a quadratic fit to lattices with $N_t \geq 5$. The error bands give the combined statistical and systematic error from all similar fits performed in the analysis.

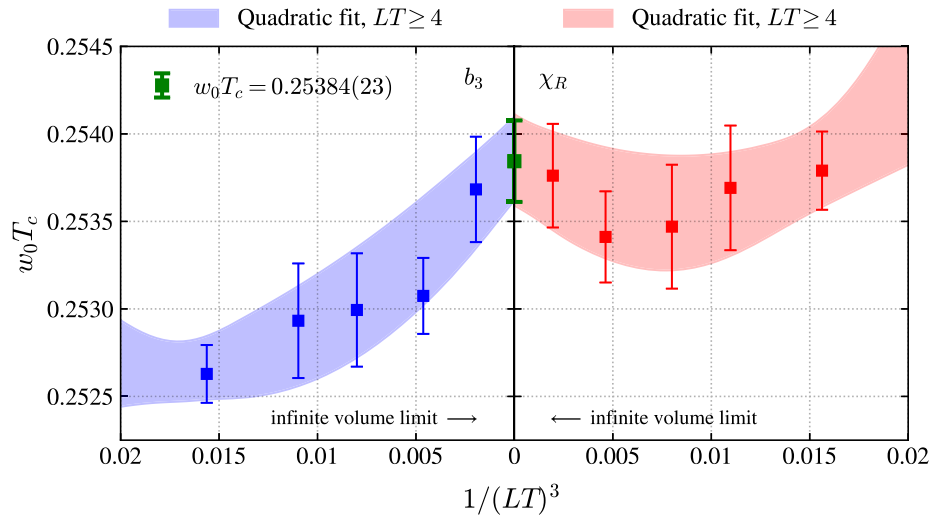


FIG. 7. A summary plot of the infinite volume extrapolation of the continuum transition temperature $w_0 T_c$ computed (left) from the zero crossing of b_3 and (right) from the peak of χ_R . Each point is the median value of $w_0 T_c$ computed for each aspect ratio LT in the systematic analysis, and the error bars give the combined statistical and systematic error. Error bands are shown for the quadratic fit to $LT \geq 4$. The error bands give the combined statistical and systematic errors from all similar fits in the analysis. The extrapolations from b_3 and χ_R are in agreement, as expected. The final result, $w_0 T_c = 0.25384(23)$, is shown in green.

$$F(w_0 T_c) = \frac{1}{2} + \frac{1}{2 \cdot 256} \sum_{i=1}^{256} w_i \operatorname{erf} \left(\frac{w_0 T_c - \mu_i}{\sqrt{2} \sigma_i} \right), \quad (11)$$

where μ_i and σ_i are the central value and statistical uncertainty, respectively, of the i th analysis, and w_i is the weight of the CDF of the i th result. (This equation then assumes that the statistical result of each analysis is well described by a normal distribution with a mean of μ_i and standard deviation of σ_i .) We have no prior assumptions as to the relative statistical significance of any of the analyses, and so we take an agnostic position by weighting all analyses equally, i.e., by setting $w_i = 1$. The result is shown in blue in Fig. 8. The final result is then the median value of

the smooth CDF, implicitly defined by $F(w_0 T_c) = 0.5$, and the central 68% width is taken as twice the total error. This yields a final value of $w_0 T_c = 0.25384(23)$, which is shown in green in Fig. 7, with an error budget of

Median	0.25384	
Statistical error	0.00011	0.043%
Full systematic error	0.00021	0.082%
Observable (b_3, χ_R)	1.1×10^{-5}	0.0042%
Fit order (3, 4, 5, 6)	2.1×10^{-5}	0.0085%
Fit type (corr, uncorr)	1.2×10^{-5}	0.0047%
Scale setting (WSC, SSC)	1.9×10^{-5}	0.0075%
Scale fit order (6, 7)	2.4×10^{-6}	0.0010%
Continuum limit range	1.2×10^{-4}	0.0487%

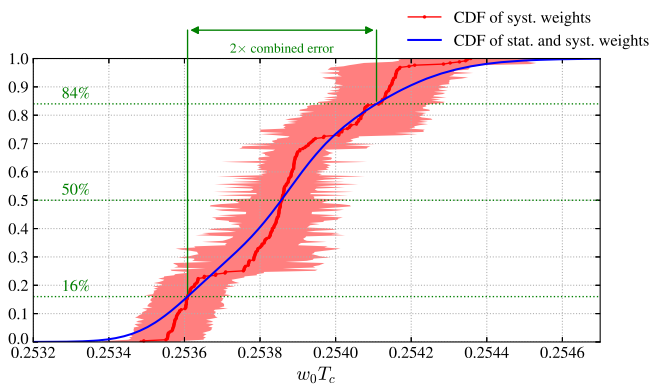


FIG. 8. The cumulative distribution function (CDF) of all the results from the systematic analysis of the infinite volume limit of $w_0 T_c$ in the continuum. The red curve shows the central value for each analysis, and the error band shows the statistical error. The blue curve gives the synthesized CDF defined in Eq. (11), combining the statistical and systematic errors. The final result is $w_0 T_c = 0.25384(23)$.

IV. EVIDENCE FOR A FIRST ORDER TRANSITION

In this section, we use our simulations to demonstrate that the thermodynamic transition of the SU(3) Yang-Mills theory is first order, as anticipated. We can perform this in two ways.

First, we study the finite volume scaling of the susceptibility (scaled variance) of the order parameter. As we did in the case of the transition temperature study, we again use the susceptibility (χ_R) of the renormalized Polyakov loop. If the transition is of first order, the leading volume dependence must be $\chi_R^{-1} \sim V^{-1}$. Such a study would not be new and also would not give a true evidence, if we used data from a fixed lattice resolution. Here, we first calculate the continuum limit of the peak height for various volumes and then demonstrate the expected volume scaling. Moreover, we show the continuum extrapolation of χ_R

as a function of w_0T around the critical value w_0T_c , and display the expected volume dependence.

Perhaps the most convincing argument for the first order nature of a transition is the existence of a latent heat in the thermodynamic and continuum limit. Here, we are building on the work of the WHOT collaboration [44,52]. Aided by the tempering algorithm and the improved action, we arrive at a combined continuum and infinite volume limit, 29 standard deviations away from zero.

A. Finite volume scaling of the Polyakov loop susceptibility

To obtain the finite volume scaling of a susceptibility, it has to be continuum extrapolated for several fixed physical volumes. This raises the question of renormalization, a topic that one could ignore as long as only one lattice resolution is relied upon. As discussed before, $\chi_R(L, T) = Z^{N_t}(\beta)\chi(\beta, N_x, N_t)$ with $L = a(\beta)N_x$ and $T^{-1} = a(\beta)N_t$. The renormalization condition for Z is defined at a fixed volume; for us, the choice is $LT = 4$ (meaning $N_x = 4N_t$). Scanning through the resolutions $N_t = 4, 5, 6, 7, 8, 10, 12$ and setting the renormalization condition $\chi_R(L = 4T_c^{-1}, T = T_c) \equiv 1$, one can easily obtain $Z(\beta)$ for the whole range of interest. Clearly, in this setting, we lose any information about an overall constant in χ_R , but this being a volume-independent factor, the actual finite size scaling is left intact.

We select fixed physical volumes ($LT = 4.5, 5, 6, 8$) and interpolate $Z(\beta)$ to the actual volume dependent β_c of each simulation with $N_t = 5, 6, 8$ and 10. The renormalized result $\chi_R(L, T) = Z^{N_t}(\beta)\chi(\beta, N_x, N_t)$ is simply continuum

extrapolated as N_t^2 . Such continuum extrapolation is shown in Fig. 9 for the $\chi_R(w_0T)$ curves at fixed volume and in Fig. 10 for the inverse peak height. In Fig. 9 (left), each curve is continuum extrapolated at the indicated fixed volume, and the bands include both statistic and systematic uncertainties. Note that had we subtracted the infinite volume limit value $\chi_R(w_0T_c)$ instead of the volume-dependent peak values $\chi_R(w_0T_{c,\chi}(LT))$, the curves would not have been exactly centered around zero, but a trend would have appeared where larger volumes would peak closer to zero. However, as can be seen in the right panel of Fig. 7, the difference would be negligible since the continuum extrapolated peak position shows a very mild volume dependence. The right panel shows the same curves, but scaled with the volume in order to highlight the volume dependence. We can see that the bands corresponding to $LT = 5, 6, 8$ are overall almost indistinguishable, clearly showing that they fall in the linear volume scaling regime typical of a first order transition.

In Fig. 10, we actually show two extrapolations, one for each definition of T_c , using the zero of the third Binder cumulant or the susceptibility peak in red and blue, respectively. The extrapolations do not differ at all depending on this ambiguity, and the main systematic uncertainty comes from the possibility to include a inverse quadratic volume term in the infinite volume extrapolations. We find that the inverse susceptibility is extrapolated to be vanishing, [or, actually, strongly constrained: $\chi_R^{-1}(V = \infty) = 0.0023(58)_{\text{stat}}(65)_{\text{sys}}$] in the infinite volume limit. (A crossover transition is signaled by a positive value of the infinite volume extrapolated inverse susceptibility).

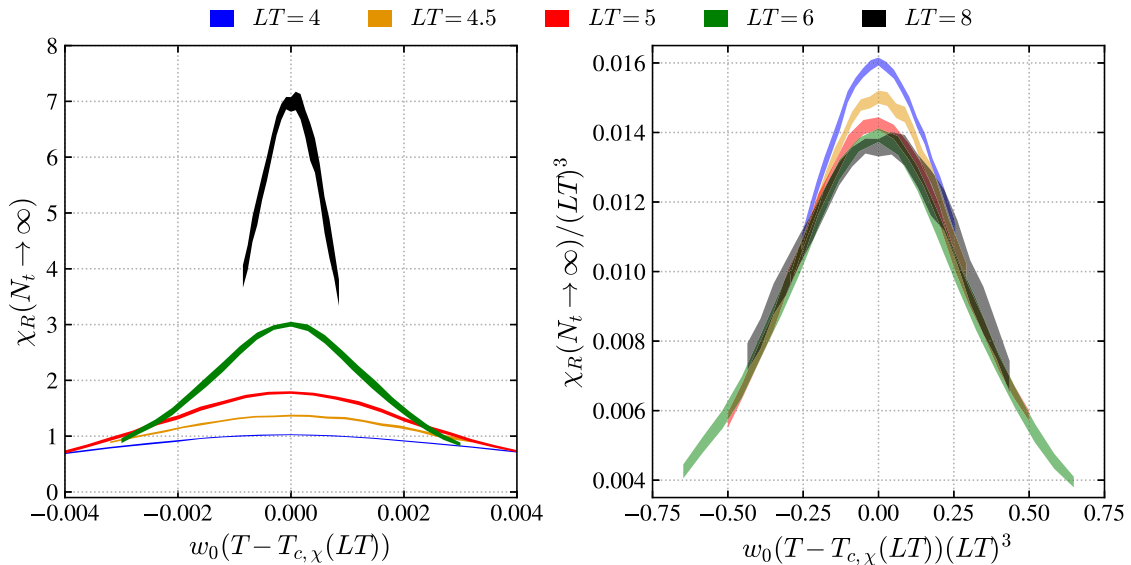


FIG. 9. (Left) The continuum extrapolated renormalized Polyakov loop susceptibility at different volumes, centered around the volume-dependent peak position $w_0T_{c,\chi}(LT)$. The bands include both statistic and systematic uncertainties. (Right) Same curves as in the left panel, but scaled with appropriate powers of the volume, in order to highlight the linear volume dependence typical of a first order transition.

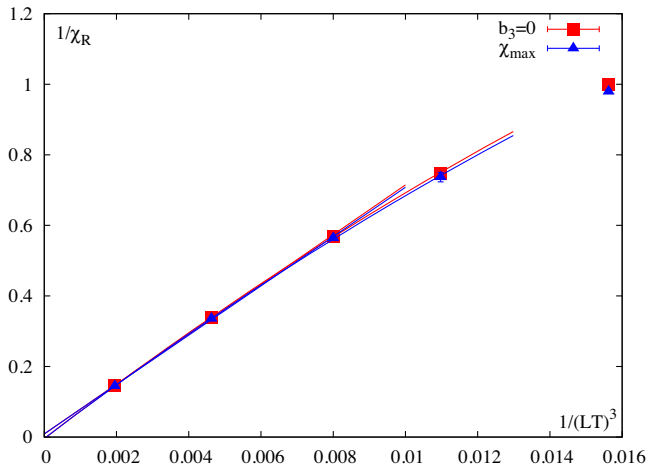


FIG. 10. Infinite volume extrapolation of the inverse Polyakov loop susceptibility. The results are renormalized and continuum extrapolated, systematic errors are included. The renormalization condition was $\chi_R(L = 4T_c^{-1}, T_c) = 1$. We use red symbols if we defined T_c through the zero of the third Binder cumulant and blue symbols when we defined T_c as the peak of the susceptibility. Depending on the details of the infinite volume extrapolation, we get a slightly positive or negative result. All in all, we get the result $\chi_R^{-1}(V = \infty) = 0.0023(58)_{\text{stat}}(65)_{\text{sys}}$. The linear scaling sets in for $LT \geq 5$. The linear convergence of χ_R^{-1} to zero indicates a first order transition.

B. Calculation of the latent heat

A nonvanishing latent heat identifies a transition to be first order beyond doubt. Its calculation, however, is much less trivial than observing a transition peak in a susceptibility. Most conveniently, the latent heat can be understood as the discontinuity of the trace anomaly, $(\epsilon - 3p)/T^4$, as the transition is crossed. This assumes a continuous pressure, which is assumed in all transition scenarios. The difficulty in quantifying this gap arises from two sources:

- To calculate a discontinuity, one must get very close to T_c from both sides. The trace anomaly happens to have a pronounced temperature dependence just above and just below T_c .
- The trace anomaly has a quartic divergence; thus, brute force statistics must scale with $\sim N_t^8$ to compensate. Even in quenched QCD, this can quickly become the driving factor in the costs for $N_t \geq 8$.

When the continuum limit is desired, point (b) determines the amount of statistics needed, but point (a) must be addressed in all cases. Just “reading off the jump” could give any value between zero and 2 for $\Delta(\epsilon - 3p)/T_c$, as shown in Fig. 1 of our earlier study [53]. One could define the latent heat as the surface under the peak of the energy susceptibility; however, the signal-to-noise ratio is diminishing in the infinite volume limit.

The most successful method to define the latent heat, the one we also use here, is to simulate right at β_c and classify

the lattice configurations into the cold and hot phases (see Refs. [44,52]). The trace anomaly can then be evaluated for both the hot and cold halves of the ensemble; the difference gives the latent heat up to a renormalization factor. The absolute value of the Polyakov loop shows a clear double peaked histogram. Selecting the Polyakov loop magnitude at the minimum of this histogram between the two peaks as the cut value, we introduce a phase categorization that has a systematic error decreasing exponentially with growing volume. This criterion works well on smeared (flowed) configurations as well [44].

We illustrate this approach on our $N_t = 5$ ensembles at β_c . We reweighted our closest β ensemble to the point where $b_3(\beta) = 0$ for the particular simulation volume. The bare Polyakov loop histograms are shown in Fig. 11 for three of our simulation volumes. We show the standard jackknife errors on the histograms. For this lattice spacing, we could reach an aspect ratio of $LT = 16$, where the separation of the two peaks are very clear, indeed. Yet, even near the minimum, we have enough hits in the Polyakov loop bins to quantify the trace anomaly bin by bin.

The trace anomaly shows no discontinuity between the phases, and its Polyakov loop dependence can be modeled by a polynomial. In Fig. 11, we fit a third order polynomial with finite volume corrections proportional to $1/L^3$ to obtain the infinite volume extrapolation (red curve). This picture connects the latent heat with the hot phase value of Polyakov loop. The cold phase peak moves to zero as $1/L^3$, hence, the y -intercept of the red curve points to the trace anomaly at $T_c - \epsilon$. Similarly, the nontrivial position of second histogram peak in the thermodynamic limit is translated by the same curve to the trace anomaly at $T_c + \epsilon$.

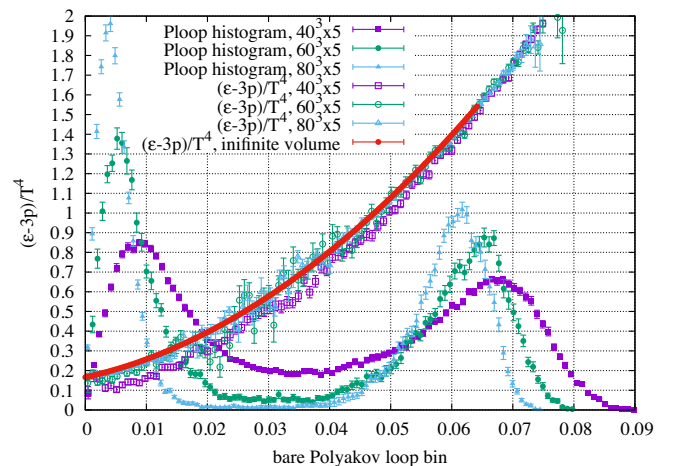


FIG. 11. We show the Polyakov loop histograms for three volumes ($N_t = 5$) along with the trace anomaly expectation values for each Polyakov loop bin. The double peak structure and the sharpening of the peaks with increasing volume match our expectations for a first order transition. The trace anomaly shows no discontinuity in this representation, and the infinite volume limit can be found using a polynomial model.

As observed in Fig. 11, one may define the latent heat at a fixed ensemble by translating the peak positions into a pair of trace anomaly values through the red curve (the infinite volume extrapolation of the trace anomaly as function of the magnitude of the Polyakov loop) and then taking the difference. While this is clearly possible, it is to be seen in practice if the volume scaling turns linear in $1/L^3$ for practical lattice sizes. Alas, this is not the case. Both the peak positions and the trace anomaly they relate to, are heavily nonlinear in the range $4 \leq LT \leq 8$, where most of our statistics is collected. However, integrating the trace anomaly curve with the histogram peak's weight for both peaks gives a linear result. This is then equivalent to the approach in Ref. [52] where the trace anomalies were simply averaged for both cold and hot subensembles. In summary, we first define a cut value in the modulus of the Polyakov loop for each lattice size individually as the minimum position of the discussed histogram. These bins play no role in the second step, where we calculate the trace anomaly for the configurations above and below this cut, separately. The trace anomaly difference between these two subensembles gives the latent heat normalized to T^4 .

Let us now come to point (b). Reference [44] introduces a truly innovative way to separate the UV noise of the physics result in the trace anomaly. The results are calculated at small gradient flow time, where the square root of the flow time replaces the lattice spacing in its role as a cutoff scale. The presence of a small, physical flow time eliminates the quartic divergence, and the continuum limit can be taken. Using larger flow time allows the usage of coarser lattices. This then needs to be related to the real theory at zero flow time using small flow time expansion [54], which gets harder as the selected flow time increases.

In this study, we analyzed the configurations at zero flow time directly. To mitigate the quartic divergence, we used the tree-level Symanzik improved action. This choice does not improve the quartic scaling of the simulation noise, yet it endorses coarser lattices within the linear range continuum scaling. We use lattices up to $N_t = 10$ instead of the $N_t = 16$ of Ref. [44]; this immediately allows for simulating a factor $43 \approx (16/10)^8$ less statistics due to the quartic divergence of the trace anomaly. (This factor must be considered together with the fact that the improved action comes at a 2.1 times higher cost [55].)

The trace anomaly in Fig. 11 was renormalized using the standard scheme of Ref. [56],

$$\frac{\epsilon - 3p}{T^4} = N_t^4 a \frac{d\beta}{da} (\langle S_g \rangle_{N_s^3 \times N_t} - \langle S_g \rangle_{\text{infinite } N_t}), \quad (12)$$

where S_g is the gauge action without the factor of the inverse coupling β . The logarithmic derivative of the β —lattice spacing (a) relation was obtained using our w_0 scan, the same that was already used as high precision scale setting in Sec. III. The same runs were used to calculate the

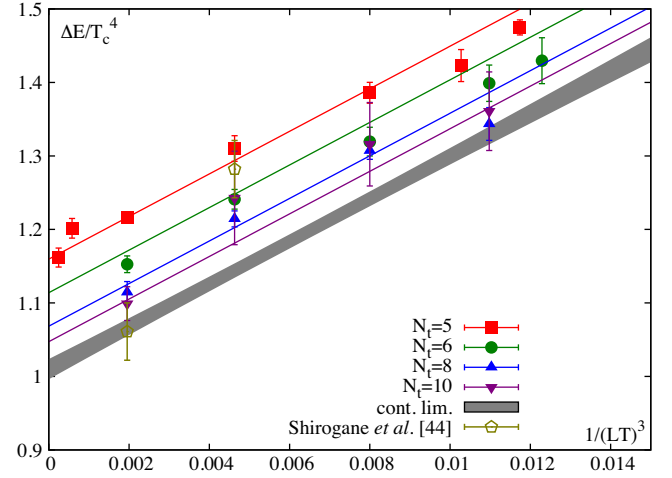


FIG. 12. Combined and continuum limit of the trace anomaly difference between the hot and cold phases. The extrapolation is linear both in $1/N_t^2$ and $1/L^3$; in this example, we exclude all ensembles with $N_t < 6$ or $LT < 5$. The continuum extrapolated but finite volume results in Ref. [44] are shown for comparison.

zero temperature gauge action $\langle S_g \rangle_{\text{infinite } N_t}$, which is not necessary for latent heat itself, only if one is interested in trace anomalies at $T_c \pm \epsilon$.

We calculated the trace anomaly differences for the ensembles that we generated. We find that a combined linear continuum and infinite volume fit is possible, and thus, we get the result in Fig. 12.

For the final results, we consider the following sources of systematic uncertainties. The continuum limit is calculated both using and not using $N_t = 5$, including or not including the aspect ratio $LT = 4.5$ into the infinite volume extrapolation. We use various number of Polyakov bins (100, 200 or 400) and two different log-polynomial models for fitting the negative logarithm of the histogram so that cut value separating the phases can be calculated. We have performed two separate analyses: first using reweighting to β_c from the closest available β ensemble, second, using the multihistogram method to calculate averages at β_c . In the end, we obtain

$$\Delta \left[\frac{\epsilon - 3p}{T^4} \right] = 1.025(21)_{(\text{stat})}(27)_{(\text{sys})}, \quad (13)$$

with an error budget of

	Median	1.0249	
Statistical error	0.021		2.1%
Full systematic error	0.027		2.7%
Histogram fitting	0.0030		0.29%
Histogram binning	0.0002		0.02%
β_c definition	0.0135		1.32%
w_0 interpolation	0.0007		0.07%
LT range	0.0121		1.18%
N_t range	0.0157		1.53%
Analysis method	0.0019		0.18%

V. CONCLUSION

In this paper, we have studied the phase transition in an SU(3) pure gauge theory. First we have investigated the efficiency of parallel tempering simulations to measure properties of the system near the phase transition point. The parallel tempering allows one to swap updates between simultaneous calculations at different temperatures. One observes a sizable decrease of the autocorrelation times compared to simulations at a fixed temperature. As observed the efficiency of parallel tempering is superior to even brute force simulations using multihistogram analysis to combine the data of all the ensembles.

Next we have determined the critical temperature of the theory. We used two definitions to find the transition point: the peak of the Polyakov loop susceptibility and the zero of the third Binder cumulant of the Polyakov loop. The ensembles in this study consisted of lattices with temporal lattice size $N_t = 5, 6, 8, 10$, and 12 with aspect ratios $LT = 4, 4.5, 6, 8, 10, 12$, using the Symanzik gauge action. First, we carried out a continuum extrapolations for fixed aspect ratios, giving the continuum extrapolated critical temperatures for finite physical volumes. Second, we calculated the infinite volume extrapolated critical temperatures and observed the coincidence of the two definitions extrapolated to the thermodynamic limit in the continuum, as expected. The systematic errors of the calculation were thoroughly investigated by exploring many possible fitting and extrapolation procedures. Our final result for the critical temperature with combined statistical and systematic error is $w_0 T_c = 0.25384(23)$.

Finally, we have investigated the properties of the phase transition. We have confirmed the first order nature of the transition by first investigating the finite volume scaling of the peak of the Polyakov loop susceptibility. For this, we have calculated the continuum limit of the renormalized Polyakov loop susceptibility for different physical system sizes. As one observes, both the height and the width of the susceptibility follow the appropriate scaling in the infinite volume limit, confirming that the phase transition is first order. Second, we have calculated the latent heat of the transition, with the help of the discontinuity of the trace anomaly. The discontinuity was measured for each ensemble by averaging the trace anomaly at the critical coupling for “hot” and “cold” configurations, where the classification is based on the value of the Polyakov loop. Finally, we carried out a combined infinite volume and continuum extrapolation, yielding the value $\Delta\epsilon/T_c^4 = 1.025(21)(27)$.

The tempering algorithm is not limited to quenched QCD. The same first order transition and the corresponding critical end point can be addressed using tempering updates in gauge coupling in the heavy quark region [18]. Introducing pseudofermions, the tempering updates can be extended to mass, imaginary chemical potential, or other parameters of the fermionic action. This algorithm will

probably play an important role in the exploration of the QCD phase diagram.

ACKNOWLEDGMENTS

The project received support from the BMBF Grant No. 05P21PXFCA. The authors gratefully acknowledge the Gauss Centre for Supercomputing e.V. [57] for funding this project by providing computing time on the GCS Supercomputer HAWK at HLRS, Stuttgart. Part of the computation was performed on the QPACE3 funded by the DFG and hosted by JSC and on the cluster at the University of Graz. S. B. thanks Julius Kuti for discussions on the topic.

APPENDIX A: CORRELATED MULTIHISTOGRAM

Here, we adopt the well-known reweighting method by Ferrenberg and Swendsen [34] and extend it to correlated data.

The weighted average a of data sets x_i , $i = 1, 2, \dots, m$ can be calculated by minimizing

$$\chi^2 = \sum_{i,j}^m (x_i - a)(C^{-1})_{i,j}(x_j - a), \quad (\text{A1})$$

with C^{-1} being the inverse covariance matrix [58]. The minimum is found for

$$a = \frac{\sum_{i,j=1}^m (C^{-1})_{i,j} x_j}{\sum_{i,j=1}^m (C^{-1})_{i,j}}, \quad (\text{A2})$$

which has minimal fluctuations compared to other possible linear combinations.

Interpolating observables such as the Polyakov loop to a target coupling β demands a precise knowledge of the partition function $Z(\beta)$ and thus, of the density of states W ,

$$Z(\beta) = \sum_E W(E) e^{-\beta E}. \quad (\text{A3})$$

Since $W(E)$ does not depend on β , it can be estimated by the densities of states of the individual simulations $W_i(E)$, according to Eq. (A2). Thus, i labels the simulation at β_i , which is correlated with the other ones due to the β tempering algorithm. The histograms of the energy $N_i(E)$ can be related to the density of states by

$$W_i(E) \frac{e^{-\beta_i E}}{Z(\beta_i)} = \frac{N_i(E)}{n_i}, \quad (\text{A4})$$

where n_i is the total amount of entries of the histogram. In the same manner, the exact density of states $W(E)$ is estimated by multiple simulations performed at the same coupling,

$$W(E) = \frac{\overline{N}_i(E)Z(\beta_i)}{n_i e^{-\beta_i E}}. \quad (\text{A5})$$

$$\text{cov}(W_i, W_j) = W^2 \frac{\text{cov}(N_i, N_j)}{\overline{N}_i \overline{N}_j} = W^2 \frac{\text{corr}(N_i, N_j) g_i g_j}{\sigma_i \sigma_j}. \quad (\text{A6})$$

It is important to note that \overline{N}_i is the averaged histogram of these simulations; thus, its error is $\sigma_i = \sqrt{g_i \overline{N}_i}$. The factor g_i takes the correlation time τ_i of the ensemble into account and reads $g_i = 1 + 2\tau_i$. The covariance for a certain energy E between the densities of states of individual simulations W_i and W_j can be calculated as

Here, $\text{cov}(N_i, N_j)$ and $\text{corr}(N_i, N_j)$ stand for components of the covariance and correlation matrix, respectively, which are estimated according to the standard Jackknife resampling method.

With Eq. (A2), the estimator of the exact density of states $W(E)$ can be written as

$$W(E) = \frac{\sum_{i,j=1}^m \sigma_i \sigma_j (g_i g_j)^{-1} \text{corr}^{-1}(N_i, N_j) W_j}{\sum_{i,j=1}^m \sigma_i \sigma_j (g_i g_j)^{-1} \text{corr}^{-1}(N_i, N_j)} = \frac{\sum_{i,j=1}^m (g_i g_j)^{-1/2} \sqrt{\frac{n_i Z_j}{n_j Z_i}} e^{E(\beta_j - \beta_i)/2} N_j(E) \text{corr}^{-1}(N_i, N_j)}{\sum_{i,j=1}^m (g_i g_j)^{-1/2} \sqrt{\frac{n_i n_j}{Z_j Z_i}} e^{-E(\beta_i + \beta_j)/2} \text{corr}^{-1}(N_i, N_j)}. \quad (\text{A7})$$

In some cases, the correlation matrix is singular, caused by bins that only contain a few entries. Since these bins have a very small contribution to the end result of $W(E)$, we neglect the correlation of them with other bins. In the case

of uncorrelated simulations, the correlation matrix is the unity matrix, and we get back to the standard Ferrenberg-Swendsen equation for the density of states,

TABLE I. Critical couplings calculated using the fitting method explained in Sec. III and by the multihistogram analysis.

Lattice	β_c from b_3 , fit	β_c from b_3 , multihist	β_c from χ^R , fit	β_c from χ , fit	β_c from χ , multihist
15 × 5	4.19667(15)	4.19676(13)	4.20246(73)	4.20019(46)	4.19996(16)
20 × 5	4.19855(9)	4.19858(8)	4.20048(47)	4.19987(14)	4.19998(10)
22 × 5	4.19917(5)	4.19915(6)	4.20050(28)	4.20020(9)	4.20020(6)
23 × 5	4.19926(9)	4.19930(9)	4.20073(13)	4.20040(12)	4.20026(9)
25 × 5	4.19995(11)	4.19996(11)	4.20084(21)	4.20064(16)	4.20069(11)
30 × 5	4.20056(11)	4.20052(10)	4.20103(13)	4.20095(13)	4.20090(10)
40 × 5	4.20112(3)	4.20110(4)	4.20126(6)	4.20123(3)	4.20123(4)
18 × 6	4.30962(39)	4.30986(41)	4.31706(116)	4.31412(97)	4.31324(60)
21 × 6	4.31081(33)	4.31090(29)	4.31470(72)	4.31325(53)	4.31319(33)
24 × 6	4.31210(19)	4.31190(12)	4.31434(28)	4.31355(28)	4.31361(12)
27 × 6	4.31233(9)	4.31230(12)	4.31389(16)	4.31350(14)	4.31347(12)
30 × 6	4.31291(13)	4.31290(15)	4.31402(19)	4.31376(17)	4.31374(16)
36 × 6	4.31379(11)	4.31373(10)	4.31427(17)	4.31415(10)	4.31417(10)
48 × 6	4.31432(3)	4.31432(5)	4.31447(4)	4.31445(3)	4.31447(5)
24 × 8	4.50511(26)	4.50506(24)	4.51309(77)	4.50954(47)	4.50936(38)
28 × 8	4.50633(23)	4.50639(20)	4.51191(46)	4.50980(43)	4.50964(37)
32 × 8	4.50756(14)	4.50758(13)	4.51060(34)	4.50965(19)	4.50964(12)
36 × 8	4.50788(6)	4.50795(7)	4.50985(7)	4.50934(9)	4.50936(8)
40 × 8	4.50846(9)	4.50847(10)	4.50985(16)	4.50951(12)	4.50952(11)
48 × 8	4.50929(10)	4.50930(9)	4.50996(12)	4.50982(11)	4.50982(9)
64 × 8	4.51028(7)	4.51024(7)	4.51048(8)	4.51046(8)	4.51041(7)
30 × 10	4.66673(71)	4.66695(68)	4.67641(231)	4.67065(160)	4.67175(182)
35 × 10	4.66990(24)	4.67002(30)	4.67634(44)	4.67382(35)	4.67362(43)
40 × 10	4.67099(12)	4.67098(14)	4.67421(47)	4.67299(25)	4.67306(20)
45 × 10	4.67166(13)	4.67159(15)	4.67371(22)	4.67309(15)	4.67318(19)
50 × 10	4.67188(30)	4.67199(31)	4.67295(54)	4.67253(52)	4.67267(43)
60 × 10	4.67244(19)	4.67259(21)	4.67326(29)	4.67314(28)	4.67318(25)
80 × 10	4.67365(10)	4.67364(11)	4.67389(12)	4.67386(11)	4.67383(12)
48 × 12	4.80980(17)	4.80980(17)	4.81303(48)	4.81199(20)	4.81205(26)

$$W(E) = \frac{\sum_{i=1}^m g_i^{-1} N_i(E)}{\sum_{i=1}^m g_i^{-1} \frac{n_i}{Z_i} e^{-\beta_i E}}. \quad (\text{A8})$$

Inserting Eq. (A7) in Eq. (A3), the partition functions Z_i of the individual simulations can be calculated iteratively. Assuming constant autocorrelation times τ_i , the g_i factors are the same for every simulation. According to Eqs. (A7) or (A8), they represent a constant factor for all Z_i which can be dropped.

APPENDIX B: CRITICAL COUPLINGS

The critical couplings for the ensembles used in this study are listed in Table I. The quoted errors are combined systematic and statistical errors for the β_c values calculated using fitting as described in Sec. III. Note that the critical couplings calculated using the multihistogram analysis have no systematic error, as they are simply solutions of the equations $b_3(\beta_b) = \partial_{\beta} \chi(\beta_{\chi}) = 0$.

-
- [1] R. Hagedorn, Statistical thermodynamics of strong interactions at high-energies, *Nuovo Cimento Suppl.* **3**, 147 (1965).
- [2] N. Cabibbo and G. Parisi, Exponential hadronic spectrum and quark liberation, *Phys. Lett.* **59B**, 67 (1975).
- [3] A. M. Polyakov, Thermal properties of gauge fields and quark liberation, *Phys. Lett.* **72B**, 477 (1978).
- [4] Y. Aoki, G. Endrodi, Z. Fodor, S. Katz, and K. Szabo, The order of the quantum chromodynamics transition predicted by the standard model of particle physics, *Nature (London)* **443**, 675 (2006).
- [5] B. Svetitsky and L. G. Yaffe, Critical behavior at finite temperature confinement transitions, *Nucl. Phys.* **B210**, 423 (1982).
- [6] L. Yaffe and B. Svetitsky, First order phase transition in the SU(3) gauge theory at finite temperature, *Phys. Rev. D* **26**, 963 (1982).
- [7] F. Y. Wu, The Potts model, *Rev. Mod. Phys.* **54**, 235 (1982); Erratum, *Rev. Mod. Phys.* **55**, 315 (1983).
- [8] C. S. Fischer, J. Luecker, and J. M. Pawłowski, Phase structure of QCD for heavy quarks, *Phys. Rev. D* **91**, 014024 (2015).
- [9] K. Kashiwa, R. D. Pisarski, and V. V. Skokov, Critical endpoint for deconfinement in matrix and other effective models, *Phys. Rev. D* **85**, 114029 (2012).
- [10] R. D. Pisarski and F. Wilczek, Remarks on the chiral phase transition in chromodynamics, *Phys. Rev. D* **29**, 338 (1984).
- [11] F. R. Brown, F. P. Butler, H. Chen, N. H. Christ, Z.-h. Dong, W. Schaffer, L. I. Unger, and A. Vaccarino, On the Existence of a Phase Transition for QCD with Three Light Quarks, *Phys. Rev. Lett.* **65**, 2491 (1990).
- [12] F. Cuteri, O. Philipsen, and A. Sciarra, On the order of the QCD chiral phase transition for different numbers of quark flavours, *J. High Energy Phys.* **11** (2021) 141.
- [13] H. Saito, S. Ejiri, S. Aoki, T. Hatsuda, K. Kanaya, Y. Maezawa, H. Ohno, and T. Umeda (WHOT-QCD Collaboration), Phase structure of finite temperature QCD in the heavy quark region, *Phys. Rev. D* **84**, 054502 (2011).
- [14] S. Ejiri, S. Itagaki, R. Iwami, K. Kanaya, M. Kitazawa, A. Kiyohara, M. Shirogane, and T. Umeda (WHOT-QCD Collaboration), End point of the first-order phase transition of QCD in the heavy quark region by reweighting from quenched QCD, [arXiv:1912.10500](https://arxiv.org/abs/1912.10500).
- [15] A. Kiyohara, M. Kitazawa, S. Ejiri, and K. Kanaya, Finite-size scaling around the critical point in the heavy quark region of QCD, [arXiv:2108.00118](https://arxiv.org/abs/2108.00118).
- [16] M. Fromm, J. Langelage, S. Lottini, and O. Philipsen, The QCD deconfinement transition for heavy quarks and all baryon chemical potentials, *J. High Energy Phys.* **01** (2012) 042.
- [17] F. Cuteri, O. Philipsen, A. Schön, and A. Sciarra, The deconfinement critical point of lattice QCD with $N_f = 2$ Wilson fermions, [arXiv:2009.14033](https://arxiv.org/abs/2009.14033).
- [18] R. Kara, S. Borsanyi, Z. Fodor, J. N. Guenther, P. Parotto, A. Pasztor, and D. Sexty, The upper right corner of the Columbia plot with staggered fermions, in *Proceedings of the 38th International Symposium on Lattice Field Theory*, [arXiv:2112.04192](https://arxiv.org/abs/2112.04192).
- [19] J. Kuti, J. Polonyi, and K. Szlachanyi, Monte Carlo study of SU(2) gauge theory at finite temperature, *Phys. Lett.* **98B**, 199 (1981).
- [20] L. D. McLerran and B. Svetitsky, A Monte Carlo study of SU(2) Yang-Mills theory at finite temperature, *Phys. Lett.* **98B**, 195 (1981).
- [21] K. Kajantie, C. Montonen, and E. Pietarinen, Phase transition of SU(3) gauge theory at finite temperature, *Z. Phys. C* **9**, 253 (1981).
- [22] T. Celik, J. Engels, and H. Satz, The order of the deconfinement transition in SU(3) Yang-Mills theory, *Phys. Lett.* **125B**, 411 (1983).
- [23] S. A. Gottlieb, J. Kuti, D. Toussaint, A. D. Kennedy, S. Meyer, B. J. Pendleton, and R. L. Sugar, The Deconfining Phase Transition and the Continuum Limit of Lattice Quantum Chromodynamics, *Phys. Rev. Lett.* **55**, 1958 (1985).
- [24] J. B. Kogut, H. Matsuoka, M. Stone, H. W. Wyld, S. H. Shenker, J. Shigemitsu, and D. K. Sinclair, Quark and Gluon Latent Heats at the Deconfinement Phase Transition in SU(3) Gauge Theory, *Phys. Rev. Lett.* **51**, 869 (1983).
- [25] B. Svetitsky and F. Fucito, Latent heat of the SU(3) gauge theory, *Phys. Lett.* **131B**, 165 (1983).
- [26] F. R. Brown, N. H. Christ, Y. F. Deng, M. S. Gao, and T. J. Woch, Nature of the Deconfining Phase Transition in SU(3) Lattice Gauge Theory, *Phys. Rev. Lett.* **61**, 2058 (1988).
- [27] M. Fukugita, M. Okawa, and A. Ukawa, Order of the Deconfining Phase Transition in SU(3) Lattice Gauge Theory, *Phys. Rev. Lett.* **63**, 1768 (1989).

- [28] N. A. Alves, B. A. Berg, and S. Sanielevici, Binder energy cumulant for SU(3) lattice gauge theory, *Phys. Lett. B* **241**, 557 (1990).
- [29] P. Bacilieri *et al.*, A new computation of the correlation length near the deconfining transition in SU(3), *Phys. Lett. B* **224**, 333 (1989).
- [30] B. Lucini, M. Teper, and U. Wenger, The deconfinement transition in SU(N) gauge theories, *Phys. Lett. B* **545**, 197 (2002).
- [31] B. Lucini, M. Teper, and U. Wenger, Properties of the deconfining phase transition in SU(N) gauge theories, *J. High Energy Phys.* **02** (2005) 033.
- [32] U. Wolff, Collective Monte Carlo Updating for Spin Systems, *Phys. Rev. Lett.* **62**, 361 (1989).
- [33] A. Ferrenberg and R. Swendsen, New Monte Carlo Technique for Studying Phase Transitions, *Phys. Rev. Lett.* **61**, 2635 (1988).
- [34] A. M. Ferrenberg and R. H. Swendsen, Optimized Monte Carlo Analysis, *Phys. Rev. Lett.* **63**, 1195 (1989).
- [35] E. Marinari and G. Parisi, Simulated tempering: A new Monte Carlo scheme, *Europhys. Lett.* **19**, 451 (1992).
- [36] R. H. Swendsen and J.-S. Wang, Replica Monte Carlo Simulation of Spin-Glasses, *Phys. Rev. Lett.* **57**, 2607 (1986).
- [37] K. Hukushima and K. Nemoto, Exchange Monte Carlo method and application to spin glass simulations, *J. Phys. Soc. Jpn.* **65**, 1604 (1996).
- [38] G. Burgio, M. Fuhrmann, W. Kerler, and M. Muller-Preussker, Modified SO(3) lattice gauge theory at $T \neq 0$ with parallel tempering: Monopole and vortex condensation, *Phys. Rev. D* **75**, 014504 (2007).
- [39] G. Burgio, M. Fuhrmann, W. Kerler, and M. Muller-Preussker, Vortex free energy and deconfinement in center-blind discretizations of Yang-Mills theories, *Phys. Rev. D* **74**, 071502 (2006).
- [40] G. Boyd, Tempered fermions in the hybrid Monte Carlo algorithm, *Nucl. Phys. B, Proc. Suppl.* **60A**, 341 (1998).
- [41] 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).
- [42] 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).
- [43] S. Borsanyi and D. Sexty, Topological susceptibility of pure gauge theory using density of states, *Phys. Lett. B* **815**, 136148 (2021).
- [44] M. Shirogane, S. Ejiri, R. Iwami, K. Kanaya, M. Kitazawa, H. Suzuki, Y. Taniguchi, and T. Umeda (WHOT-QCD Collaboration), Latent heat and pressure gap at the first-order deconfining phase transition of SU(3) Yang-Mills theory using the small flow-time expansion method, *Prog. Theor. Exp. Phys.* **2021**, 013B08 (2021).
- [45] E. Marinari, Optimized Monte Carlo methods, in *Advances in Computer Simulation* (Springer, New York, 1998), pp. 50–81.
- [46] D. J. Earl and M. W. Deem, Parallel tempering: Theory, applications, and new perspectives, *Phys. Chem. Chem. Phys.* **7**, 3910 (2005).
- [47] Y. Aoki, Z. Fodor, S. Katz, and K. Szabo, The QCD transition temperature: Results with physical masses in the continuum limit, *Phys. Lett. B* **643**, 46 (2006).
- [48] S. Borsanyi, S. Durr, Z. Fodor, C. Hoelbling, S. D. Katz *et al.*, High-precision scale setting in lattice QCD, *J. High Energy Phys.* **09** (2012) 010.
- [49] M. Luscher, Properties and uses of the Wilson flow in lattice QCD, *J. High Energy Phys.* **08** (2010) 071.
- [50] Z. Fodor, K. Holland, J. Kuti, S. Mondal, D. Nogradi, and C. H. Wong, The lattice gradient flow at tree-level and its improvement, *J. High Energy Phys.* **09** (2014) 018.
- [51] S. Borsanyi *et al.*, Leading-order hadronic vacuum polarization contribution to the muon magnetic moment from lattice QCD, [arXiv:2002.12347](https://arxiv.org/abs/2002.12347).
- [52] M. Shirogane, S. Ejiri, R. Iwami, K. Kanaya, and M. Kitazawa, Latent heat at the first order phase transition point of SU(3) gauge theory, *Phys. Rev. D* **94**, 014506 (2016).
- [53] S. Borsanyi, G. Endrodi, Z. Fodor, S. Katz, and K. Szabo, Precision SU(3) lattice thermodynamics for a large temperature range, *J. High Energy Phys.* **07** (2012) 056.
- [54] H. Suzuki, Energy-momentum tensor from the Yang-Mills gradient flow, *Prog. Theor. Exp. Phys.* **2013**, 083B03 (2013).
- [55] The 2.1 factor was measured without communication on an AMD EPYC node using the AVX-2 instructions. The operation count needed to calculate the staple is 7 times higher with the 2×1 term in the gauge action than without.
- [56] G. Boyd, J. Engels, F. Karsch, E. Laermann, C. Legeland, M. Lütgemeier, and B. Petersson, Thermodynamics of SU(3) lattice gauge theory, *Nucl. Phys.* **B469**, 419 (1996).
- [57] www.gauss-centre.eu.
- [58] M. Schmelling, Averaging correlated data, *Phys. Scr.* **51**, 676 (1995).