Konrad-Zuse-Zentrum
für Informationstechnik Berlin

Takustraße 7
D-14195 Berlin-Dahlem
Germany

Ernst-Michael Ilgenfritz[1], Werner Kerler[2],
Michael Müller-Preussker[2], Hinnerk Stüben

Parallel tempering in full QCD with Wilson fermions

[1]Research Center for Nuclear Physics, Osaka University, Japan
[2]Institut für Physik, Humboldt-Universität zu Berlin, Germany

Accepted for publication in Physical Review D

ZIB-Report 01–33 (November 2001, revised February 2002)

Parallel tempering in full QCD with Wilson fermions

E.-M. Ilgenfritz

Research Center for Nuclear Physics,

Osaka University, Osaka 567-0047, Japan

W. Kerler

Institut für Physik, Humboldt Universität zu Berlin, D-10115 Berlin, Germany

M. Müller-Preussker

Institut für Physik, Humboldt Universität Berlin, D-10115 Berlin, Germany

H. Stüben

Konrad-Zuse-Zentrum für Informationstechnik Berlin, D-14195 Berlin, Germany

(Dated: 13 February 2002)

Abstract

We study the performance of QCD simulations with dynamical Wilson fermions by combining the Hybrid Monte Carlo algorithm with parallel tempering on 10^4 and 12^4 lattices. In order to compare tempered with standard simulations, covariance matrices between sub-ensembles have to be formulated and evaluated using the general properties of autocorrelations of the parallel tempering algorithm. We find that rendering the hopping parameter κ dynamical does not lead to an essential improvement. We point out possible reasons for this observation and discuss more suitable ways of applying parallel tempering to QCD.

PACS numbers: 11.15.Ha, 12.38.Gc

I. INTRODUCTION

Improving Hybrid Monte Carlo (HMC) simulations of QCD with dynamical fermions is a long standing problem. While better decorrelation is, of course, highly desirable for all observables, it is of crucial importance for the ones which are sensitive to topological sectors. In fact, it has been observed that the η' correlator is definitely dependent on the topological charge Q [1]. Thus it is quite important to look for simulation methods that produce realistic Q-distributions. From this point of view the topological charge appears to be a good touchstone when looking for improvements by new methods.

For staggered fermions an insufficient tunneling rate of the topological charge Q has been observed [2, 3]. For Wilson fermions the tunneling rate has been claimed to be adequate in many cases [4, 5]. However, since the comparison is somewhat subtle, the reason for this could also be that one is not as far in the critical region as with staggered fermions. One could fear that simulating closer to the chiral limit, insufficient tunneling could become for Wilson fermions as severe as for staggered ones. Indeed, for Wilson fermions on large lattices and for large values of κ near the chiral limit the distribution of Q is not symmetric even after more than 3000 trajectories (see *e.g.* Figure 1 of Ref. [4]).

In the method of *simulated* tempering first proposed in Ref. [6] the inverse temperature is made a dynamical variable in the simulations. More generally, any parameter in the action can be made dynamical. Let us suppose that, depending on this particular coupling parameter, the chosen algorithm has a largely different tunneling rate between certain metastable states (in configuration space). Augmenting the algorithm with the tempering method means that now the system is updated in an enlarged configuration space including the coupling. Instead of overcoming a high barrier at an unfortunate parameter value, a detour in parameter space is now opened to be an easier route. This results in better decorrelation.

Considerable improvements have been obtained with dynamical number of the degrees of freedom in the Potts-Model [7], with dynamical inverse temperature for spin glass [8] and with dynamical monopole coupling in U(1) lattice theory [9]. With dynamical mass of staggered fermions in full QCD [10] it has been indicated that by tempering a better sampling of the configuration space (with respect to topological charge) can be achieved. However, simulated tempering requires the determination of a weight function in the generalized action, and an efficient method of estimating it [8, 9] turns out crucial for successfully accelerating the simulation.

A major progress was the proposal of the *parallel* tempering method (PT) [11, 12], in which no weight function needs to be determined. This method has allowed large improvements in the case of spin glass [11]. In QCD improvements have been reported with staggered fermions [13], applying PT to subensembles characterized by different values of the quark mass. This has led to the expectation that, analogously in the case of Wilson fermions, introducing various values of the hopping parameter κ might be the right choice of parameter for applying the idea of PT. In a first study of this problem, simulations of QCD with $\mathrm{O}(a)$-improved Wilson fermions [14], no computational advantage has been found. Because only two coupled sub-ensembles, both at relatively small κ, had been used, this could not be the final answer concerning the potential capabilities of the PT method. In a previous work [15], with more ensembles and (standard) Wilson fermions on an 8^4 lattice, we have observed a considerable increase of the transitions between topological sectors. We have extended this study to larger lattices (10^4 and 12^4) in [16].

In the present paper, our task will be to compare standard HMC with HMC combined with PT in a more elaborate, quantitative manner. In order to really compare algorithms one has to relate the computational effort (computer time) to errors of final results (e.g. particle masses). In the case of PT the calculation of errors becomes more complicated, because one has to take cross correlations between ensembles into account. Cross correlations lead to the technical problem of calculating full covariance matrices from (auto-) correlation functions. In this paper, we carry out such an analysis for the average plaquette and the topological charge.

At the time when we were doing our autocorrelation analysis SESAM published autocorrelation results for all their runs [17]. Qualitatively the SESAM results are very similar to ours, i.e. they also do not observe a clear increase of autocorrelation times with increasing κ (lower quark mass) in the interval $[0.1560, 0.1575]$. One might have expected some mass dependence from the positive experience with tempering in the case of staggered fermions. In the light of this fact, the final conclusion of our analysis, that the PT method with respect to the hopping parameter does not lead to an improved performance of the HMC algorithm, is not really astonishing. Nevertheless, formulating the tools for the correlation analysis is a major part of this paper which should prove useful for possible future applications of PT.

The paper is organized as follows. In Section II we describe the method of PT to be applied to the HMC algorithm. Our simulation results are discussed in Section III. In Section IV we give general properties of covariances and autocorrelations which – in view of the moderate statistics available – must be employed to fully exploit the autocorrelation data. In Section V we present and discuss our results on integrated autocorrelation times. The respective results for off-diagonal elements of the covariances are presented and discussed in Section VI. In Section VII we compare the efficiency of the two simulation algorithms (HMC without and with PT) and study the effect of cross correlations. In Section VIII we comment on swap acceptance rates and give a new formula for that rate extending the one used in Ref. [14]. We conclude in Section IX and point out potentially more promising applications of PT to QCD.

II. PARALLEL TEMPERING

In standard Monte Carlo simulations one deals with one parameter set α and generates a sequence of field configurations $F(s)$, where s denotes the Monte Carlo time. In our case the set α includes the physical parameters β, κ and algorithmic parameters like the leapfrog time step and the number of time steps per trajectory. One field configuration $F(s)$ comprises the gauge field and the pseudo fermion field.

In the parallel tempering (PT) approach [11, 12] one updates K field configurations F_ν with $\nu = 1, \ldots, K$ in the same run. The characteristic feature is that the assignment of the parameter sets α_j with $j = 1, \ldots, K$ to the field configurations F_ν changes in the course of a tempered simulation. The total configuration at time s thus consists of $B(s)$, $F_1(s)$, $F_2(s)$,..., $F_K(s)$ where the permutation

$$B(s) = \begin{pmatrix} \nu_1(s) & \nu_2(s) & \ldots & \nu_j(s) & \ldots & \nu_K(s) \\ 1 & 2 & \ldots & j & \ldots & K \end{pmatrix} \tag{1}$$

describes the assignment of the field configurations $F_{\nu_j(s)}(s)$ to the parameter sets α_j. For short this approach is called PT with K ensembles.

The update of the F_ν occurs by the usual HMC procedure using the parameter sets α_j as assigned at a given time. The update of B is achieved by swapping pairs according to the Metropolis acceptance condition with probability

$$P_{\mathrm{swap}}(i,j) \;=\; \min\left(1, e^{-\Delta H}\right) \,,$$
$$\Delta H \;=\; H(\alpha_i, F_{\nu_i}) + H(\alpha_j, F_{\nu_j}) - H(\alpha_i, F_{\nu_j}) - H(\alpha_j, F_{\nu_i}) \tag{2}$$

where H denotes the Hamiltonian of the HMC dynamics for the parameter set α_j and the field configurations F_{ν_j}. The total update of the Monte Carlo algorithm, after which its time s increases by one, then consists of the updates of all F_μ followed by the full update of B which consists of a sequence of attempts to swap pairs.

Detailed balance for the swapping follows from (2). Ergodicity is obtained by updating all F_ν and by swapping pairs in such a way that all permutations (1) can be reached. There remains still the freedom of choosing the succession of the individual steps. All such choices lead to legitimate algorithms, which might differ in efficiency.

Our choice of steps is such that the updates of all F_ν and that of B alternate. Our criterion for choosing the succession of swapping pairs in the update of B has been to minimize the average time it takes for the association of a field configuration to the parameters to travel from the lowest to the largest κ-value. This has led us to swap pairs belonging to neighboring κ-values and to proceed with this from smaller to larger κ-values.

The observables of interest are associated to a specific parameter set α_j. We denote them as

$$\mathcal{O}_j(s) \equiv \mathcal{O}(F_{\nu_{j(s)}}(s)), \quad j = 1, \ldots, K \,. \tag{3}$$

III. SIMULATION RESULTS

We have simulated lattice QCD with standard Wilson fermions and one-plaquette action for the gauge fields. In the HMC program the standard conjugate gradient inverter with even/odd preconditioning was used. The trajectory length was always 1. The time steps were adjusted to get acceptance rates of about 70% in the HMC Metropolis step. In all cases (standard HMC, tempered runs for all lattice sizes and ensemble sizes) 1000 trajectories were generated, with additional 50–100 trajectories for thermalization.

We have performed tempered runs using 6 and 7 ensembles, all at $\beta = 5.6$ on 10^4 and 12^4 lattices, as well as standard HMC runs for comparison. Our simulations cover the κ-range investigated by SESAM ($\kappa = 0.156, 01565, 0.157, 0.1575$) [4]. In a large scale PT simulation analogous to that by SESAM one would be interested to use all subensembles (separated in κ) anyway, such that additional computational effort seems to be affordable. For instance, in the run with 6 ensembles we extended the κ-range by adding lower values of κ, while in the run with 7 ensembles we have used a denser spacing of the κ-values. Our κ-values are listed in Table I.

The observables determined were the average plaquette P and the topological charge Q. The topological charge was measured using its naively discretized plaquette form after doing 50 cooling steps of Cabibbo-Marinari type. This method gives multiples of a unit charge. The value of the unit charge is close to 1 and has to be determined from the measurements.

Figure 1 shows typical time series of Q obtained for standard HMC and for tempered HMC with 6 and with 7 ensembles. One sees that tempering makes Q fluctuate much stronger. Such behavior is indicative for the decreasing of correlations between subsequent trajectories. The time series on the 12^4 lattice exhibits a richer pattern of transitions as that on the 10^4 lattice, and the width of the topological-charge distribution increases. But the rate of fluctuations decreases with increasing κ.

In our previous investigations [15, 16] we have considered the mean absolute change of Q (called mobility in [4]) to account quantitatively for these rates of fluctuations. However, this quantity does not provide a quantitative measure of the computational gain obtainable with the tempering method in comparison with standard HMC.

In the present investigation, we therefore base the comparison on the full account of the non-diagonal covariance matrix for different observables to be introduced in Eq. (5). The covariance matrix will be calculated from general correlation functions which, for an observable $\mathcal{O}$ and a number N of updates, are defined as

$$R_{jk}(t) = \frac{1}{N}\sum_{s=1}^{N} \mathcal{O}_j(s)\mathcal{O}_k(s+t) - \Big(\frac{1}{N}\sum_{s'=1}^{N}\mathcal{O}_j(s')\Big)\Big(\frac{1}{N}\sum_{s''=1}^{N}\mathcal{O}_k(s'')\Big) . \tag{4}$$

For $j = k$ they are the usual autocorrelation functions, while for $j \neq k$ they describe cross correlations between different ensembles.

Typical examples of normalized autocorrelation functions $\rho_j(t) = R_{jj}(t)/R_{jj}(0)$ are presented in Figure 2. It can be seen that for the tempered runs the decay is considerably faster than for the standard ones. Among the tempering runs it is fastest for the run with 7 ensembles. In the latter case the remarkable fast decay occurring in the interval $t \in [0, 1]$ should be noticed.

With the statistics available, the correlation functions decay below the Monte-Carlo noise for relatively small t. Although the much faster decay of the functions for tempering is apparent, giving numbers for the autocorrelation times and cross correlations is clearly a formidable task. In principle, in order to estimate the autocorrelation times within 10 % one would need trajectory numbers higher than ours by roughly an order of magnitude. To reach a conclusion about the expected improvement we do not attempt such a precision measurement of the autocorrelation times. Moreover, we want to apply some a priori knowledge on the spectral properties in order to exploit our simulation data in the most efficient way possible. Therefore, in the following we first have to elaborate on some theoretical issues concerning covariances.

IV. COVARIANCE MATRIX AND MARKOV SPECTRUM

We obtain the covariance matrix by using the general correlation function (4) and generalizing the derivation given in Ref. [18] for the case $j = k$

$$C_{jk} = \frac{1}{N}\left(R_{jk}(0) + \sum_{t=1}^{N-1}\left(1 - \frac{t}{N}\right)\Big(R_{jk}(t) + R_{kj}(t)\Big)\right) . \tag{5}$$

The diagonal elements of (5) are the variances of $\mathcal{O}_j$ which are traditionally written in the form

$$\mathrm{var}(\mathcal{O}_j) = \frac{R_{jj}(0)}{N}\, 2\tau_j \ , \tag{6}$$

introducing the integrated autocorrelation times

$$\tau_j = \frac{1}{2} + \sum_{t=1}^{N-1} \rho_j(t) \ , \tag{7}$$

where $\rho_j(t) = R_{jj}(t)/R_{jj}(0)$.

When evaluating practical simulations the summation in (7) up to $N-1$ makes no sense since $\rho_j(t)$ is buried in the Monte Carlo noise already for relatively small t. Therefore, it has been proposed [18, 19] to sum up only to some smaller value M of t. However, in practice that procedure is not stable against the choice of M and neglecting the rest is a bad approximation. The proposal to estimate the neglect by an extrapolation based on the t values M and $M-1$ [20] is still inaccurate in general. A more satisfying procedure is to describe the rest by a fit function based on the (reliable) terms of (5) for $t \leq M$ and on the general knowledge about the Markov spectrum. This procedure has led to perfect results in other applications [21].

In order to apply the latter strategy also for determining off-diagonal entries in (5) we have to look how spectral properties enter the parallel–tempering case. For such considerations it is convenient to introduce a Hilbert space [19, 22] with an inner product $(\phi, \chi) = \sum_{\mathcal{C}} \mu(\mathcal{C})\phi^*(\mathcal{C})\chi(\mathcal{C})$, where $\mu(\mathcal{C})$ is the equilibrium distribution of the system and $\mathcal{C}$ denotes the configurations $\mathcal{C} = \{B,\ F_1,\ F_2, \ldots, F_K\}$. Using this notation we can write the expectation values $\langle \mathcal{O}_j \rangle = \sum_{\mathcal{C}} \mu(\mathcal{C})\mathcal{O}_j(\mathcal{C}) = \Big(1, \mathcal{O}_j\Big)$ and the two-time correlation functions

$$\langle \mathcal{O}_j(0)\mathcal{O}_k(t)\rangle = \sum_{\mathcal{C}^{[0]},\mathcal{C}^{[t]}} \mu(\mathcal{C}^{[0]})\mathcal{O}_j(\mathcal{C}^{[0]})W^t(\mathcal{C}^{[0]};\mathcal{C}^{[t]})\mathcal{O}_k(\mathcal{C}^{[t]}) = \Big(\mathcal{O}_j, W^t\mathcal{O}_k\Big) \tag{8}$$

as inner products, where

$$W^t(\mathcal{C}^{[0]};\mathcal{C}^{[t]}) = \sum_{\mathcal{C}^{[1]},\ldots,\mathcal{C}^{[t-1]}} W(\mathcal{C}^{[0]};\mathcal{C}^{[1]})W(\mathcal{C}^{[1]};\mathcal{C}^{[2]})\ldots W(\mathcal{C}^{[t-1]};\mathcal{C}^{[t]}) \tag{9}$$

is the t-step transition matrix constructed from the one-step transition matrix $W(\mathcal{C};\mathcal{C}')$ of the Markov process considered. In the spectral representation

$$W = \sum_{r\geq 1} \lambda_r P_r \tag{10}$$

with eigenvalues λ_r and projection operators P_r, one has $\lambda_1 = 1$ for the equilibrium eigenvector $\mu(\mathcal{C})$ and $|\lambda_r| < 1$ for the other modes. Obviously only P_1 contributes to the stationarity relation $\sum_{\mathcal{C}'} \mu(\mathcal{C}')W(\mathcal{C}';\mathcal{C}) = \mu(\mathcal{C})$, and $P_1(\mathcal{C}',\mathcal{C}) = \mu(\mathcal{C})$ follows. With this notation one can rewrite

$$\langle \mathcal{O}_j\rangle\langle \mathcal{O}_k\rangle = \Big(\mathcal{O}_j, P_1\mathcal{O}_k\Big)\ . \tag{11}$$

Using (8), (10) and (11) we obtain for the general correlation function

$$R_{jk}(t) = \langle \mathcal{O}_j(0)\mathcal{O}_k(t)\rangle - \langle \mathcal{O}_j\rangle\langle \mathcal{O}_k\rangle = \sum_{r>1} \lambda_r^t \Big(\mathcal{O}_j, P_r\mathcal{O}_k\Big)\ , \tag{12}$$

where, due to the subtraction, the term with $\lambda_1 = 1$ cancels out. We thus get the general representation

$$R_{jk}(t) = \sum_{r>1} a_{jkr}\lambda_r^t \quad \text{with} \quad |\lambda_r| < 1 \tag{13}$$

where only the coefficients a_{jkr} depend on the particular pair of observables while the eigenvalues λ_r are universal and characterizing the chosen simulation algorithm. It is important to realize that this also holds for the observables of form (3) used in PT.

V. AUTOCORRELATION RESULTS

For the numerical evaluation of (5) we apply the method explained in Section IV using the fact that after some time only the slowest mode in (13) survives. Our method is to sum up the simulation data only up to some t before the noisy region and determining the rest of the sum from a fit assuming that the fit describes the slowest mode well. For τ_j the rest typically amounts maximally to about 25 %. The proper choice of the fit intervals in t (excluding the region of fast contributions and the noisy region) was controlled by inspection of the graphs and watching the resulting χ^2 values. Examples of such fits are shown in Figure 2.

In view of the moderate statistics available we additionally have made use of the universality of the Markov spectrum implying that in (13) for given algorithm and lattice size only the coefficients a_{jkr} can vary with the observables. We have verified that a collective fit for the whole diagonal with a universal slowest mode gives results comparable to individual fits (see Table I). That motivated us to perform collective fits with one single mode to diagonal and non-diagonal terms. In fact that method greatly helped to get stable fits which will be further discussed in section VI.

To obtain errors for the covariances one can generalize the derivations given in Ref. [18] for the diagonal case to calculate covariances of covariances from the $R_{jk}(t)$ data only. However, such calculation is impractical with the statistics available. Therefore, we have to rely on the comparison of measurements of covariances at different parameter values and on consistency checks to get some idea about the size of the errors.

Comparing the results for $R_{jj}(0)$ from the different simulation algorithms, which should give the same numbers, one sees that this ingredient of a calculation of τ_j has errors of about 20 %. The exponential autocorrelation time $\tau_{\text{exp}} = -1/\ln\lambda$, where λ denotes the slowest mode, corresponds within a good approximation to the integrated autocorrelation time which one would get taking only the slowest mode into account. The presence of faster modes then renders τ_j smaller than τ_{exp} . We always find consistency with this requirement.

The integrated autocorrelation times τ_j obtained in this way are given in Table I. One can see that there is little difference between the results from individual and collective fits. The fluctuations of values found for neighboring κ-values indicate relatively large errors. Judging from the observed noise levels the errors are expected to be largest for the standard case and smallest for tempering with 7 ensembles. Despite these errors two unexpected features are clearly visible: i) there is no sizable increase in τ_j with κ and ii) there is nevertheless gain in terms of τ_j when using tempering.

The lack of a sizable increase in τ_j with κ, which one would have expected for the standard runs, firstly indicates that valuing time histories by eye (as of Q in Figure 1) can be misleading. It secondly shows the important fact that the usual precondition of successful tempering, connecting regions with considerably different τ, is not fulfilled here.

In the light of this it comes with some surprise that nevertheless gain in terms of τ_j is observed. The reduction of τ_j for Q turns out to be larger than for P.

VI. OFF-DIAGONAL COVARIANCES

For the discussion of cross correlations in the following Section we need the off-diagonal elements of the covariance matrices. As in the diagonal case, the use of the simulation data in the sum (5) makes only sense outside the noisy region. One finds that the off-diagonal elements of the general correlation functions are decreasing with the distance from the diagonal. Therefore, the evaluation of the sum (5) in the off-diagonal case is more difficult because beyond some distance $|j-k|$ the elements are completely indiscernible within the noise. See Figure 3 for an illustration.

In tempering with 7 ensembles we generally find that three off-diagonals can be determined while for tempering with 6 ensembles only one. For the off-diagonal elements $R_{jk}(t)$ which clearly show a signal above the noise we generally observe a maximum at $t = |j-k|$ (see Figure 3). We therefore look for a prediction of their functional form. Rewriting (12) as $R_{jk}(t) = \Big(\mathcal{O}_j(s), (1-P_1)W^t\mathcal{O}_k(s)\Big)$ their qualitative behavior can be discussed. Obviously at each time the observables of type (3) depend only on one field configuration. This form of $R_{jk}(t)$ suggests that for $j \neq k$ a sizable contribution only arises when, under the action of the transition matrix by $W^t\mathcal{O}_k(s)$, a contribution also depending on the field configuration entering $\mathcal{O}_j(s)$ has been generated. For our type of swapping this situation occurs for $t \geq |j-k|$ so that

$$R_{jk}(t) \approx \left\{ \begin{array}{cll} \sum_{r>1} \tilde{a}_{jkr}\lambda_r^t & \text{for} & |j-k| \leq t \\ 0 & \text{for} & 0 \leq t < |j-k| \end{array} \right\} \quad \text{for} \quad j \neq k \tag{14}$$

should be the approximate behavior. We indeed generally see this behavior within errors in our data.

For the numerical evaluation of (5) we again apply the method explained in Section IV restricted to a t interval where the respective $R_{jk}(t)$ signal is sufficiently above the noise. To get stable off-diagonal results we use the method of the collective fit, using the existence of a universal slowest mode described in Section V. Table II shows an example of a numerical result for C_{jk} (remember that C_{jk} is symmetric). Generally the off-diagonal elements obtained, especially the smaller ones, are likely to be overestimated because of possible contributions of the noise.

In the applications to be described in Section VII the full covariance matrix C_{jk} is needed. This excludes, for this purpose, the consideration of the PT results in our case of 6 ensembles, because we were able to determine elements only in the first subdiagonal $|j-k|=1$. They have much larger errors than in our PT studies with 7 ensembles. In this case, in contrast, we got 3 sub-diagonals. We find that putting the remaining ones equal to zero or using various extrapolations in $|j-k|$ for them makes only little difference in the results. This reflects the fact that, although the elements close to the diagonal are not small, there is nevertheless a faster decrease farther away from the diagonal.

VII. COMPARISON OF SIMULATION ALGORITHMS

At the end of Section V we have pointed out that there is no increase of autocorrelation times with κ. Because of this observation the usual mechanism of tempering – which is to provide an easier detour through parameter space for the suppressed transitions – is not available here. If this mechanism is working, tempering with several parameter points can be advantageous even if one is interested only in the result at one point. This holds, in particular, for systems where in the region of interest otherwise almost no transition occurs [7, 9]. Unfortunately our region is not of this type.

As also observed in Section V there is nevertheless a reduction of autocorrelation times. The effect of this is, however, not large enough to get generally gain if one is interested in only one point. In fact, dividing the reduction factors of the τ_j in Table I by the number of ensembles it can be seen that at best in the case of Q some gain remains.

We now turn to the question whether gain remains if one is interested in the results at all parameter values. In this case it is necessary to account for cross correlations between the ensembles. To be able to do this one has to rely on fits to the data. The respective fit method is well know from the treatment of indirect measurements (see e.g. Ref. [23]). For proper comparison this method, which leads to improved errors, has to be applied to the tempering case as well as to the standard case.

Final results then are obtained from fits to mean values from individual ensembles. In the case of PT the full covariance matrix enters the fit. Although there are difficulties to account for the full matrix numerically, we have tried to develop some feeling for its influence by making fits to the observables we have measured.

It is known that $\langle Q\rangle = 0$. Therefore our fit ansatz for $\langle Q\rangle$ is a constant, i.e., the fit procedure is a weighted mean using the full covariance matrix. For the plaquette we observed that our data are consistent with a linear dependence on κ. Therefore we have used a linear fit ansatz in κ.

In the following we outline the fit method for the case of the plaquette. The linear fit ansatz just mentioned is $\langle P\rangle = x_1 + x_2\kappa$, where x_1 and x_2 are the fit parameters. In matrix notation we have $\eta = -Ax$ with

$$\eta = \begin{pmatrix} \langle P\rangle_1 \\ \vdots \\ \langle P\rangle_K \end{pmatrix}, \quad A = -\begin{pmatrix} 1 & \kappa_1 \\ \vdots & \vdots \\ 1 & \kappa_K \end{pmatrix}, \quad x = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} . \tag{15}$$

We also introduce the vector of measured values $p = (\bar{P}_1, \ldots, \bar{P}_K)^T$ and call the corresponding covariance matrix C_p. The result of the fit is the minimum of $(\eta - p)^T C_p^{-1}(\eta - p)$ which lies at $\tilde{x} = -(A^T C_p^{-1} A)^{-1} A^T C_p^{-1} p$. The errors of the result are square roots of the diagonal entries of $C_{\tilde{x}} = (A^T C_p^{-1} A)^{-1}$. For fitting a constant to measured values of $\langle Q\rangle$ corresponding formulae apply with $x_2 \equiv 0$.

Actually we are interested in the variances of the fit-function values. To get them we insert $\tilde{x}$ into the fit function,

$$\tilde{\eta} = -A\tilde{x} , \tag{16}$$

and using (16) in the transformation law of covariances we obtain

$$C_{\tilde{\eta}} = A C_{\tilde{x}} A^T . \tag{17}$$

The diagonal elements of (17) are the variances of interest and the square roots of them the improved errors.

For standard HMC, in these calculations we have used the measurements for the 5 selected κ values (see Table I). In the case of tempering the number of measurements is equal to the number K of ensembles.

Table III gives data with usual statistical errors and fit results (16) with improved errors (17). We denote the errors by e_0 for the usual statistical errors (in column 2), by e_f for the improved errors taking into account the *full* covariance matrix (column 3), and by e_d for those obtained only with the *diagonal* elements of the covariance matrix (column 4), respectively.

The factor $(e_f/e_d)^2$ describes the influence of cross correlations. We typically find values of about 2 to 3 for it. As compared to the reduction factors apparent from Table I this appears not large. However, one has to be aware that proper comparison here needs consideration of the improved errors in the standard *and* in the tempering case.

The relevant computational gain factor for the comparison standard *vs.* tempering case is given by the improved errors and the numbers of ensembles as

$$(e_d^{standard}/e_f^{tempering})^2 \; N_{standard}/N_{tempering} \; . \tag{18}$$

In the example in Table III this factor appears close to one. However, because of the inaccuracy of the standard data the respective fit results are not reliable (giving factors from about 0.5 to 4 in other cases). Thus we are not able to give definite numerical results for (18).

A further quantity to be mentioned is the reduction factor e_0/e_f for the errors. This reduction in general is larger for larger e_0. Thus the comparison based on the fit procedure favors the standard case, which reduces a possible gain.

Altogether it looks that even with more accurate data it might be difficult to present evidence for computational gain in our case, where the autocorrelation times for standard HMC do not vary within the κ-range considered.

VIII. SWAP ACCEPTANCE RATES

For the effect of enlarging the lattice size on the efficiency of the algorithm the change of the swap acceptance rate $\langle P_{\mathrm{swap}}\rangle$ resulting from (2) appears to be relevant. In Ref. [14] agreement has been reported between the behavior observed in PT with the expression

$\mathrm{erfc}\left(\frac{1}{2}\sqrt{\langle\Delta H\rangle}\right)$ derived in Ref. [24]. The derivation given there, however, relies on the area-preserving property of the HMC algorithm, implying $\langle\exp(-\Delta H)\rangle = 1$, which does not hold in the case of swapping.

To get the relation appropriate for swapping, we again neglect higher cumulants in the cumulant expansion $\langle\exp(-\Delta H)\rangle = \exp\Big(-\langle\Delta H\rangle + \frac{1}{2}\langle(\Delta H - \langle\Delta H\rangle)^2\rangle \mp \dots\Big)$. However, in contrast to Ref. [24], we put $\langle\exp(-\Delta H)\rangle = \exp(-\delta)$ with an unknown δ. By convexity of the exponential function one finds that $\langle\Delta H\rangle \geq \delta$ holds. The relation between mean and width of the gaussian used in Ref. [24] then generalizes to

$$\langle\Delta H\rangle = \frac{1}{2}\sigma^2 + \delta \quad , \quad \sigma^2 = \langle(\Delta H - \langle\Delta H\rangle)^2\rangle \ . \tag{19}$$

The evaluation of the integral $\frac{1}{\sqrt{2\pi}\sigma}\int_{-\infty}^{\infty}\mathrm{d}x\ \min(1,\exp(-x))\exp\Big(-\frac{(x-\langle\Delta H\rangle)^2}{2\sigma^2}\Big)$ gives then

$$\langle P_{\mathrm{swap}}\rangle = \frac{1}{2}\mathrm{erfc}\left(\frac{1}{2}\Big(u + \frac{\delta}{u}\Big)\right) + \frac{\exp(-\delta)}{2}\mathrm{erfc}\left(\frac{1}{2}\Big(u - \frac{\delta}{u}\Big)\right) \tag{20}$$

where $u = \sqrt{\langle\Delta H\rangle - \delta}$.

Within errors our values of $\langle\Delta H\rangle$ turn out to scale with the volume L^4 (being roughly 1.4 and 2.8 for 6 ensembles and 0.35 and 0.7 for 7 ensembles for 10^4 and 12^4, respectively). While the values of $\langle\exp(-\Delta H)\rangle$ for the 10^4 lattice within errors conform with 1, on the 12^4 lattice they deviate substantially from this (increasing from 0.6 to 1.4).

Our data agree with (20) using δ as found from our simulations. In the cases where we find that $\delta = 0$ is not true, despite of this using $\delta = 0$ in the acceptance formulae within errors gives still consistency. However, for larger lattices, where further increase of the deviations of $\langle\exp(-\Delta H)\rangle$ from 1 is to be expected, this might be no longer so. The indicated deviations tend to improve the situation on larger lattices. Since the behavior of δ is not known quantitatively, detailed predictions on $\langle\Delta H\rangle$ at present are not possible.

IX. CONCLUSIONS

In this paper we have compared PT with standard HMC quantitatively on 10^4 and 12^4 lattices. We have described the steps of such an analysis and carried them out for the average plaquette and the topological charge. In a quantitative analysis one has to look at the size of errors and therefore in principle also at errors of the errors. While the first is at the limits of feasibility, the latter is definitely beyond our statistics. For that purpose we have demonstrated how the cross-correlation functions between the sub-ensembles have to be taken into account, using their general, algorithm independent properties. This has allowed us to make consistency checks, and we found consistent behavior of the results. We believe that this part of our analysis is of general interest.

The choice of the κ-range in this paper was guided by recent large-scale QCD simulations with dynamical Wilson fermions like SESAM [4] and triggered by the suppression of tunneling rates for the topological charge observed at the largest κ-values available. However, this range might be still too far from the chiral limit with the consequence of no dramatic change of the autocorrelation times for the standard Hybrid Monte Carlo method. This would explain why PT in our case, *i.e.* for the κ range covered by the work of SESAM, did not provide a considerable computational gain.

At stronger coupling we know that approaching the chiral limit we arrive at a second order transition into a phase with broken combined parity-flavor symmetry (so-called Aoki-phase) [25]. At this transition we expect a critical slowing down and therefore strongly increasing autocorrelations. This is the reason, why we nevertheless believe that PT should become efficient if one is sufficiently close to the chiral limit. An even more promising scenario might be the application of the PT method towards the continuum limit, in particular along lines of constant physics in the $\beta-\kappa$-plane. Already in pure $SU(3)$-Yang-Mills theory the tunneling rate between different topological sectors becomes strongly suppressed with increasing β.

The tools developed in the present paper will be very useful for such future applications.

Acknowledgments

The simulations were done on the CRAY T3E at Konrad-Zuse-Zentrum für Informationstechnik Berlin. E.-M. I. gratefully appreciates the support by the Ministry of Education, Culture and Science of Japan (Monbu-Kagaku-sho) and thanks H. Toki for the hospitality at RCNP.

[1] T. Struckmann, K. Schilling, P. Ueberholz, N. Eicker, S. Güsken, Th. Lippert, H. Neff, B. Orth and J. Viehoff, `hep-lat/0006012`.

[2] M. Müller-Preussker, Proc. of the XXVI Int. Conf. on High Energy Physics, Dallas, Texas (1992), p. 1545.

[3] B. Allés, G. Boyd, M. D'Elia, A. Di Giacomo and E. Vicari, Phys. Lett. **B 389** (1996) 107.

[4] B. Allés, G. Bali, M. D'Elia, A. Di Giacomo, N. Eicker, K. Schilling, A. Spitz, S. Güsken, H. Hoeber, Th. Lippert, T. Struckmann, P. Ueberholz and J. Viehoff, Phys. Rev. **D 58** (1998) 071503.

[5] A. Ali Khan, S. Aoki, R. Burkhalter, S. Ejiri, M. Fukugita, S. Hashimoto, N. Ishizuka, Y. Iwasaki, K. Kanaya, T. Kaneko, Y. Kuramashi, T. Manke, K. Nagai, M. Okawa, H. P. Shanahan, A. Ukawa and T. Yoshié, Nucl. Phys. B. (Proc. Suppl.) **83–84** (2000) 162.

[6] E. Marinari and G. Parisi, Europhys. Lett. **19** (1992) 451.

[7] W. Kerler and A. Weber, Phys. Rev. **B 47** (1993) R11563.

[8] W. Kerler and P. Rehberg, Phys. Rev. **E 50** (1994) 4220.

[9] W. Kerler, C. Rebbi and A. Weber, Phys. Rev. **D 50** (1994) 6984; Nucl. Phys. **B 450** (1995) 452.

[10] G. Boyd, `hep-lat/9701009`.

[11] K. Hukushima and K. Nemoto, `cond-mat/9512035`.

[12] For a review see: E. Marinari, `cond-mat/9612010`.

[13] G. Boyd, Nucl. Phys. B (Proc. Suppl.) **60 A** (1998) 341.

[14] B. Joó, B. Pendleton, S. M. Pickles, Z. Sroczynski, A. C. Irving and J. C. Sexton, Phys. Rev. **D 59** (1999) 114501.

[15] E.–M. Ilgenfritz, W. Kerler and H. Stüben, Nucl. Phys. B. (Proc. Suppl.) **83–84** (2000) 831; H. Stüben in *Lattice Fermions and Structure of the Vacuum*, V. Mitrjushkin and G. Schierholz (eds.), Kluwer Academic Publishers, Dordrecht (2000), p. 211.

[16] E.–M. Ilgenfritz, W. Kerler, M. Müller–Preussker and H. Stüben, `hep-lat/0007039`.

[17] G. S. Bali, N. Eicker, T. Lippert, H. Neff, B. Orth, K. Schilling, T. Struckmann and J. Viehoff, Phys. Rev. **D 64** (2001) 054502.

[18] M. B. Priestley, *Spectral analysis and time series*, Academic Press, London (1987).

[19] N. Madras and A. D. Sokal, J. Stat. Phys. **50** (1988) 109.

[20] U. Wolff, Phys. Lett. **B 228** (1989) 379.

[21] W. Kerler, Phys. Rev. **D 47** (1993) R1285.

[22] W. Kerler, Phys. Rev. **D 48** (1993) 902.

[23] S. Brandt, *Datenanalyse*, BI, Mannheim (1992) (3rd ed., Chapters 9.2 and 9.3).

[24] S. Gupta, A. Irbäck, F. Karsch and B. Petersson, Phys. Lett. **B 242** (1990) 437.

[25] S. Aoki and A. Gocksch, Phys. Rev. **D 45** (1992) 3845;
K. M. Bitar, Phys. Rev. **D 56** (1997) 2736.

TABLE I: Integrated autocorrelation times $\tau_{int,P}$ for plaquette and $\tau_{int,Q}$ for topological charge from individual fits (see Section V). For comparison results from collective fits are given in brackets. For each case 1000 trajectories were generated with trajectory length 1.

	standard HMC		tempered HMC			
			6 ensembles $\Delta\kappa = 0.0005$		7 ensembles $\Delta\kappa = 0.00025$	
	10^4	12^4	10^4	12^4	10^4	12^4
κ	$2\tau_{\mathrm{int},P}$					
0.15500	14.0	16.7	5.9 [6.7]	8.8 [9.5]		
0.15550			6.3 [6.9]	10.2 [8.6]		
0.15600	10.4	13.3	3.7 [4.0]	5.5 [6.2]	4.2 [4.6]	4.6 [5.1]
0.15625					2.8 [2.6]	2.8 [3.4]
0.15650	9.3	12.9	5.2 [5.6]	6.1 [6.9]	2.9 [2.5]	3.4 [3.4]
0.15675					2.7 [2.4]	4.5 [3.7]
0.15700	8.6	14.2	5.8 [5.6]	7.8 [8.1]	2.9 [2.4]	6.1 [4.8]
0.15725					4.1 [2.9]	5.2 [4.9]
0.15750	9.0	8.1	8.4 [8.5]	10.9 [10.8]	3.4 [3.9]	8.2 [7.9]
κ	$2\tau_{\mathrm{int},Q}$					
0.15500	42	22	15 [17]	17 [23]		
0.15550			13 [11]	11 [12]		
0.15600	37	74	7 [7]	20 [19]	16.6 [16.5]	4.8 [7.7]
0.15625					9.7 [9.7]	5.2 [6.3]
0.15650	41	48	8 [7]	16 [17]	5.7 [6.1]	5.8 [6.5]
0.15675					3.3 [3.1]	6.4 [6.8]
0.15700	45	38	16 [12]	8 [10]	2.3 [2.5]	6.3 [6.5]
0.15725					5.3 [4.9]	5.3 [5.5]
0.15750	46	14	6 [8]	35 [26]	6.8 [7.6]	5.5 [7.0]

TABLE II: Covariance matrix of Q for tempering with 7 ensembles on 12^4 lattice.

j	κ_j	$10^3 \cdot R_{j,j}$	$10^3 \cdot R_{j,j+1}$	$10^3 \cdot R_{j,j+2}$	$10^3 \cdot R_{j,j+3}$
1	0.15600	12.68	10.34	6.87	3.14
2	0.15625	9.89	8.28	5.77	2.04
3	0.15650	9.90	8.44	4.42	1.04
4	0.15675	9.27	6.30	3.22	1.23
5	0.15700	7.02	4.46	2.39	
6	0.15725	4.34	2.82		
7	0.15750	4.26			

TABLE III: Data and fit results for P from standard runs and from tempering with 7 ensembles on the 12^4 lattice.

κ	data	fit with full covariance matrix	fit with diagonal covariance matrix
	standard HMC		
0.15500	0.431825 (135)		0.432002 (112)
0.15600	0.429872 (136)		0.429968 (66)
0.15650	0.429343 (132)		0.428952 (55)
0.15700	0.428309 (138)		0.427935 (59)
0.15750	0.426695 (91)		0.426919 (77)
	tempering (7 ensembles)		
0.15600	0.430024 (79)	0.429973 (69)	0.430064 (50)
0.15625	0.429494 (64)	0.429516 (56)	0.429578 (38)
0.15650	0.429138 (63)	0.429060 (48)	0.429093 (30)
0.15675	0.428749 (67)	0.428603 (48)	0.428607 (29)
0.15700	0.428166 (81)	0.428146 (56)	0.428121 (35)
0.15725	0.427560 (82)	0.427690 (68)	0.427636 (46)
0.15750	0.427034 (102)	0.427233 (84)	0.427150 (60)

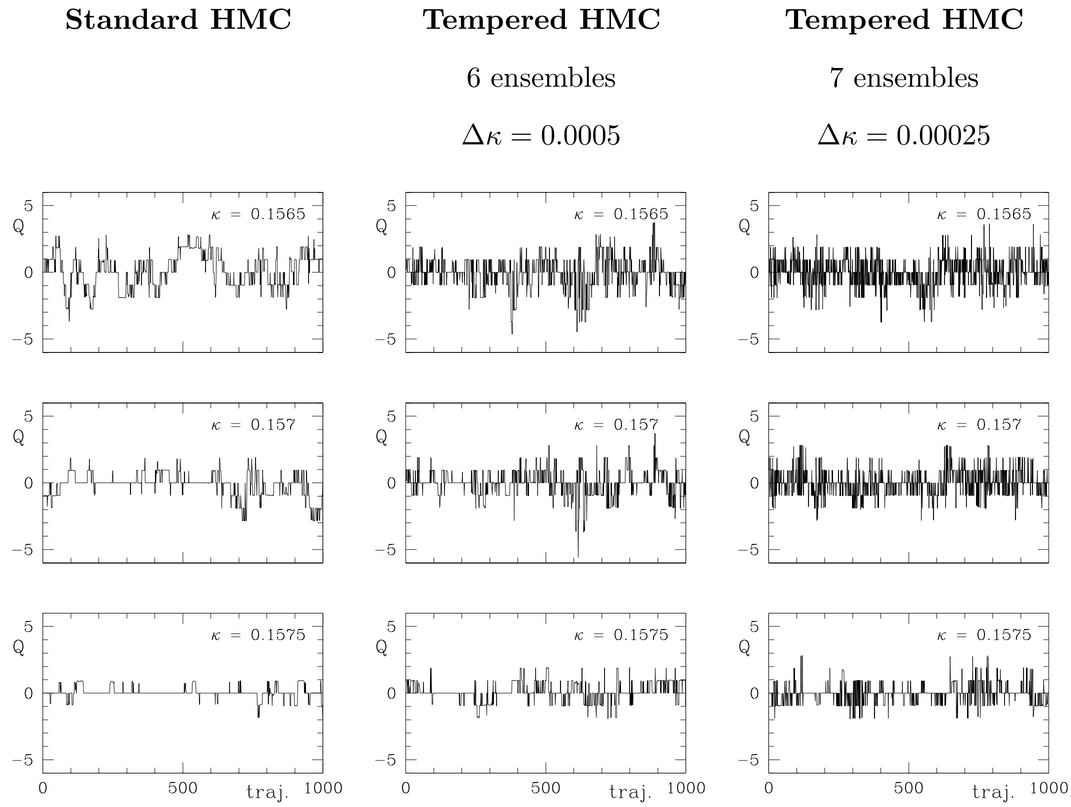

FIG. 1: Time series of Q for standard and tempered HMC on 12^4 lattice at $\beta = 5.6$ (for part of κ-values only, see Table I for full list).

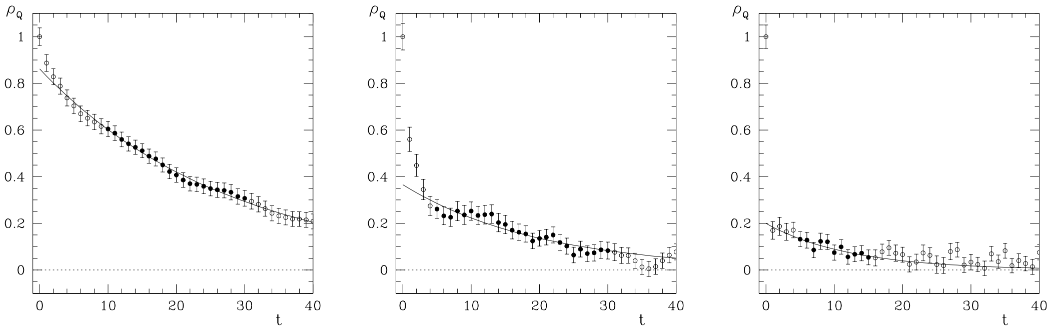

FIG. 2: Normalized autocorrelation functions for Q for standard HMC (left) and tempering with 6 ensembles (center) and 7 ensembles (right) on a 12^4 lattice for $\beta = 5.6$, $\kappa = 0.1565$. The errors indicated are the purely statistical ones. The lines represent fits to the subset of data points with full symbols.

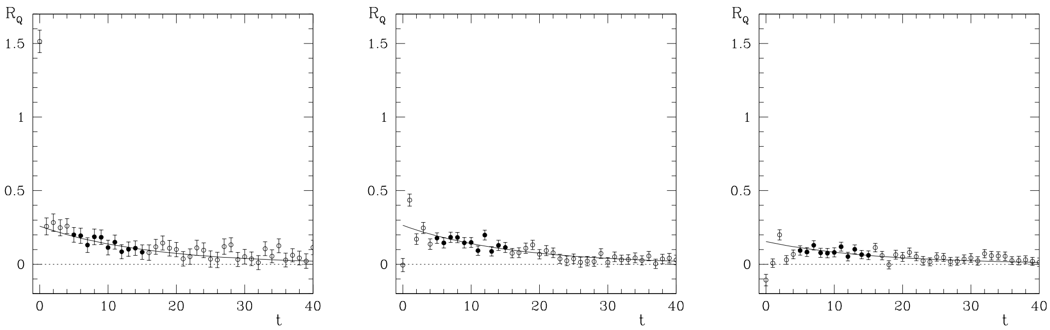

FIG. 3: Correlation functions for Q for tempering with 7 ensembles on a 12^4 lattice for $\beta = 5.6$. Shown are the autocorrelation function at $\kappa = 0.1565$ (left) and the cross-correlation functions for $\kappa = 0.1565, 0.15675$ (center) and $\kappa = 0.1565, 0.157$ (right). The errors indicated are purely statistical ones. The lines represent one combined fit as explained in the text. Full data symbols indicate the fit interval.