Parallel tempering in lattice QCD with $O(a)$-improved Wilson fermions

Bálint Joó, Brian Pendleton, Stephen M. Pickles, and Zbigniew Sroczynski
Department of Physics and Astronomy, The University of Edinburgh, The King's Buildings, Edinburgh EH9 3JZ, Scotland

Alan C. Irving
Theoretical Physics Division, Department of Mathematical Sciences, University of Liverpool, P.O. Box 147, Liverpool L69 3BX, United Kingdom

James C. Sexton
School of Mathematics, Trinity College and Hitachi Dublin Laboratory, Dublin 2, Ireland

(UKQCD Collaboration)
(Received 19 October 1998; published 19 April 1999)

We present details of our investigations of the parallel tempering algorithm. We consider the application of action matching technology to the selection of parameters. We then present a simple model of the autocorrelations for a particular parallel tempered system. Finally we present results from applying the algorithm to lattice QCD with $O(a)$-improved dynamical Wilson fermions for twin sub-ensemble systems. [S0556-2821(99)08109-6]

PACS number(s): 12.38.Gc, 02.70.Lq, 11.15.Ha,

I. INTRODUCTION

The computational cost of lattice QCD has always been enormous. During the last few years the power of supercomputers has grown immensely but simulations with dynamical fermions are still very time consuming.

One of the most popular algorithms for dynamical fermion simulations is the hybrid Monte Carlo (HMC) algorithm [1]. However it has been suggested that the HMC algorithm is not very efficient at decorrelating some long range observables such as the topological charge [2]. On the other hand, results from the SESAM Collaboration [3] indicate that HMC simulations using Wilson fermions seem to tunnel between topological sectors at an adequate rate. The results of SESAM indicate an autocorrelation time for the topological charge of about 50 HMC trajectories.

With such high computational costs it is always necessary to keep an eye open for alternative algorithms. Parallel tempering (PT) or exchange Monte Carlo was proposed in [4] to assist decorrelation in spin-glass systems. A lucid description of PT and related algorithms such as simulated tempering and their applications to spin-glass and other systems may be found in [5,6].

Recently PT has been applied to simulations of lattice QCD with staggered fermions [7] and this preliminary study indicated that the autocorrelation times for some observables were significantly improved over the normal HMC results.

In this paper we present our study of the PT algorithm using 2 flavors of degenerate $O(a)$-improved Wilson fermions [8] with a non-perturbatively determined improvement coefficient [9].

PT simulates several lattice QCD ensembles concurrently, hereafter referred to as *sub-ensembles*, with different simulation parameters. PT exploits the fact that the equilibrium distributions of the configurations in individual sub-ensembles have an overlap, and occasionally tries to swap configurations between pairs of sub-ensembles, while keeping all sub-ensembles in equilibrium. This is done in such a way that the factorization of the joint equilibrium distribution of configurations into the individual distributions for each sub-ensemble is not disturbed by the swapping.

The acceptance of these swap attempts depends on how close the sub-ensembles are to each other in parameter space. The concept of distance in parameter space is formalized in [10–12] by the machinery of action and observable matching. In theory, this technology should allow the selection of an optimal set of parameters to maximize the swap acceptance rate between the sub-ensembles.

Another possibility is to use the action matching technology to define curves in parameter space on which some observable such as r_0 [13] is constant. PT can, in principle be used to simulate numerous points on such a curve in one simulation. However it must be stressed that this scenario is different from the one above. Matching observables is not the same as matching the action [12]. Hence in this case one does not in general have as good control over the swap acceptance rate as in the situation outlined previously.

The remainder of this paper is organized as follows:

Our particular variant of the PT algorithm is described in detail in the following section, where we show that it satisfies detailed balance and present a formula for the acceptance rate of the swap attempts. We then relate this formula to the distance in parameter space as defined in the context of action matching technology.

The swapping of configurations between sub-ensembles is expected to reduce the autocorrelation times of observables within individual sub-ensembles with respect to their HMC autocorrelation times. In Sec. III we discuss the simple case of a PT system consisting of two sub-ensembles. We suggest a model for the autocorrelation function in the PT sub-ensembles in terms of that of the original HMC ensembles.

Our simulations are discussed in Sec. IV and our results

are presented in Sec. V. We show that indeed our acceptance rate formula of Sec. II is borne out by the simulation results. We estimate the autocorrelation time of the plaquette for several swap acceptance rates and compare these estimates with the prediction of the model outlined in Sec. III.

Our summary and conclusions are presented in Sec. VI.

II. THE PARALLEL TEMPERING ALGORITHM

Algorithm description

Normal Monte Carlo integration in lattice QCD generates an ensemble of configurations distributed according to a specified probability measure determined by the action of the system. In practice, in a modern lattice QCD simulation, a number of ensembles are generated for varying lattice sizes and action parameters. This is done to analyze parameter dependence of results, to measure volume effects and to carry out extrapolations to infinite volume and zero quark mass.

In PT, at least two separate sub-ensembles are generated in parallel. At regular periods within the parallel sub-ensemble generation process, the current configurations within pairs of sub-ensembles are subjected to a Metropolis acceptance test, the outcome of which is used to determine whether or not a swap of these configurations between the pair of sub-ensembles should take place.

In lattice QCD with dynamical fermions the fundamental degrees of freedom are the lattice gauge variables, U, defined on the lattice links. Normally a configuration is defined to consist only of the values of all the link variables. However, the algorithm which we use to integrate single sub-ensembles is HMC. This algorithm enlarges the sub-ensemble phase space by introducing additional pseudofermion, ϕ, and link conjugate momentum, π, degrees of freedom. The conjugate momenta are included as an additive term in the action, and play no role in the version of PT which we have implemented. The pseudofermion fields, however, introduce an important complication. In normal HMC, pseudofermions allow the replacement of the logarithm of the fermion determinant term in the lattice action with a pseudofermion bilinear term, which contains the inverse of the fermion hopping matrix. Computationally, the pseudofermion bilinear is much simpler to calculate than the determinant which it replaces.

The consequence of the above for our implementation of PT was, that we were faced with having to choose between the following two schemes:

(1) A ''configuration'' in the PT sense is defined to include links only. The acceptance test then requires the evaluation of the fermion determinant for each parameter set for each configuration.

(2) A ''configuration'' is defined to include both links and pseudofermions. The acceptance test here requires the evaluation of the pseudofermion bilinear for each parameter set for each configuration.

In implementing the PT algorithm we chose to adopt the second approach here, and swapped both links and pseudofermion fields at each successful PT swap step. The fundamental reason for this choice was simply that the additional coding requirement to add parallel tempering to the UKQCD HMC code was minimized. Choosing to swap the links only would have required the development and integration of significant additional code to calculate fermion determinants.

In principle, implementing only one possibility could potentially have left a gap in the analysis of the performance of the PT method for lattice QCD. However, as we report below, the single implementation we have executed has provided us with sufficient detail to estimate also the performance of the second possibility. We find, in fact, that the links-only swap would have performed better than the links plus pseudofermion swap (providing we ignore the additional computational work required to evaluate the fermion determinant). This gain in performance is not sufficient however to change the basic conclusions of this paper.

Notation

Let each sub-ensemble be labelled by an index i and let the phase space of sub-ensemble i be Γ_i. Each sub-ensemble has an action $\mathcal{S}_i$ which depends upon the set of parameters and the fields of the sub-ensemble.

We simulate two flavors of dynamical fermions using the standard pseudofermionic action

$$\mathcal{S}_i = -\beta_i \mathcal{W}_\Box(U) + \phi^\dagger (M^\dagger(\kappa_i, c_i) M(\kappa_i, c_i))^{-1} \phi \quad (1)$$

where $\mathcal{W}_\Box$ is the Wilson plaquette action, U are the gauge fields, ϕ are the pseudofermion fields, and M is the $O(a)$-improved fermion matrix with hopping parameter κ and clover coefficient c. In addition, for HMC algorithms we need to introduce momentum fields π_i and construct Hamiltonian functions $\mathcal{H}_i = \pi_i^2 + \mathcal{S}_i$. A state in sub-ensemble i is then represented by the triple $a_i = (U_i, \pi_i, \phi_i)$ while the parameter set for sub-ensemble i is the triple of real numbers (β_i, κ_i, c_i). Note that the subscript i serves only to distinguish ensembles and will be dropped when discussing a single sub-ensemble.

Each sub-ensemble has the phase space

$$\Gamma_i = \{U_i\} \otimes \{\pi_i\} \otimes \{\phi_i\}. \quad (2)$$

We note at this stage that each Γ_i is an identical copy of the same fundamental phase space, and that the only distinguishing features of individual sub-ensembles are the parameter set choices used in their generation, and the consequent differences in the distributions of configurations which result.

A PT simulation state is thus the collection of states $\{a_i | i = 1 \ldots n\}$, where n is the number of sub-ensembles. The overall PT phase space is the direct product of the phase spaces of the sub-ensembles

$$\Gamma_{\mathrm{PT}} = \prod_{i=1}^{n} \Gamma_i. \quad (3)$$

Detailed Balance

In a PT simulation one needs to construct a Markov process which has (joint) equilibrium probability distribution:

$$P_{\mathrm{PT}}^{\mathrm{eq}}=\prod_i P_i^{\mathrm{eq}}(U,\pi,\phi) \tag{4}$$

where $P^{\mathrm{eq}}(U,\pi,\phi)$ is the desired equilibrium probability distribution of the individual sub-ensemble i. In our case

$$P_i^{\mathrm{eq}}(U,\pi,\phi)=\frac{1}{Z_i}e^{-\mathcal{H}_i(U,\pi,\phi)} \tag{5}$$

$$Z_i=\int [dU][d\pi][d\phi][d\phi^\dagger]e^{-\mathcal{H}_i(U,\pi,\phi)}. \tag{6}$$

Equation (4) formalizes our notion of simulating ensembles independently. To be more precise, the Markov steps within any individual sub-ensemble are independent of those in the others, but the resulting sub-ensembles are not independent as they are coupled by the swapping steps. However the overall joint equilibrium distribution of the PT system is not affected by the swapping, and remains the product of the individual equilibrium distributions of the sub-ensembles.

We define two kinds of Markov transitions:

(1) **Transitions within a single sub-ensemble:** These transitions can be made with any desired Markovian update procedure that satisfies detailed balance with respect to P^{eq} for its sub-ensemble. In our case such transitions are made with HMC. We refer to the set of HMC trajectories that are performed between swaps as an HMC step.

(2) **Transitions between sub-ensembles:** These transitions are used to connect the phase spaces of the sub-ensembles. Such a transition would be a proposed swap between any two sub-ensembles i and j. Let a be a configuration in sub-ensemble i and b be a configuration in sub-ensemble j. The swap transition can be denoted

$$(a,b)\rightarrow\begin{cases}(b,a) & \text{if swap is accepted,}\\ (a,b) & \text{if swap is rejected.}\end{cases} \tag{7}$$

Let us denote by $P_s(i,j)$ the probability that the swap succeeds. The detailed balance condition is

$$P_s(i,j)e^{-\mathcal{H}_i(a)}e^{-\mathcal{H}_j(b)}=P_s(j,i)e^{-\mathcal{H}_j(a)}e^{-\mathcal{H}_i(b)} \tag{8}$$

as the contributions from the other ensembles cancel on both sides. A suitable choice for P_s is the simple Metropolis [14] acceptance probability

$$P_s(i,j)=\min(1,e^{-\Delta\mathcal{H}}) \tag{9}$$

where

$$\Delta\mathcal{H}=\{\mathcal{H}_j(a)+\mathcal{H}_i(b)\}-\{\mathcal{H}_i(a)+\mathcal{H}_j(b)\} \tag{10}$$

which satisfies the detailed balance condition by construction.

The required overall Markov transition should be constructed of a number of both kinds of transitions. HMC steps within all the sub-ensembles are necessary and sufficient for convergence of the individual sub-ensembles to the required equilibrium probability distributions. Transitions between sub-ensembles are not essential for convergence to the correct distributions, but without such transitions PT would be identical to running several independent HMC simulations.

Swap acceptance rate and action matching

Any extra decorrelation of observables in PT over and above normal HMC must necessarily come from the swapping transitions. Understanding the factors which determine the acceptance rate for swapping transitions is therefore of fundamental importance in determining any possible improvements which might result from applying PT. The swapping probability between two sub-ensembles is determined by the energy change $\Delta\mathcal{H}$ (9) which would result if the swap was accepted. In the approximation where $\Delta\mathcal{H}$ is small, the average acceptance rate, $\langle A\rangle$, for Metropolis-like algorithms of this kind is easily shown to be [15]

$$\langle A\rangle=\mathrm{erfc}\left(\frac{1}{2}\sqrt{\langle\Delta\mathcal{H}\rangle_{12}}\right), \tag{11}$$

where $\langle\Delta\mathcal{H}\rangle$ is the expectation value of $\Delta\mathcal{H}$ in the joint measure of the two sub-ensembles for which the swap is attempted.

To make the notation explicit, let $S_1[U]$ and $S_2[U]$ be the actions of two lattice gauge theories defined on the same phase space but with different parameter values. The individual partition functions corresponding to these actions are

$$\mathcal{Z}_i=\int [dU]\exp\{-S_i[U]\} \qquad i=1,2 \tag{12}$$

and the expectation of an observable $\mathcal{O}$ in ensemble i is

$$\langle\mathcal{O}\rangle_i=\frac{1}{\mathcal{Z}_i}\int [dU]\,\mathcal{O}(U)\exp\{-S_i[U]\}. \tag{13}$$

The partition function for the joint system containing two PT sub-ensembles, one simulated with action S_1, and the second simulated with action S_2 is defined over the direct product of the single system phase space with itself, and is given as

$$\mathcal{Z}_{12}=\int [dU]_a[dU]_b\exp\{-S_1[U_a]\}\exp\{-S_2[U_b]\}. \tag{14}$$

Here the a and b subscripts on the measures label the individual copies of the integration phase space and on the link variables they indicate the copy of the phase space from which the link variables have been drawn.

Expectations of observables defined on the product phase space are denoted as

$$\langle\mathcal{O}\rangle_{12}=\frac{1}{\mathcal{Z}_{12}}\int [dU]_a[dU]_b\,\mathcal{O}(U_a,U_b)\times\exp\{-S_1[U_a]\}\exp\{-S_2[U_b]\}. \tag{15}$$

Straightforward generalizations apply if we include pseudofermions and/or link conjugate momenta as fundamental phase space degrees of freedom.

If the parameters defining the two actions, S_1 and S_2 differ by a small amount only, then the expectation values of an observable in the two measures can be related by a cumulant expansion which, to first order, takes the form

$$\langle \mathcal{O} \rangle_2 = \langle \mathcal{O} \rangle_1 + \langle \tilde{\mathcal{O}} \tilde{\Delta}_{12} \rangle_1 + \cdots \tag{16}$$

where $\Delta_{12} \equiv S_1 - S_2$ and $\tilde{\mathcal{O}} \equiv \mathcal{O} - \langle \mathcal{O} \rangle$ etc.

An interesting feature of this formula is that the first order cumulant expansion term, $\langle \tilde{\mathcal{O}} \tilde{\Delta}_{12} \rangle_1$ is actually measurable numerically. (The first order term is a correlation which has fluctuations of order the system size, so finite but large work is required to evaluate these correlations). This allows the development of numerical techniques to calculate the formal distance in parameter space between two different actions. It also allows the numerical tuning of action parameters to maximize the overlap of specified features of two ensembles generated by different actions. There are a number of applications of such ''action matching'' technology [10].

For example, given two actions with different structures (e.g. one might include a Wilson fermion term, while the second might include a Kogut Susskind fermion term), consider the problem of adjusting the parameters of the second action in any of the following ways:

(1) to match the values of some subset of the observables i.e. require that $\langle \mathcal{O} \rangle_1 = \langle \mathcal{O} \rangle_2$,

(2) to minimize the formal distance between two actions in parameter space,

(3) to maximize the acceptance in an exact algorithm for S_2 constructed via accept/reject step applied to configurations generated with action S_1.

It was shown in [10] that the last two conditions are equivalent to lowest order in a cumulant expansion. Under special circumstances the first condition is also equivalent to the other two to lowest order. The prescriptions differ in a calculable way at the next order.

The relevance of this action matching technology for PT is that it allows us, firstly, to generate estimates for the average acceptance of PT swaps, and secondly, to analyze the dependence of the average acceptance rate on the structure of the actions for the PT sub-ensembles, and on the difference in parameter values between sub-ensembles.

The acceptance of PT swaps is determined by the energy differences before and after the swap. This energy difference is

$$\delta = \Delta \mathcal{H}. \tag{17}$$

The momentum fields cancel exactly in the Hamiltonian terms and one can deal directly with the actions

$$\delta = S_1(U_b, \phi_b) + S_2(U_a, \phi_a) - S_1(U_a, \phi_a) - S_2(U_b, \phi_b). \tag{18}$$

Collecting the terms depending on the same fields one obtains

$$\delta = \Delta_{12}(U_b, \phi_b) - \Delta_{12}(U_a, \phi_a). \tag{19}$$

We now identify δ with $-\delta$ in (3.15) in [10]. Following the analysis of [10] one may obtain the acceptance rate formula of the action matching mechanism

$$\langle A \rangle = \mathrm{erfc}\left(\frac{1}{2} \sqrt{\sigma^2(\Delta_{12})} \right). \tag{20}$$

One can then deduce that

$$\sigma^2(\Delta_{12}) = \langle \Delta \mathcal{H} \rangle \approx \frac{1}{2} \sigma^2(\Delta \mathcal{H}) \tag{21}$$

where the second approximate equality is required to derive the acceptance rate (11).

Our PT parameters were tuned using the action matching technology to maximise the acceptance between two sub-ensembles using the action

$$\mathcal{S}_i = -\beta_i \mathcal{W}_\Box - T_i \tag{22}$$

with

$$T_i = \mathrm{Tr} \ln(Q_i^{-1}) \tag{23}$$

and

$$Q_i = (M^\dagger(\kappa_i) M(\kappa_i))^{-1}. \tag{24}$$

The tuning was carried out before performing the PT simulation using configurations from a preliminary HMC run at the desired reference parameter set. This tuning would have maximized acceptance in the PT swap steps for the implementation where only the links are swapped. [As discussed above, the swap steps for this implementation depend on the differences in the link action including a fermion determinant term, (22).]

However, as also previously discussed the PT implementation we chose involved swapping both links and pseudofermions. The action to be evaluated for this case is

$$\mathcal{S}_i = -\beta_i \mathcal{W}_\Box + \phi^\dagger Q_i \phi. \tag{25}$$

We note that tuning parameters to maximise the PT swap acceptance for Eq. (22) does not necessarily optimize the swap acceptance for the link plus pseudofermion swap case. Since the details are of considerable importance in the following discussions, we present them here.

Consider first the distance σ^2 between actions $\mathcal{S}_i$ where the $\mathcal{S}_i$ are as given by Eq. (22). Then

$$\Delta_{12} = \Delta\beta \mathcal{W}_\Box + \Delta T \tag{26}$$

with

$$\Delta\beta = \beta_2 - \beta_1 \tag{27}$$

$$\Delta T = T_2 - T_1. \tag{28}$$

The variance of Δ_{12} in an individual sub-ensemble is

$$\sigma^2(\Delta_{12})_i = \langle (\Delta\beta \tilde{\mathcal{W}}_\Box + \widetilde{\Delta T})^2 \rangle_i. \tag{29}$$

One can see that for a given ΔT one can tune $\Delta\beta$ to minimize this variance.

However when one examines the case of the pseudofermionic action of Eq. (25) one finds that

$$\Delta_{12}=\Delta\beta\mathcal{W}_{\Box}+\phi^{\dagger}(Q_1-Q_2)\phi. \tag{30}$$

When calculating the variance of Δ_{12} one encounters the quadratic term

$$\langle\phi^{\dagger}(Q_1-Q_2)\phi\phi^{\dagger}(Q_1-Q_2)\phi\rangle_i. \tag{31}$$

This term gives rise to both connected and disconnected pieces when the integration over the pseudofermion fields is carried out

$$\begin{aligned}\langle\phi^{\dagger}(Q_1-Q_2)\phi\phi^{\dagger}(Q_1-Q_2)\phi\rangle_i &=\langle \mathrm{Tr}^2((Q_1-Q_2)Q_i^{-1})\rangle_i^U \\ &+\langle \mathrm{Tr}(Q_1-Q_2)Q_i^{-1}(Q_1-Q_2)Q_i^{-1}\rangle_i^U.\end{aligned} \tag{32}$$

Here the superscript U on the expectations indicates that they are now to be carried out over the gauge fields only. Hence one finds that

$$\begin{aligned}\sigma_i^2(\Delta_{12})=\langle[\Delta\beta\tilde{\mathcal{W}}_{\Box}+\overbrace{\mathrm{Tr}((Q_1-Q_2)Q_i^{-1})}]^2\rangle_i^U \\ +\langle \mathrm{Tr}(Q_1-Q_2)Q_i^{-1}(Q_1-Q_2)Q_i^{-1}\rangle_i^U.\end{aligned} \tag{33}$$

We also note that to first order in Q_1-Q_2

$$\Delta T\approx\mathrm{Tr}((Q_1-Q_2)Q_i^{-1}). \tag{34}$$

Comparing Eq. (29) and (33) it can be seen that using a pseudofermionic action gives rise to a connected piece in $\sigma_i^2(\Delta_{12})$ which one would not get using the action of Eq. (22). This connected piece cannot be tuned away by changing $\Delta\beta$ and it increases the distances in parameter space compared to when the action of Eq. (22) is used. If parameters are tuned using the action of Eq. (22) and the simulation is carried out using pseudofermions the acceptance rate of the swaps will not be optimized.

III. AUTOCORRELATIONS

The cost of measuring observables

The gain from PT is expected to come from the swapping of configurations between sub-ensembles. This reduction in autocorrelation time is supposed to occur due to the fact that the sub-ensembles are simulated (between swaps) with independent Markov processes. However the swaps couple the ensembles and include cross correlations between them. Thus care must be taken when using results from separate sub-ensembles together.

According to [16,17] if successive measurements of $\mathcal{O}$ are correlated, the sample mean $\overline{\mathcal{O}}$ is given (we use the convention of [16]) by

$$\overline{\mathcal{O}}=\langle\mathcal{O}\rangle\pm\sqrt{\frac{2\tau_{\mathcal{O}}+1}{N}\sigma^2(\mathcal{O})}. \tag{35}$$

Here, $\sigma^2(\mathcal{O})$ is the variance of operator $\mathcal{O}$ given by

$$\sigma^2(\mathcal{O})=\langle\mathcal{O}^2\rangle-\langle\mathcal{O}\rangle^2 \tag{36}$$

and $\tau_{\mathcal{O}}$ is the integrated autocorrelation time, defined as

$$\tau_{\mathcal{O}}=\sum_{t=1}^{\infty}C_{\mathcal{O}}(t) \tag{37}$$

and where $C_{\mathcal{O}}(t)$ is the normalized autocorrelation function

$$C_{\mathcal{O}}(t)=\frac{1}{\sigma^2(\mathcal{O})}\langle(\mathcal{O}(t+I)-\langle\mathcal{O}\rangle)(\mathcal{O}(I)-\langle\mathcal{O}\rangle)\rangle \tag{38}$$

and the expectation values are over all pairs of $\mathcal{O}_i$ separated by an interval t. From now on we shall drop the subscript $\mathcal{O}$ from these formulas except where necessary. Furthermore the term ''autocorrelation time'' will always be used to refer to the integrated autocorrelation time.

The practical meaning of the statements above is that $2\tau+1$ correlated measurements of $\mathcal{O}$ are needed in order to reduce the error in $\overline{\mathcal{O}}$ by the same amount as if two uncorrelated measurements were used. Markov methods in general produce correlated sequences of configurations, and hence correlated sequences of measured observables. The integrated autocorrelation time τ is therefore an important indicator of the performance of a Monte Carlo simulation that is carried out with the intention of measuring observable $\mathcal{O}$.

In particular, if one assumes that the autocorrelation function decays exponentially

$$C(t)=\exp\{-kt\} \tag{39}$$

with $k>0$, one finds that

$$\exp\{-k\}=\frac{\tau}{\tau+1} \tag{40}$$

which is a result we shall use later.

Autocorrelations in twin sub-ensemble PT

We are interested in whether or not PT will reduce the integrated autocorrelation time of an observable measured on an ensemble with some parameter set relative to the corresponding autocorrelation time of the same observable measured on an ensemble generated at the same parameters using HMC. We refer to the former of these autocorrelation times as the PT autocorrelation time and the latter as the HMC autocorrelation time.

Let us examine the situation of a PT system with two sub-ensembles. Sub-ensemble 1 has the desired parameter set, and the other sub-ensemble has its parameters chosen so as to give some acceptance rate $\langle A\rangle$. We assume that the HMC autocorrelation functions of both ensembles are the same. We demonstrate in Sec. V that over the distances in parameter space for which we can use PT, and with the statistics available, we cannot differentiate between the autocorrelation times of the plaquette operator between sub-ensembles, so we regard the above assumption as reasonable.

Having made the above assumption, the changes in the autocorrelation time due to PT are now controlled solely by the number of successful swaps between the sub-ensembles. The swap probability in general depends on the particular PT state at which the swap is attempted, but for simplicity we assume that we can replace individual swap probabilities with the overall average swap probability which is none other than the acceptance rate $\langle A\rangle$.

Let the HMC autocorrelation function be denoted $C_H(t)$, and the PT autocorrelation function of the sub-ensemble of interest be denoted $C_{PT}(t)$. Consider the connected autocorrelation function (38) written in the more compact notation introduced in Eq. (16)

$$C_H(t)=\frac{1}{\langle\tilde{\mathcal{O}}^2\rangle}\sum_{i=0}^{n-t}\tilde{\mathcal{O}}_{i+t}\tilde{\mathcal{O}}_i \tag{41}$$

where n is the number of samples of $\tilde{\mathcal{O}}_i$.

The autocorrelation function in the PT ensemble of interest can now be written as

$$C_{PT}(t)=\frac{1}{\langle\tilde{\mathcal{O}}^2\rangle}\{S_e+S_o\} \tag{42}$$

where

$$S_e=\sum_{\text{even}}\tilde{\mathcal{O}}_{i+t}\tilde{\mathcal{O}}_i, \tag{43}$$

$$S_o=\sum_{\text{odd}}\tilde{\mathcal{O}}_{i+t}\tilde{\mathcal{O}}_i. \tag{44}$$

By the even sum we mean that the only terms contributing to the sum are those where an even number of swaps succeeded out of the t tried between the measurements of $\tilde{\mathcal{O}}_{i+t}$ and $\tilde{\mathcal{O}}_i$.

Given some configuration in one sub-ensemble, after an odd number of successful swaps it can only be in the other one. As the HMC steps are independent in different sub-ensembles, we expect (to a first approximation) no correlation between configurations in a sub-ensemble that are separated by an odd number of swaps. Hence we assume that S_o sums to zero and we consider only the S_e term.

We then rewrite Eq. (42) as

$$C_{PT}=P_eC_H(t) \tag{45}$$

where P_e is the probability that an even number of successful swaps occur in t trials. P_e is given by

$$P_e=\sum_i C_i^t(1-\langle A\rangle)^{t-i}\langle A\rangle^i \tag{46}$$

where the index i runs from 0 to the largest even integer less than or equal to t, i is even and C_i^t is the number of ways of choosing i swaps from t.

Carrying out the sum in Eq. (46) one finds

$$P_e=\frac{1}{2}\{1+(1-2\langle A\rangle)^t\} \tag{47}$$

leading to the result

$$C_{PT}(t)=\frac{1}{2}\{1+(1-2\langle A\rangle)^t\}C_H(t). \tag{48}$$

We consider three separate cases:

(i) $\langle\mathbf{A}\rangle=\mathbf{0}$: In this case $C_{PT}(t)=C_H(t)$, which is what we expect when we do not carry out any successful swaps.

(ii) $\mathbf{0}<\langle\mathbf{A}\rangle\leqslant\frac{1}{2}$: In this case $C_{PT}\in[\frac{1}{2}C_H,C_H)$ and we can see a reduction in the autocorrelation function of at most a factor of 2.

(iii) $\frac{1}{2}<\langle\mathbf{A}\rangle\leqslant\mathbf{1}$: In this case the term $(1-2\langle A\rangle)^t$ in Eq. (48) becomes oscillatory. In particular if $\langle A\rangle=1$ (every swap succeeds) it is impossible to get an even number of successful swaps out of an odd number of trials, whereas it is a certainty for an even number of trials.

If one models the autocorrelation function by an exponential decay as in Eq. (39), it is possible to calculate the PT integrated autocorrelation time for the ensemble:

$$\tau_{PT}=\sum_1^\infty C_{PT}(t) \tag{49}$$

$$=\frac{1}{2}\tau_H+\frac{1}{2}\sum_1^\infty((1-2\langle A\rangle)\exp\{-k\})^t \tag{50}$$

$$=\frac{\tau_H[1+\langle A\rangle(\tau_H-1)]}{1+2\langle A\rangle\tau_H} \tag{51}$$

where the last line follows from using Eq. (40), summing the resulting geometric series and simplifying. The ratio of τ_{PT} to τ_H is then

$$\frac{\tau_{PT}}{\tau_H}=\frac{1+\langle A\rangle(\tau_H-1)}{1+2\langle A\rangle\tau_H}. \tag{52}$$

We remark on several features of the ratio in Eq. (52):

(i) When $\langle A\rangle=0$, one is, in effect, carrying out two uncoupled HMC simulations and the autocorrelation times in each sub-ensemble remain the same as they would be for HMC simulations.

(ii) For a fixed $\langle A\rangle\in(0,\frac{1}{2})$ increasing τ_H from 0 has the effect that the ratio of Eq. (52) approaches the value of $\frac{1}{2}$ *from above*. The closer $\langle A\rangle$ is to $\frac{1}{2}$, the faster this limit is approached. If one is interested in both sub-ensembles this is still a gain. If one of the two ensembles serves only to decorrelate the other and is not otherwise interesting (it is thrown away at the end) then one would lose over HMC as one would have done twice the work, but gained less than a factor of two.

(iii) For $\langle A\rangle=\frac{1}{2}$ the ratio is exactly $\frac{1}{2}$ and a break-even is reached, in the sense that one does the work of two simulations, but in each sub-ensemble the integrated autocorrelation is halved. This is the stage when a sub-ensemble which

originally served no other purpose than to help decorrelate the other one may be thrown away without losing out.

(iv) For $\langle A\rangle \in (\frac{1}{2},1]$ the ratio approaches $\frac{1}{2}$ rapidly *from below*. In this case one clearly wins even if one is only interested in a single sub-ensemble. However the gain is not much, as for any reasonable value of τ_H the ratio will have already approached the asymptotic limit of $\frac{1}{2}$ to a good level of accuracy.

One can therefore win most with PT when the acceptance rate is very high, and the observable of interest has a very short autocorrelation time. In such a situation it is possible to gain more than a factor of two over the HMC autocorrelation time in each ensemble if the swap acceptance rate is greater than $\frac{1}{2}$. However if an observable has such a short HMC autocorrelation time, it may not be worthwhile employing PT. In a typical situation, it would be expected that the gain in each ensemble is very close to a factor of 2.

IV. SIMULATION DETAILS

Our PT simulations were carried out on the PPARC Cray T3E facility in Edinburgh. Code for performing the HMC trajectories was taken from the GHMC code written for the UKQCD Dynamical Fermions project, described in [18].

Program features

The PT code ran trajectories on each sub-ensemble in *series*. Sub-ensembles were labelled from 0 to $N-1$, where N was the total number of sub-ensembles. Swaps of configurations between sub-ensembles were attempted according to a boolean *plan matrix M*. If, after carrying out the HMC trajectories in sub-ensemble i, the element M_{ij} was found to contain *true*, the code would attempt to swap configurations j and $j+1$. ($j\in[0,N-2]$) The default matrix had all its elements set to *false* except for the last row which had all its elements set to *true*. This way the program would perform all the HMC trajectories on all the ensembles and would then attempt a chain of pairwise swaps.

The number of HMC trajectories per sub-ensemble was controlled through an independent parameter file for each sub-ensemble. This way a sub-ensemble could be equilibrated with the GHMC code and if desired, it could easily be taken and further evolved on its own using the GHMC code. Likewise each sub-ensemble had an associated set of log files for the plaquette and for solver statistics. The overall driver routine kept a log file of the success or failure of swap attempts and the swap energies.

Simulation parameters

Five PT simulations $S1$, $S2$, $S3$, $S4$ and $S5$ were performed, each of which comprised two sub-ensembles. The parameters for these simulations are shown in Table I. In all five simulations one sub-ensemble had parameters ($\beta=5.2, c=2.0171, \kappa=.13300$). The parameters for the second sub-ensemble were given by action matching for $S1$, $S2$ and $S3$, while for $S4$ and $S5$ only κ was varied. Thus we could investigate the PT swap acceptance rate for different distances in parameter space.

TABLE I. Simulation parameters used for twin ensemble runs and the reference HMC run.

[h] Simulation	(β_1,c_1,κ_1)	(β_2,c_2,κ_2)
HMC	(5.2,2.0171,0.133)	
$S1$	(5.2,2.0171,0.133)	(5.2060,2.01002,0.13280)
$S2$	(5.2,2.0171,0.133)	(5.2105,2.00471,0.13265)
$S3$	(5.2,2.0171,0.133)	(5.2150,1.99940,0.13250)
$S4$	(5.2,2.0171,0.133)	(5.2,2.0171,0.13280)
$S5$	(5.2,2.0171,0.133)	(5.2,2.0171,0.13265)

We also had data from a previous HMC simulation with parameters ($\beta=5.2, c=2.0171, \kappa=.13300$) on lattices of volume $8^3\times16$ and $8^3\times24$.

The results from the reference run on the $8^3\times16$ lattice were used to validate the PT code. Our PT simulations were also carried out on lattices of this size. Furthermore, it was possible to compare the autocorrelation times of the plaquette from this HMC run with the autocorrelation times of the plaquette from the first sub-ensembles of the PT runs. For the second sub-ensembles, the GHMC code was used only to achieve equilibration. Thus there is insufficient data to calculate the HMC autocorrelation times of the second sub-ensembles.

In the PT simulations each HMC step was one trajectory long. The plan matrix used was the default one described earlier. Simulations $S1$, $S2$ and $S3$ ran for 6000 swap attempts giving 6000 trajectories for each sub-ensemble, while $S4$ and $S5$ ran for only 1000 swap attempts due to time constraints.

The matching procedure was performed using reference HMC results from $8^3\times24$ lattices, and the methods outlined in [12].

Analysis

We examined the acceptance rate as a function of the average swap energy change $\langle\Delta\mathcal{H}\rangle$, and of $\Delta\kappa=\kappa_2-\kappa_1$, the change in the hopping parameters. We investigated the autocorrelation time of the average plaquette.

Errors in ensemble averages were estimated using the bootstrap method. Autocorrelations were estimated using the sliding window scheme of Sokal *et al.* [17].

V. RESULTS

A summary of our results is shown in Table II. We show for each simulation $\Delta\beta=\beta_2-\beta_1$, the corresponding $\Delta\kappa$, $\langle\Delta\mathcal{H}\rangle$, the acceptance rate $\langle A\rangle$, the integrated autocorrelation time τ for the plaquette in sub-ensemble 1 and the autocorrelation time in sub-ensemble 1 divided by the HMC autocorrelation time, $\tau_1/\tau_{\mathrm{HMC}}$.

Swap acceptance rate

Figure 1 shows the measured swap acceptance rates of the simulations. The solid line is the acceptance rate formula in

TABLE II. Results from the PT simulations showing the appropriate results from HMC for comparison.

Simulation	$\Delta\beta(\times 10^{-3})$	$\Delta\kappa(\times 10^{-4})$	$\langle\Delta\mathcal{H}\rangle$	$\langle A\rangle$	τ_1	$\tau_1/\tau_{\mathrm{HMC}}$
HMC					26(6)	1
$S1$	6	-2.0	1.23(2)	0.43(1)	12(3)	0.5(2)
$S2$	10.5	-3.5	3.76(4)	0.17(1)	19(4)	0.7(2)
$S3$	15	-7.5	7.64(6)	0.051(2)	24(6)	0.9(3)
$S4$	0	-2.0	0.91(4)	0.49(1)	9(4)	0.3(2)
$S5$	0	-3.5	2.29(7)	0.26(2)	18(10)	0.7(4)

Eq. (11). It can be seen that the measured results are in excellent agreement with its predictions.

Calibration and matching

It can be seen from Table II that simulations $S2$ and $S3$ which had parameters given by matching the Tr ln actions of Eq. (22) have lower acceptance rates than $S4$ and $S5$ for which tempering was carried out only in κ. We expect that this is due to the noise term of Eq. (33) and is the result of using the pseudofermion action for calculating the swap energy differences.

To see how large the effect of this noise term is, we can compare the residual variance $\sigma^2(\Delta_{12})$ from the matching procedure [12], using the Tr ln action with the variance as measured in our PT simulations through $\langle\Delta\mathcal{H}\rangle$. Note that we only have biased estimators for $\sigma^2(\Delta_{12})$ from the matching procedure, and that we have calculated the residual variance estimate only for $\Delta\kappa=0.0005$.

Table III contains our comparison of the Tr ln matching predictions and pseudofermionic measurements for simulation $S3$. We can see in column 2, our biased estimate of the residual variance on matching and in column 4 the corresponding predicted acceptance rate. In column 3 we see the actual variance as measured in the simulation and in column 5 the corresponding measured acceptance rate. We expect the difference in the variances to be due to the four point term in equation Eq. (33). We can therefore numerically estimate the four point term to be

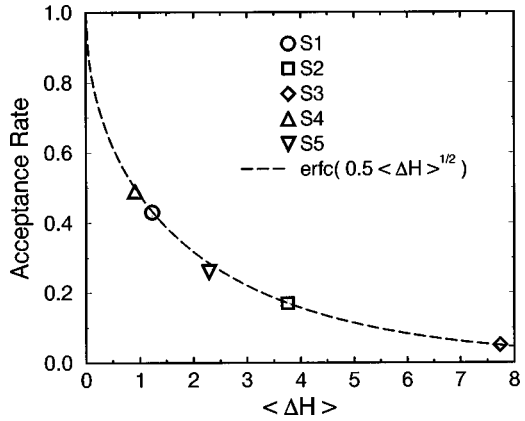

FIG. 1. Acceptance rate against $\langle\Delta\mathcal{H}\rangle$. Error bars are smaller than the symbols.

$$\langle \mathrm{Tr}(Q_2-Q_1)Q_i^{-1}(Q_2-Q_1)Q_i^{-1}\rangle_i^U = 6.6(2) \qquad (53)$$

for simulation $S3$.

Note that if during our swap acceptance steps, we were to discard the pseudofermion fields, and calculate the energy change using the Tr ln action by the methods outlined in [12], we would suffer a workload hit due to the additional work required to calculate the necessary fermion determinants, but would expect an acceptance rate of around 48% in the case of simulation $S3$. Thus, using pseudofermions was a poor way to proceed originally. However as the action difference scales like the lattice volume, going to larger lattices would effectively cancel all the gain one could obtain by using the Tr ln action to evaluate the swap action or energy difference.

Autocorrelation times and efficiency

The autocorrelation times of the plaquette operator on the sub-ensembles with parameter $\kappa=.1330$ are shown in column 5 of Table II. We also show for comparison the autocorrelation time estimated from our independent HMC run at the same parameter set. In Table IV we gather some estimates of the integrated autocorrelation time of the plaquette for some independent HMC runs at similar parameters to our PT runs. It can be seen that the HMC autocorrelation times agree with each other within estimated errors, justifying the assumptions of our model of Sec. II.

Figure 2 shows the ratio of PT to HMC autocorrelation times. The errors on the ratios were obtained by simple error combination. The line superimposed on the data in Fig. 2 is the prediction of the model in Sec. II [cf. Eq. 52]. It can be seen that it is not inconsistent with the data.

VI. SUMMARY AND CONCLUSIONS

In this paper we presented our study of the parallel tempering algorithm applied to lattice QCD with $O(a)$-improved Wilson fermions. We showed how the algo-

TABLE III. Comparison of Tr ln matching and acceptance with pseudofermionic acceptance.

Simulation	$\sigma^2(\Delta_{12})_{\mathrm{Tr\,ln}}$	$\sigma^2(\Delta)_{\mathrm{p.f}}=\langle\Delta\mathcal{H}\rangle$	$\langle A\rangle_{\mathrm{Tr\,ln}}$	$\langle A\rangle_{\mathrm{p.f}}$
$S3$	1.02(20)	7.64(6)	0.48(5)	0.051(2)

TABLE IV. The integrated autocorrelation times of some other simulations.

β	c	κ	τ_{HMC}
5.2	1.99	0.1335	18(8)
5.2	2.0171	0.1330	26(6)
5.232	1.98	0.1335	20(6)

rithm satisfies detailed balance, and gave a formula for the swap acceptance rate in terms of the swap energy change $\Delta\mathcal{H}$. We highlighted the connection between PT and the technology of action matching. We presented and discussed a simple model of autocorrelations in a twin sub-ensemble PT system, and found that the algorithm is unlikely to improve autocorrelation times by more than a factor of two for such a system. We verified our simple model assumptions by gathering autocorrelation time data from previous simulations.

We carried out a numerical study where we verified the acceptance formula and the predictions of the autocorrelation model within statistical errors. We also obtained information on how the acceptance rate of the algorithm falls with increasing $\Delta\kappa$.

We found that using the pseudofermions from HMC in the swap attempt is a poor way to proceed if the parameters are matched for the Tr ln action. We have shown analytically that there is an extra noise term in the definition of the distance between actions when pseudofermions are used. We have attempted to estimate the size of this noise term numerically.

We conclude that parallel tempering does not seem to give any real gain over HMC at the present time for simulating lattice QCD. We were unable to use PT to simulate sub-ensembles sufficiently far apart in parameter space. The acceptance rate drops too quickly with $\Delta\kappa$. This situation could be alleviated somewhat if the swap action or energy differences were to be calculated using the Tr ln action, for simulations with parameters matched with that action. However in the end the real problem is that the swap action or energy change scales with the volume for a fixed kappa, and that when employing the PT algorithm on a realistic sized (e.g. $16^3\times 32$) lattice, the scaling of the swap energy change would lower the acceptance rate and lose all that could be gained by using the Tr ln action.

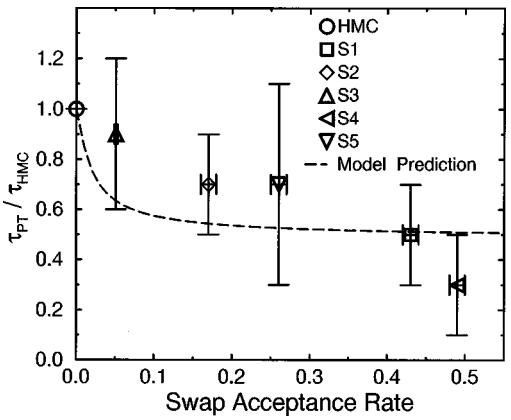

FIG. 2. Integrated autocorrelation times for the plaquette normalized by that from GHMC simulations.

We note that the above failings are problems inherent to lattice QCD and its simulation by HMC. They are not inherent problems of the PT approach. Indeed given a system where some interesting observable has autocorrelation times which decrease rapidly with increases in the distance between sub-ensembles (with distance being defined in the context of action matching), the PT approach may be highly successful. However lattice QCD appears not to be such a system.

Thus we were unable to take advantage of the fact that in one region of parameter space autocorrelation times are short while in another they are long. With our parameter values, the HMC autocorrelation times of our sub-ensembles are the same within experimental errors and the predictions of our model apply. A chain of sub-ensembles that would span the required distance in parameter space can be constructed, but would take an unfeasibly large number of sub-ensembles for lattices of interesting size.

ACKNOWLEDGMENTS

We gratefully acknowledge support from PPARC grant no. GR/L22744, and EPSRC for funding under grant number GR/K41663. We also wish to thank Tony Kennedy and Stephen Booth for helpful discussions.

[1] S. Duane, A.D. Kennedy, B.J. Pendleton, and D. Roweth, Phys. Lett. B **195**, 216 (1987).
[2] B. Allés, G. Boyd, M. D'Elia, A. Di Giacomo, and E. Vicari, Phys. Lett. B **389**, 107 (1996).
[3] B. Allés, G. Bali, M. D'Elia, A. Di Giacomo, N. Eicker, S. Güesken, H. Höeber, Th. Lippert, K. Schilling, A. Spitz, T. Struckmann, P. Ueberholz, and J. Viehoff, Phys. Rev. D **58**, 071503 (1998).
[4] K. Hukushima and J. Nemoto, cond-mat/9512035.
[5] E. Marinari, cond-mat/9612010.
[6] E. Marinari, G. Parisi, and J. Ruiz-Lorenzo, ''Spin Glasses and Random Fields,'' edited by P. Young, cond-mat/9701016.
[7] G. Boyd, Nucl. Phys. B (Proc. Suppl.) **60A**, 341 (1998).
[8] B. Sheikholeslami and R. Wohlert, Nucl. Phys. **B259**, 572 (1985).
[9] K. Jansen and R. Sommer, Nucl. Phys. B (Proc. Suppl.) **63A-C**, 853 (1998).
[10] A.C. Irving and J.C. Sexton, Phys. Rev. D **55**, 5456 (1997).
[11] A.C. Irving, J.C. Sexton, and E. Cahill, Nucl. Phys. B (Proc. Suppl.) **63A-C**, 967 (1998).
[12] UKQCD Collaboration, A.C. Irving, J.C. Sexton, E. Cahill, J. Garden, B. Joó, S.M., Pickles and Z. Sroczynski, Phys. Rev. D **58**, 114504 (1998).
[13] R. Sommer, Nucl. Phys. **B411**, 839 (1994).

[14] N. Metropolis, A.W. Rosenbluth, M.N. Rosenbluth, A.H. Teller, E. Teller, J. Chem. Phys. **21**, 1087 (1953).
[15] S. Gupta, A. Irbäck, F. Karsch, and B. Petersson, Phys. Lett. B **242**, 437 (1990).
[16] I. Horváth, and A.D. Kennedy, Nucl. Phys. **B510**, 367 (1998).
[17] N. Madras, and A.D. Sokal, J. Stat. Phys. **50**, 109 (1988).
[18] UKQCD Collaboration Z. Sroczynski, S.M. Pickles, and S.P. Booth, Nucl. Phys. B (Proc Suppl.) **63A-C**, 949 (1998).