Tuning the hybrid Monte Carlo algorithm

Rajan Gupta
Theoretical Division, Los Alamos National Laboratory, Los Alamos, New Mexico 87545

Gregory W. Kilcup
Physics Department, Brown University, Providence, Rhode Island 02912
and Boston University, Boston, Massachusetts 02215

Stephen R. Sharpe
Stanford Linear Accelerator Center, Stanford University, Stanford, California 94309
(Received 4 April 1988)

We apply the hybrid Monte Carlo algorithm to pure gauge QCD and to QCD with four flavors of staggered dynamical fermions. We show how the acceptance in the global Metropolis step depends upon the parameters of the algorithm. By tuning the values of the coupling constants in the algorithm to be different from those in the global Metropolis step, we find that the acceptance can be kept large without having to make the step size prohibitively small. We give an analytic discussion of the tuning, and argue that the algorithm requires computer time growing proportional to $V^{5/4}$.

I. INTRODUCTION

The greatest obstacle to the simulation of QCD is the inclusion of dynamical fermions. This can be done either with approximate or with exact algorithms. The advantage of the approximate algorithms is that the time required grows with a smaller power of the volume. The disadvantage is that these algorithms have systematic errors which can only be studied by sending $\epsilon \to 0$, ϵ being the small parameter characterizing the approximation. Unless we can prove that mass ratios are not affected by these ϵ-dependent errors, we have to make several runs at different values of ϵ and extrapolate $\epsilon \to 0$. The volume dependence may be worse for the exact algorithms, but the fact that there is no overhead to be paid for extrapolation implies that an exact algorithm can be competitive on lattice up to a certain size. To decide whether this is true for lattices accessible in current numerical simulations, one must resort to field testing.

To this end we have investigated the hybrid Monte Carlo algorithm[1] (HMCA). This is an exact fermion algorithm of the kind proposed in Ref. 2. The basic evolution through phase space is controlled by the hybrid algorithm[3,4] (HA). The HA is an efficient blend of molecular dynamics[5] (MD) and Langevin algorithms.[6,7] The evolution of the molecular dynamics equations provides fast motion through phase space. The noise in the Langevin equation ensures ergodicity. The HMCA is made exact by using the HA to propose changes in the links of the entire lattice and then doing a global Metropolis accept and/or reject.

The question we attempt to answer in this paper is whether by tuning the HMC algorithm we can maintain reasonable acceptance rates on lattices large enough to extract useful physical information. We would like to provide simple rules for choosing the tunable parameters. We have performed extensive numerical tests for the SU(3) pure gauge theory and for QCD with four staggered flavors of dynamical quarks. For the pure gauge theory we have worked on lattices of sizes 4^4, 6^4, and 9^4. For QCD we have data on lattices on size 4^4, $6^3\times 4$, and $8^3\times 4$.

In order to understand how our results will extrapolate to larger volumes we have carried out an analytic investigation of the dependence of the acceptance on ϵ and the volume V. This extends the work of Creutz.[8] We find that ϵ must be scaled as $V^{-1/4}$ in order to keep constant acceptance, and that the computer time needed to generate uncorrelated configurations scales as $V^{5/4}$. This should be compared to the linear dependence on V expected for the approximate algorithms. As a spin-off from this analysis we are able to predict the direction in which the parameters must be tuned for the different variants of the HMC algorithm which we use.

Using the tuned HMC algorithm we have studied four-flavor QCD at finite temperature and our results confirm a first-order chiral transition. These are reported in the following paper.[9]

II. THE HYBRID MONTE CARLO ALGORITHM

In this section we describe the HMC algorithm for full QCD. The algorithm for pure gauge theory can be obtained by simply deleting the fermionic parts of the equations. We are using the Φ algorithm of Gottlieb *et al.*,[10] and we refer the reader to their work for a complete discussion.

The partition function for QCD in Euclidean space is

$$Z = \int \mathcal{D}\psi\,\mathcal{D}\psi^\dagger \mathcal{D}U \exp(-S_G + \bar{\psi} M \psi)\ , \tag{2.1}$$

which can be rewritten in terms of pseudofermion fields living only on even sites:

$$Z=\int \mathcal{D}U\,\mathcal{D}\phi_e\mathcal{D}\phi_e^\dagger \exp[-S_G-\phi_e^\dagger(M^\dagger M)^{-1}\phi_e] \equiv \int \mathcal{D}U\,\mathcal{D}\phi_e\mathcal{D}\phi_e^\dagger \exp(-S_G-S_{\rm PF})\ . \tag{2.2}$$

S_G is the gauge action (which in the present study is the simple Wilson action), and M for staggered fermions is given by

$$M[U]_{i,j}=m\delta_{ij}+\tfrac{1}{2}\sum_\mu \eta_{i,\mu}(U_{i,\mu}\delta_{i,j-\mu}-U^\dagger_{i-\mu,\mu}\delta_{i,j+\mu})\ , \tag{2.3}$$

where m is the quark mass, $U_{i\mu}$ are the link variables, and $\eta_{i,\mu}$ are the staggered fermion phases. In all of the following we shall absorb these phases into the definition of the gauge fields.

The HMC algorithm generates gauge configurations and ϕ_e fields which are distributed exactly according to the measure given in (2.2). This is done by first suggesting a change to the entire lattice using a "preprocessor," in this case, the Φ algorithm. This change is then globally accepted or rejected in the Metropolis step. We describe these two parts of the algorithm in turn.

A. The hybrid algorithm

In the Φ algorithm one rewrites the partition function as

$$Z=\int \mathcal{D}\phi_e\mathcal{D}U\,\mathcal{D}P\, e^{-H_\Phi}\ , \tag{2.4}$$

where the Hamiltonian is

$$H_\Phi=\tfrac{1}{2}{\rm Tr}\sum P_{i,\mu}^2+\frac{\beta}{3}{\rm Re\,Tr}\sum(1-U_p) +\phi_e^\dagger(M^\dagger M)^{-1}\phi_e\ . \tag{2.5}$$

Here U_p is the plaquette. The form (2.4) clearly yields the same correlation functions of U and ϕ_e as does (2.2). The traceless Hermitian $P_{i,\mu}$, which live on sites, are interpreted as the momenta conjugate to $U_{i,\mu}$. The ϕ fields are given no dynamics and thus have no conjugate momenta. Setting the ϕ fields only on the even sites reduces the number of fermion flavors from eight to four, which is the minimum number one can describe with a local staggered Hamiltonian.

The partition function (2.4) is simulated in three stages. First, pseudofermion fields are generated according to the distribution in (2.5). This is done by generating a vector of random complex numbers r living on all sites and distributed according to $P(r)\propto\exp(-r^\dagger r)$. The ϕ_e are then given by

$$\phi=M^\dagger r \Longrightarrow P(\phi_e)\propto e^{-\phi_e^\dagger(M^\dagger M)^{-1}\phi_e}\ . \tag{2.6}$$

Thus, the contribution of fermions to the action is represented by Gaussian noise.

The second stage of the algorithm is the generation of the conjugate momenta. This is done by writing

$$P_{i,\mu}=\sum_{a=1}^{8} r^a_{i,\mu}\lambda_a\ , \tag{2.7}$$

where the real variables $r^n_{i\mu}$ are drawn from the probability distribution $P(r)\propto\exp(-r^2)$, and the Gell-Mann matrices are normalized as ${\rm Tr}(\lambda_a\lambda_b)=2\delta_{ab}$. This "refreshing" of the momenta ensures ergodicity.

The third and final stage of the algorithm is the MD evolution. The gauge fields and their conjugate momenta are evolved, keeping ϕ_e fixed, in a manner which preserves H_Φ and the differential volume in configuration space. The MD equation for the gauge fields is

$$\dot U_{i,\mu}(t)=iP_{i,\mu}(t)U_{i,\mu}(t)\ , \tag{2.8}$$

where t is the MD time. This preserves U as an element of SU(3). The equation for P breaks up into two parts. For links starting at even sites,

$$i\dot P_{i,\mu}=\left[-\frac{\beta}{3}U_{i,\mu}V_{i,\mu} +2U_{i\mu}\left[\sum_\nu U_{\iota+\mu,\nu}T_{i+\mu+\nu,i} -\sum_\nu U^\dagger_{i+\mu-\nu,\nu}T_{i+\mu-\nu,i}\right]\right]_{\rm TA}\ , \tag{2.9}$$

while for links starting at odd sites,

$$i\dot P_{i,\mu}=\left[-\frac{\beta}{3}U_{i,\mu}V_{i,\mu}+2U_{i,\mu}\left[\sum_\nu T_{i+\mu,i-\nu}U_{i-\nu,\nu} -\sum_\nu T_{i+\mu,i+\nu}U^\dagger_{i,\nu}\right]\right]_{\rm TA}\ . \tag{2.10}$$

Here $V_{i,\mu}$ is the sum of the six staples adjoining the link $U_{i,\mu}$,

$$T_{i,j}\equiv X_iX_j^*,\quad X\equiv(M^\dagger M)^{-1}\phi_e\ ,$$

and TA stands for the traceless anti-Hermitian part.

To implement this algorithm numerically, we must discretize the differential equations. This introduces a MD time step ϵ. We use a leap-front updating scheme, which has errors of $O(\epsilon^2)$ (Ref. 10). Given an initial gauge configuration U at time t, the recipe we use is as follows.

(1) Refresh the momenta P and generate the ϕ_e fields.

(2) Update $U(t)$ to $U(t+\frac{1}{2}\epsilon)$ using

$$U(t+\tfrac{1}{2}\epsilon)=\exp[i\tfrac{1}{2}\epsilon P(t)]U(t)\ .$$

We approximate the exponential by a fourth-order polynomial and reunitarize the resulting matrix.

(3) Calculate $\dot P$ at $t+\frac{1}{2}\epsilon$ using Eqs. (2.9) and (2.10). This is the most time-consuming part of the calculation because it requires calculating $(M^\dagger M)^{-1}\phi_e$. We do this using the conjugate-gradient (CG) algorithm.

(4) Calculate $P(t+\epsilon)=P(t)+\epsilon\dot P(t+\frac{1}{2}\epsilon)$.

(5) Update U to $t+\frac{3}{2}\epsilon$ using $P(t+\epsilon)$ as in step (2). Thereafter, steps (3)–(5) are repeated $n_{\rm MD}$ times. At the end one has $P(t+n_{\rm MD}\epsilon)$ and $U(t+(n_{\rm MD}-\frac{1}{2})\epsilon)$. To complete the MD evolution, we calculate

$$U(t+n_{\rm MD}\epsilon)=\exp[i\tfrac{1}{2}\epsilon P(t+n_{\rm MD}\epsilon)]U(t+(n_{\rm MD}-\tfrac{1}{2})\epsilon)\ .$$

This entire sequence of changes is then accepted or re-

jected according to the criterion described below.

We refer to the above as "leapfrog method 1." It is the method we use for most of our studies. It differs from the method used in Refs. 4, 1, and 10, in which one first moves P forward by $\epsilon/2$, and then leapfrogs U to $t+\epsilon$, etc. This we call "leapfrog method 2." A final method, "method 1/2," is to first do $n_{\rm MD}$ steps of method 1, then $n'_{\rm MD}$ steps of method 2, and then accept or reject the entire change.

B. The global metropolis step

In the Metropolis method[1] one proposes small changes in the configurations, and accepts these changes with a probability

$$P=\min\left[1,\frac{P(C'\to C)e^{-S(C')}}{P(C\to C')e^{-S(C)}}\right]. \tag{2.11}$$

Here C' is the trial configuration, $S(C')$ its action, $P(C\to C')$ is the probability of proposing the change given that one is at C, and $P(C'\to C)$ is the reverse probability. If these two probabilities are equal, as they are in most algorithms used to simulate pure gauge theories, changes which lower the action are always accepted, while those which increase it are accepted only conditionally.

In the HMC algorithm a configuration consists of U and ϕ_e, and the action to be simulated is S_G+S_{PF}. A new configuration is suggested using the hybrid algorithm as described above. Since the MD equations are deterministic, the probability of suggesting the change is given by the probability of choosing the initial momenta

$$P(U\to U')=Ne^{-{\rm Tr}(P^2)/2}, \tag{2.12}$$

where N is a normalization constant. The leapfrog discretization of the differential equations is exactly reversible; given that (U,P) evolves into (U',P'), $(U',-P')$ will evolve into $(U,-P)$ even when using a finite step size. The probability of suggesting the inverse change is, therefore, precisely

$$P(U'\to U)=Ne^{-{\rm Tr}(P'^2)/2}, \tag{2.13}$$

where P' is the final momentum. Substituting the probabilities (2.12) and (2.13) into (2.11) one finds that the hybrid-algorithm evolution is accepted with probability

$$P=\min(1,e^{-\delta H_\Phi})=\min\left[1,\frac{e^{-H_\Phi(U',P')}}{e^{-H_\Phi(U,P)}}\right]. \tag{2.14}$$

Thus, if the MD evolution equations were solved exactly all changes would be accepted. The discretization means that $\delta H_\Phi\neq 0$, so that some of the changes are rejected, and in the process the algorithm is made exact. Note that the need to evaluate the change in a local Hamiltonian constrains one to use a multiple of four flavors of staggered fermions. For the most interesting cases $n_f=2$ or 3, one must give up the description in terms of a Hamiltonian, and therewith the exactness.

The key point which allows the use of the Monte Carlo steps is the reversibility of the MD evolution. For this to hold the exponentiation in computing $U(t+\epsilon)=\exp(i\epsilon P)U(t)$ must be exact. In practice, we have used a fourth-order expansion for most of our runs, and our implementation is therefore reversible only to $O(\epsilon^4)$. The cost of going to a sixth-order expansion is small and we have checked that doing so has no significance effect on the results. A second irreversibility can come from incomplete convergence in the calculation of X. Of course, incomplete convergence is not allowed at the end points in the calculation of δH_Φ. But for the intermediate times one has a choice. If the initial guess for X is always the same, or is random, then incomplete convergence does not cause irreversibility. The alternative is to make a good initial guess for X using an extrapolation from the values at previous MD times. This will reduce the number of CG iterations needed to calculate X to a given accuracy, but it introduces an irreversibility unless the calculation of X has no error. We use the latter method in this paper.

To summarize, the MD equations will wander off the energy shell due to the finite step size. Since the process is reversible, one can use a global Metropolis step to correct for the wandering on average. From this point of view it is clear that the parameters β_h and m_h used in the hybrid evolution do not need to be the same as the β and m used in the global accept and/or reject. This opens up the possibility of tuning the parameters β_h and m_h (Ref. 1) so as to cancel some of the effect of the systematic errors. Ideally, one should tune the parameters so as to minimize decorrelation times. In practice this requires very long runs.[12] In this paper we use an alternative criterion: maximizing the acceptance of the global accept and/or reject step. If the acceptance rate is reasonably high, then studies of decorrelations using the uncorrected hybrid algorithm[4,13] should be applicable.

In a similar way one can use a smaller number of conjugate-gradient iterations ($n_{\rm CG}$) in the intermediate calculations of X, if one is using the method in which the initial guess for X does not depend on the previous values of X. The alternative is to use the previous values to make a good initial guess for X, and in this way reduce $n_{\rm CG}$. Which choice is the better can only be decided in the field, and will depend on the parameters, particularly the quark mass and ϵ. Given the number of parameters that one can tune, we decided to forego the additional freedom of varying $n_{\rm CG}$. Instead we use the previous X as the initial guess, i.e., do a zeroth-order extrapolation, and always run to a fixed accuracy. The convergence criterion we use for all our runs is

$$|(M^\dagger M)X-\phi_e|^2<10^{-9}|X|^2.$$

We have found that acceptances are essentially unchanged if we use half the number of CG steps that result from applying this criterion. Thus, we consider this choice conservative.

The HMC algorithm defined above depends on the two parameters ϵ and $n_{\rm MD}$. The tuning of coupling constants must be done separately for each value of ϵ and $n_{\rm MD}$. It is also possible to change the HA. For example, one can, for fixed ϕ_e, repeat the sequence of refreshing the momenta and evolving the MD equations more than once

before accepting or rejecting using the accumulated δH_Φ. One can also use a higher-order discretization scheme, as long as is reversible. However, in this study we have concentrated entirely on the simplest algorithm.

III. RESULTS FOR PURE GAUGE SU(3)

In the absence of dynamical fermions, the HMC algorithm has one physical parameter β, and three tunable parameters: ϵ, $n_{\rm MD}$, and β_h. We have measured the acceptance as a function of these parameters on 4^4 and 6^4 lattices. We also have measured decorrelation times on longer runs on a 9^4 lattice,[12] though we have not made a scan of the parameter space for this lattice.

Most of our runs on 4^4 and 6^4 lattices are at $\beta=5.6$. This value lies in the crossover region where there is a peak in the specific heat. For the 4^4 lattices this is also close to the position of the finite-temperature transition. We choose this value so as to provide the HMC algorithm with the most stringent test.

We display our results as plots of the acceptance versus the shift in β: $\delta\beta=\beta-\beta_h$. The same initial thermalized lattice is used for all the data points in Fig. 1. Figures 1(a) and 1(b) show our results using leapfrog method 1, for $\epsilon=0.05$ and 0.1, and a variety of $n_{\rm MD}$. Figure 1(a) contains the 4^4 data and Fig. 1(b) the 6^4 data. The acceptance is an average over 2500 Metropolis accepts and/or rejects. Figures 2(a)–2(c) compare leapfrog methods 1 and 2 on 4^4 lattices. The three figures show, respectively, the effects of varying $n_{\rm MD}$, β, and ϵ. Notice that the data in Fig. 2(c) are from $\beta=5.7$. Notice also that the data in Fig. 2 are obtained from an average over 500 Metropolis steps, resulting in a larger statistical error than in Fig. 1. We estimate roughly a 5% error for Fig. 2. We draw the following conclusions from these figures.

(i) For fixed ϵ and $n_{\rm MD}$ the acceptance peaks at a value of β_h different from β. For leapfrog method 1, $\beta_h^{\rm peak}<\beta$, while for method 2, $\beta_h^{\rm peak}>\beta$. The shape of the peaks is

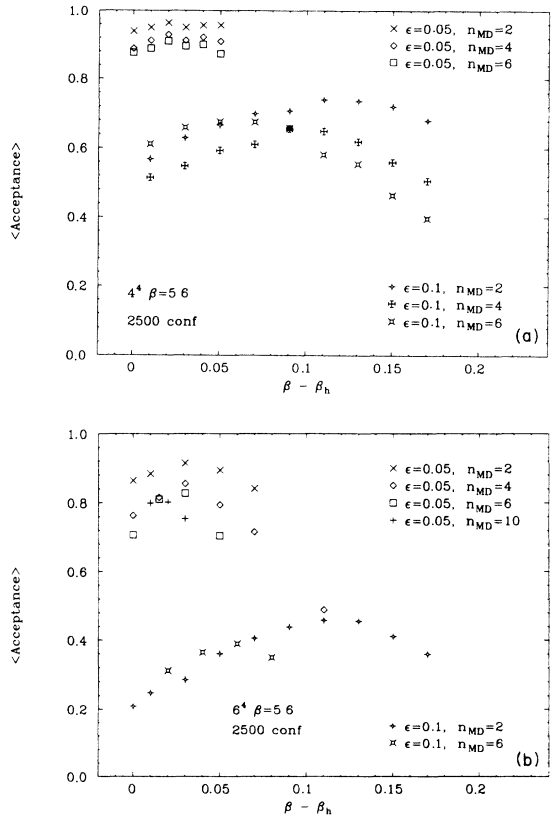

FIG. 1. Acceptance vs $\delta\beta=\beta-\beta_h$ for the pure gauge hybrid Monte Carlo algorithm (method 1) on a (a) 4^4 and (b) 6^4 lattice. The runs are at $\beta=5.6$, for two values of ϵ and for a variety of leapfrog steps $n_{\rm MD}$. The acceptance is averaged over 2500 Metropolis steps.

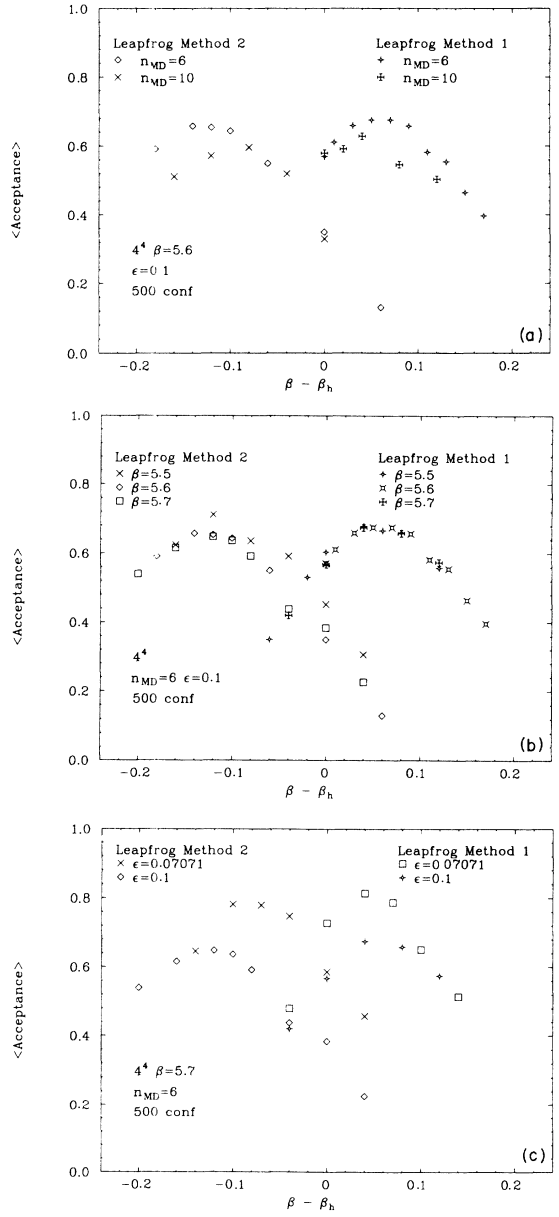

FIG. 2. Comparisons of HMCA methods 1 and 2 for pure gauge SU(3) on a 4^4 lattice. The points in the right half of the plot are for method 1. The acceptance is averaged over 500 Metropolis steps. The plots show the effect of varying (a) $n_{\rm MD}$, (b) β, and (c) ϵ.

similar for both methods, but for method 2 the peak is shifted further from $\beta_h=\beta$.

(ii) For fixed ϵ, the peak in the acceptance moves towards $\beta_h=\beta$ when $n_{\rm MD}$ is increased.

(iii) The height of the acceptance peak depends weakly on $n_{\rm MD}$.

(iv) The shape of the acceptance curve depends on $n_{\rm MD}$. It becomes more sharply peaked as $n_{\rm MD}$ is increased.

(v) The height of the acceptance curve falls rapidly with increasing ϵ, and the peak moves further from $\beta=\beta_h$.

(vi) For fixed parameters, the acceptance drops when going from the 4^4 lattice to the 6^4 lattice.

(vii) On a 4^4 lattice the acceptance profile shows almost no β dependence for leapfrog method 1 and only a small dependence for method 2.

To further investigate the difference between methods 1 and 2, we compare them both with and without the accept and/or reject step. We measure the average plaquette in high statistics runs at $\beta=5.7$, $n_{\rm MD}=6$, and $\epsilon=0.1$. The results are as follows: (a) HA using leapfrog method 1 [0.5645(6)]; (b) HA with method 2 [0.5415(7)]; (c) HMC algorithm with method 1 [0.5595(3)]; and (d) HMC algorithm with method 2 [0.5595(4)]. These results confirm that the two variants of the exact HMC algorithm are simulating the same theory. They also show that at couplings where the HMC algorithm is efficient, the lowest-order leapfrog hybrid algorithm has significant ϵ-dependent errors. We see that method 1 orders the system, while method 2 disorders it. The magnitude of the effect is larger for method 2. These results are consistent, in both direction and relative magnitude, with the tuning in β_h needed to optimize acceptances.

We have also made tests using the mixed method 1/2. Since methods 1 and 2 require β_h to be tuned in opposite directions, one might hope that a combination of the methods would improve upon either. In fact, we find the opposite: method 1/2 is considerably worse than either method 1 or 2 for the whole range of β_h. For example, using six iterations with $\epsilon=0.1$ of method 1, followed by six similar iterations of method 2, and then applying the Metropolis step, we find an acceptance of 9%, 9%, and 2% for $\beta-\beta_H=0.1$, 0.0, and -0.1, respectively.

In Sec. V we take the first steps towards an analytic understanding of these numerical results. We explain the direction of the shifts in the average plaquette in the pure HA runs. We can also estimate the magnitude of the shifts. We find a partial explanation for the signs and magnitudes of $\delta^{\rm peak}=\beta_h^{\rm peak}-\beta$ for the two methods. We find that $\delta^{\rm peak}$ depends on $n_{\rm MD}$, but we cannot calculate the form of this dependence. Naively, one expects that the acceptance should fall monotonically as $n_{\rm MD}$ increases. In fact, the data show a more complicated, though weak, dependence on $n_{\rm MD}$. For example, for some values of β_h the acceptance first rises and then falls as $n_{\rm MD}$ increases. Although we do not understand this variation in the acceptance, the fact that it is weak suggests that the bulk of the error in the MD evolution arises from the first few steps. Indeed, the only difference between methods 1 and 2 is in the initial and final half steps. This suggests that if one could reduce the errors in these half steps, one might well improve the algorithm.

The acceptance is very sensitive to the value of ϵ. Roughly speaking one should choose ϵ to maximize acceptance$\times\epsilon$, which is an estimate of the speed of progress through configuration space measured in CPU time. This criterion suggests that $\epsilon_{\rm optimal}\geq 0.1$ for the 4^4 lattice, while on a 6^4 lattice $\epsilon_{\rm optimal}\sim 0.05$–$0.1$. Clearly, the optimal ϵ decreases as the lattice size increases.

The lack of sensitivity to $n_{\rm MD}$ means that it should be chosen so as to minimize decorrelation times. Studies of the uncorrected hybrid algorithm[4,13] suggest that $\epsilon\times n_{\rm MD}\sim 0.5$–$0.1$ is optimal. For acceptances $\geq 50\%$ this result should apply to the HMC algorithm.

These results suggest that the HMC algorithm can be a practical algorithm for simulations on large lattices. We have tested this by making long runs on a 9^4 lattice. These runs are part of a study of decorrelation times using different algorithm.[12] All runs use method 1, and have $\beta_h=\beta$. At $\beta=6$, and using $\epsilon=0.04$, $n_{\rm MD}=5$, we find 59% acceptance. At $\beta=5.9$ we have two measurements. With $\epsilon=0.0375$ and $n_{\rm MD}=20$ we find 70% acceptance. With $\epsilon=0.025$ and $n_{\rm MD}=40$ the acceptance is 84%. The good news is that the values of ϵ are comparable to those used in HA simulations. For example, a study of the pure gauge theory using the method 2 HA on a $8^3\times 4$ lattice uses $\epsilon=0.025$ (Ref. 14). Thus, for lattices up to 9^4 one loses little or nothing by moving to the exact algorithm. On the other hand, the bad news is that HMC algorithm is about a factor of ≈ 20 slower in CPU time than the finite-step-size overrelaxed algorithm, as measured by decorrelations. However, for QCD with fermions, the decorrelation will be slow for all known practical algorithms, since all involve a small step size. This is a warning that we should be prepared to make very long runs to get a reliable statistical sample.

IV. RESULTS FOR QCD WITH $n_f=4$

We have geared out tests and optimization of the exact Φ algorithm towards studying the finite-temperature transition for finite quark masses. We concentrate here on the optimization of the acceptance, and present our results for the transition in the following paper.[9] The advantages of working at or near the transition are (i) that we have a large body of data to compare against and (ii) that it provides a stringent test of the algorithm. Our data is taken on $6^3\times 4$ and $8^3\times 4$ lattices.

With the introduction of dynamical fermions we have one extra parameter to vary: m_h. The proliferation of parameters forces us to study the effect of varying one or two parameters at a time. Based on our study of the pure gauge theory we decided a reasonable estimate of the acceptance for a given set of parameters would be provided by a run of 500 trajectories. Here a trajectory is a unit of MD time, i.e., $1/\epsilon$ iterations. Thus, we are averaging over $500/(\epsilon\times n_{\rm MD})$ global accepts and/or rejects. We have measured the acceptance on the $6^3\times 4$ lattice at $m=0.1$, 0.2, 0.3, and 0.5. All results are obtained using

the method-1 leapfrog algorithm.

We first fix $\epsilon=0.05$, $n_{\rm MD}=4$, and search for the optimal β_h and m_h. It turns out that the results are similar to the pure gauge case. The peak in the acceptance occurs when both β_h and m_h are chosen to disorder the system, i.e., $\delta\beta\equiv\beta-\beta_h>0$ and $\delta m\equiv m-m_h<0$. It turns out that the location of the peak is at $\delta\beta\sim0.02$ and $\delta m\sim-0.005$ for the ranges of β ($5.1<\beta<5.6$) and m that we studied. For all but the smallest quark mass, the data at different β agree within our statistical errors, and we average them. Figure 3 shows the acceptance as a function of δm for $m=0.5$ and $\delta\beta=0.0$ and 0.2. Figure 4 shows the acceptance for $m=0.2$ as a function of $\delta\beta$ for $\delta m=-0.005$ and 0.0. The curves for $m=0.3$ and 0.5 are similar except for a small overall decrease in acceptance as m is reduced. The important feature exhibited by this data is that with dynamical fermions the acceptance profile is very flat about the maximum. Thus, running without any shift in the couplings is close to optimal.

Most of our runs sit on the finite-temperature transition and we see, for all m, tunneling transitions between high- and low-temperature phases.[9] For $m=0.2$, 0.3, and 0.5 we see no significant change in the acceptance when the system jumps from one phase to the other. However, for $m=0.1$ we do see significant differences between the phases. This is shown in Fig. 5, where we label points by their phase (hot, i.e., $\beta>\beta_{\rm crit}$, or cold). Comparing this plot with Fig. 4 one sees that there is a drop in acceptance on going to $m=0.1$ from $m=0.2$ for the same ϵ. These results are consonant with the expectation that the MD evolution becomes more sensitive as the eigenvalues of the Dirac operator become smaller.

Figure 6 illustrates the dependence of the acceptance on ϵ for $\epsilon\times n_{\rm MD}$ fixed at 0.2. The data is for $m=m_d=0.3$ and $\delta\beta=0.03$. For these parameters the ϵ which maximizes acceptance$\times\epsilon$ is $\epsilon\sim0.07$. As for the pure gauge theory, we find that the acceptance is almost insensitive to the value of $n_{\rm MD}$. For example, with $m=m_d=0.3$, $\beta=5.33$, and $\delta\beta=0.1$–0.2, we find acceptances of 0.895, 0.891, 0.882 for $n_{\rm MD}=6$, 15, 30. Thus it is possible, with no loss in acceptance, to use the value of $n_{\rm MD}$ which minimizes the decorrelation lengths.[4]

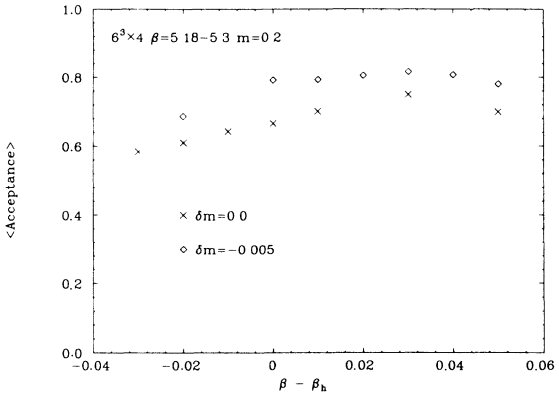

FIG. 4. Acceptance vs $\delta\beta$ for $m=0.2$, and $\delta m=0.0$ and -0.005. Results are for $\epsilon=0.05$ and $n_{\rm MD}=4$ on a $6^3\times4$ lattice.

We have extended these results onto an $8^3\times4$ lattice for $m=0.1$. For $\epsilon=0.05$ and $n_{\rm MD}=4$, the acceptance drops to 15% (cold phase) and 25% (hot phase). We thus decreased ϵ to 0.0033. Once again, we find little dependence on $n_{\rm MD}$, and so for our studies of the phase transition we use $n_{\rm MD}=15$. We find that the acceptance rises to $\sim65\%$. This is roughly an average between the hot and cold phases. As for the pure gauge theory, the optimal value of ϵ is comparable to those used in HA simulations. For example, Ref. 15, use $\epsilon=0.01$ on an $8^3\times4$ lattice.

V. ANALYTIC CONSIDERATIONS

In this section we give an analytic discussion of the HMC algorithm. We first derive the dependence of the acceptance on ϵ and the volume V. Our aim is to calculate how the CPU time varies with the volume. This analysis is an extension of the work of Creutz,[8] who considered the acceptance in the exact Langevin algorithm, i.e., HMCA with $n_{\rm MD}=1$. For simplicity we work with a

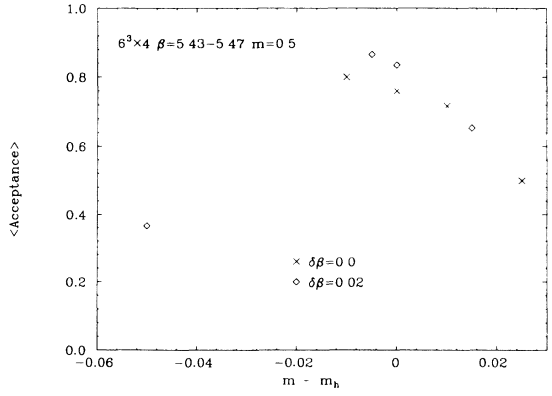

FIG. 3. Acceptance vs $\delta m=m-m_d$ for $m=0.5$, $\epsilon=0.05$, and $n_{\rm MD}=4$ on a $6^3\times4$ lattice. Results for $\delta\beta=\beta-\beta_h=0.0$ and 0.02 are given.

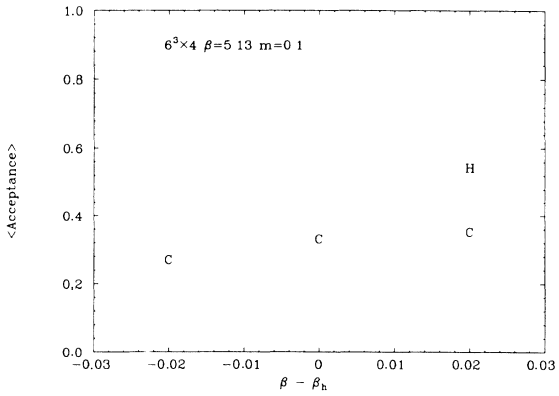

FIG. 5. Same as Fig. 3 for $m=0.1$ and $\delta\beta=0.02$. Data points in hot and cold phases are distinguished.

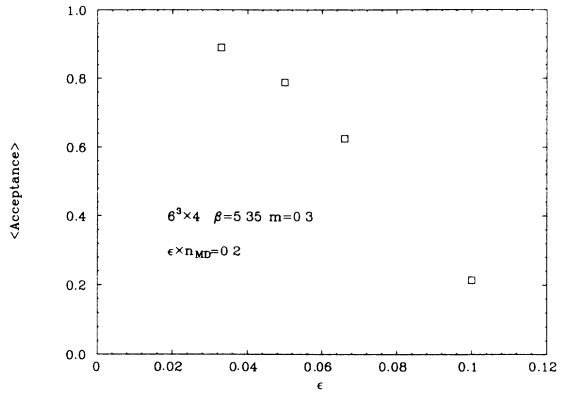

FIG. 6. Acceptance vs ϵ for $\epsilon\times n_{\rm MD}=0.2$. Data are for $\beta=5.35$, $m=0.3$, $\delta m=0.0$, and $\delta\beta=0.03$ on a $6^3\times4$ lattice.

theory of scalar fields, the Hamiltonian of which is

$$H=\tfrac{1}{2}P_iP_i+S(\phi)\ . \tag{5.1}$$

S is an arbitrary local action, and we use the notation $S_i=\partial S/\partial\phi_i$ for partial derivatives. Repetition of indices always implies summation.

The exact MD equations are

$$\dot\phi_i=P_i,\quad \dot P_i=-S_i\ . \tag{5.2}$$

We solve these approximately using the two leapfrog methods described above. These can be built up from elementary MD steps which move $\phi^{(n)}$ and $P^{(n)}$ to $\phi^{(n+1)}$ and $P^{(n+1)}$ (Ref. 16), where the superscripts indicate the total number of steps taken.

Method 1:

$$\begin{aligned}P_i^{(n+1)}&=P_i^{(n)}-\epsilon S_i(\phi^{(n)}+\tfrac{1}{2}\epsilon P^{(n)})\ ,\\ \phi_i^{(n+1)}&=\phi_i^{(n)}+\tfrac{1}{2}\epsilon P_i^{(n)}+\tfrac{1}{2}\epsilon P_i^{(n+1)}\ .\end{aligned}$$

Method 2:

$$\begin{aligned}\phi_i^{(n+1)}&=\phi_i^{(n)}+\epsilon P_i^{(n)}-\tfrac{1}{2}\epsilon^2S_i(\phi^{(n)})\ ,\\ P_i^{(n+1)}&=P_i^{(n)}-\tfrac{1}{2}\epsilon S_i(\phi^{(n)})-\tfrac{1}{2}\epsilon S_i(\phi^{(n+1)})\ .\end{aligned} \tag{5.3}$$

For both methods the measure is preserved:

$$d\phi^{(n+1)}dP^{(n+1)}=d\phi^{(n)}dP^{(n)}\ .$$

The change in the H during a single step can be written as (subscripts refer to powers of ϵ)

$$\begin{aligned}\delta H^{(n)}&\equiv H(\phi^{(n+1)},P^{(n+1)})-H(\phi^{(n)},P^{(n)})\\ &=\epsilon^3\delta H_3^{(n)}+\epsilon^4\delta H_4^{(n)}+\epsilon^5\delta H_5^{(n)}+\epsilon^6\delta H_6^{(n)}\\ &\quad+\cdots\ ,\end{aligned} \tag{5.4}$$

where

$$\delta H_3=aP_iS_{ij}S_j+bP_iP_jP_kS_{ijk}\ . \tag{5.5}$$

We do not need the explicit forms of δH_4, δH_5, and δH_6 for the present analysis. The parameters a and b distinguish methods 1 and 2. For method 1, $a=-\frac{1}{4}$ and $b=\frac{1}{24}$; for method 2, $a=\frac{1}{4}$ and $b=-\frac{1}{12}$. This means that δH_3 is almost opposite for the two methods.

Let us briefly recap the analysis of Creutz[8] for $n_{\rm MD}+1$. Recall that the probability of acceptance is

$$P_{\rm acc}=\min(1,e^{-\delta H^{(0)}})\ . \tag{5.6}$$

To get the average acceptance this must be integrated over the equilibrium probability distribution:

$$\langle P_{\rm acc}\rangle\equiv\int_\phi\int_P d\phi^{(0)}dP^{(0)}e^{-H(\phi^{(0)},P^{(0)})}P_{\rm acc}\ . \tag{5.7}$$

Each term in δH is proportional to the volume V. We want to find the first term in δH which causes the acceptance to decrease exponentially with the volume. This will then determine how large we can make ϵ. The third-order term in δH averages to zero since it is proportional to odd powers of the momenta. This means that however big δH_3 becomes, it can only lower the acceptance to 1/2. To get exponential suppression we must find a term whose average is positive definite. As shown by Creutz, $\langle\delta H_4^{(0)}\rangle=0$ for method 2. One can also demonstrate that this is true for method 1, although the expression for δH_4 differs from that found with method 2. For both methods, $\langle\delta H_5^0\rangle$ vanishes trivially, leaving the first nontrivial term to be δH_6. Creutz shows that $\langle\delta H_6^{(0)}\rangle>0$ for method 2, and we find the same to be true for method 1 (Ref. 17). Thus

$$\langle P_{\rm acc}\rangle\sim\exp(-\epsilon^6V) \tag{5.8}$$

and to maintain reasonable acceptance, $\epsilon\propto V^{-1/6}$. Since the HMCA with $n_{\rm MD}=1$ is equivalent to the Langevin algorithm, it takes $O(1/\epsilon^2)$ steps to obtain a significantly different configuration. To move this far thus requires a time $\propto V^{4/3}$, since each Langevin step requires time growing as V.

We now extend this analysis to $n_{\rm MD}\sim1/\epsilon$, i.e., the case in which we move unit distance with the MD equations before the global accept and/or reject. To do this we have to account for the fact that the measure after some number of steps differs from that for the first step. To this end, we define $H_{\rm eff}$ to be the functional which is kept constant by the single steps defined in Eq. (5.3):

$$H_{\rm eff}(\phi^{(n+1)},P^{(n+1)})=H_{\rm eff}(\phi^{(n)},P^{(n)})\ .$$

To leading order

$$H_{\rm eff}(\phi,P)=H(\phi,P)+\epsilon^2H_2(\phi,P)+O(\epsilon^3)\ , \tag{5.9}$$

where

$$H_2(\phi,P)=-(\tfrac{1}{2}a+b)S_iS_i-bS_{ij}P_iP_j\ . \tag{5.10}$$

This can be seen by noting that for the exact trajectory

$$\frac{d}{dt}H_2(\phi(t),P(t))=-\delta H_3(\phi(t),P(t))\ . \tag{5.11}$$

The result (5.10) agrees with that of Duane and Kogut,[4] who have calculated H_2 for method 2. Notice again that the result for the two methods are roughly opposite. However, for both methods H_2 has the same, positive-definite, average: $\langle H_2\rangle=\langle S_iS_i\rangle/24$.

From Eq. (5.9) it follows that the measure in terms of the fields after n steps is

$$d\phi^{(0)}dP^{(0)}\exp[-H(\phi^{(0)},P^{(0)})]=d\phi^{(n)}dP^{(n)}\exp\{-[H(\phi^{(n)},P^{(n)})+\epsilon^2H_2(\phi^{(n)},P^{(n)})-\epsilon^2H_2(\phi^{(0)},P^{(0)})+O(\epsilon^2 n)]\}\ . \tag{5.12}$$

We would like to bring down the ϵ^2 terms from the exponent. We show below that $H_2^{(n)}-H_2^{(0)}$ is a fluctuating quantity proportional to $\sqrt{V}$. Thus, provided $\epsilon^2\propto 1/\sqrt{V}$, we may make the expansion. This is precisely the scaling relation we find below. Using this we can calculate the average of δH for the $(n+1)$th step:

$$\begin{aligned}\langle\delta H^{(n)}\rangle&=\langle\delta H^{(0)}\rangle-\epsilon^2\langle\delta H^{(0)}H_2^{(0)}\rangle\\&\quad+\epsilon^2\langle\delta H^{(0)}H_2^{(-n)}\rangle+O(\epsilon^6)\\&=+\epsilon^5\langle\delta H_3^{(0)}H_2^{(-n)}\rangle+O(\epsilon^6)\\&=+\epsilon^5\langle\delta H_3^{(n)}H_2^{(0)}\rangle+O(\epsilon^6)\ .\end{aligned}\tag{5.13}$$

We have changed variables freely, and assumed that $n\le 1/\epsilon$. Since we are dropping terms of $O(\epsilon^6)$, we can replace $\phi^{(n)}$ and $P^{(n)}$ by the values they would have along the trajectory obtained by integrating the MD equations exactly. This means that averages of quantities which only depend on the fields at a single time are time independent. In particular $\langle H_2^{(n)}\rangle=\langle H_2^{(0)}\rangle$, which means that on average the $O(\epsilon^2)$ term in the measure is zero. However, the fluctuations in this term are correlated with those in $\delta H_3^{(n)}$, leading to a nonvanishing contribution to $\langle\delta H^{(n)}\rangle$, as we shall now see.

To evaluate the right-hand side (RHS) of (5.13) we use Eq. (5.11). Concentrating only on the leading term, we have $(t=n\epsilon)$

$$\langle\delta H^{(n)}\rangle=-\epsilon^5\left\langle H_2(0)\frac{dH_2(t)}{dt}\right\rangle\ . \tag{5.14}$$

What we really want is the average value of the total change in H summed over all the steps, $\langle\Delta H\rangle=\sum_n^{n_{\rm MD}}\langle\delta H^{(n)}\rangle$. To leading order, the sum can be turned into an integral, which we can integrate trivially since the RHS of (5.14) is a total derivative. The final result is

$$\begin{aligned}\langle\Delta H\rangle&=\epsilon^4\langle H_2(0)^2-H_2(0)H_2(t)\rangle\\&=\tfrac{1}{2}\epsilon^4\langle[H_2(0)-H_2(t)]^2\rangle\ ,\end{aligned}\tag{5.15}$$

where t is now $n_{\rm MD}\epsilon$. To get the second line we have used the time independence of $\langle H_2(t)^2\rangle$. The RHS of (5.15) is manifestly positive, and thus reduces the acceptance. It is of lower order in ϵ than $\langle\delta H^{(0)}\rangle$—one pays a price for running the MD equations for many steps. However, for $t\sim\epsilon$ it is easy to see that the RHS of Eq. (5.15) is $\propto\epsilon^6$, and thus matches smoothly onto the $n_{\rm MD}=1$ result.

Despite appearances, $\langle\Delta H\rangle$ is proportional to the volume, since it is the difference of two quantities, $\langle H^{(n_{\rm MD})}\rangle$ and $\langle H^{(0)}\rangle$, each linear in the volume. This is true for the RHS of (5.15) because the difference $H_2(0)-H_2(t)$, whose average is zero, is uncorrelated from site to site. For small t this can be shown explicitly. For $t\sim 1$ there should be no remaining correlations between sites.

The average acceptance for the entire trajectory scales as

$$\langle P_{\rm acc}\rangle\sim\exp(-\epsilon^4V)\ , \tag{5.16}$$

so we must scale $\epsilon\propto V^{-1/4}$ to keep the acceptance fixed. This means that to traverse a unit in MD time takes CPU time $\propto V^{5/4}$. This is a slight improvement over the exact Langevin algorithm ($n_{\rm MD}=1$).

Using this result, one can estimate the size of the fluctuations in ΔH. The leading term in ΔH is proportional to ϵ^2-1/ϵ steps each with error $\propto\epsilon^3$. This term averages to zero, but since it is fluctuating we expect that it is $\propto\sqrt{V}$. Thus, the size of the fluctuations in ΔH are $\propto\epsilon^2\sqrt{V}\sim 1$, i.e., of the same size as the contribution from the ϵ^4 terms. Since the powers of ϵ are different in the fluctuating and positive-definite parts of ΔH, we expect the acceptance to vary rapidly from 100% to 0% as ϵ is increased though the critical value. The result for the scaling of ϵ also justifies the expansion of the exponential in the measure (5.12).

We close this section with a discussion of the differences between methods 1 and 2. We will only consider the pure gauge simulations. We would like to understand why acceptance is optimized by using $\beta_h<\beta$ for method 1, while the opposite is true for method 2. A related question is why the pure HA with method 1 orders the system, while the HA with method 2 disorders. It turns out that the latter question is easier to answer so we discuss it first.

The analysis of hybrid algorithms using method 2 has been given done by Duane and Kogut.[4] Using a generalization of the Fokker-Planck equation they find that the equilibrium action for the ϕ fields is independent of $n_{\rm MD}$:

$$S_{\rm eff}^{(2)}(\phi)\equiv S+S_2^{(2)}=S+\epsilon^2(-\tfrac{1}{8}S_iS_i+\tfrac{1}{4}S_{ii})\ , \tag{5.17}$$

The superscript denotes the method. This agrees with the result from the Langevin limit.[7] It is possible to adapt the analysis of Ref. 4 to method 1, and after considerable algebra we find the simple result

$$S_{\rm eff}^{(1)}\equiv S+S_2^{(1)}=S+\epsilon^2(\tfrac{1}{8}S_iS_i-\tfrac{1}{8}S_{ii})\ . \tag{5.18}$$

The shift in the action is again independent of $n_{\rm MD}$, but it is almost opposite to that from method 2.

To apply these results to gauge theories care must be taken because U is an SU(3) matrix. The equations derived above can be used with the substitutions

$$\begin{aligned}&S(U)\to\beta\sum_p(1-\tfrac{1}{3}{\rm Re\,Tr}U_p)\ ,\\&S_iS_i\to-\tfrac{1}{9}\beta^2\sum_{i,\mu}{\rm Tr}[(U_{i,\mu}V_{i,\mu})_{\rm TA}(U_{i,\mu}V_{i,\mu})_{\rm TA}]\ ,\\&S_{ii}\to\tfrac{8}{9}\beta\sum_{i,\mu}{\rm Re\,Tr}U_{i,\mu}V_{i,\mu}\ .\end{aligned}\tag{5.19}$$

Thus, S_{ii} is of the form of the original action, up to a constant. To relate $S_i S_i$ to S one can either perform a nonlinear change of variables,[7] or use an approximation valid as the lattice spacing vanishes. The former method is exact through $O(\epsilon^2)$, but the nonlinearity in the change of variables means that the expectation values of the Wilson loops are renormalized in a loop-dependent way.[7] We are seeking a qualitative understanding so we use the approximate, but transparent, approach. The crucial observation is that the TA part of a plaquette is proportional to $F_{\mu\nu}$ at the lowest nontrivial order in the lattice spacing. Dropping higher-order terms we find

$$-\frac{\beta^2}{9}\sum_{i,\mu}\mathrm{Tr}[(U_{i,\mu}V_{i,\mu})_{\mathrm{TA}}(U_{i,\mu}V_{i,\mu})_{\mathrm{TA}}] = 4\beta\sum_{p}(1-\tfrac{1}{3}\mathrm{Re}\,\mathrm{Tr}\,U_p)+O(a^6)\ . \quad (5.20)$$

Plugging these correspondences into (5.17) and (5.18) we find

$$S_{\mathrm{eff}}=\beta_{\mathrm{eff}}\sum_{p}(1-\tfrac{1}{3}\mathrm{Re}\,\mathrm{Tr}\,U_p)\ ,$$
$$\beta_{\mathrm{eff}}=\beta(1+\epsilon^2 c^{(i)})\ , \quad (5.21)$$

where $c^{(1)}=\frac{5}{6}$ and $c^{(2)}=-\frac{7}{6}$. Thus, we expect method 1 to increase the expectation values of Wilson loops compared to the exact algorithm, and for method 2 to decrease them by a somewhat larger magnitude. These expectations are in agreement with the results of the high-statistics runs described in Sec III.

These results suggest a simple strategy for tuning the HMC algorithm: shift β_h such that $\beta_{\mathrm{eff}}=\beta$. For then the pure HA will be producing gauge configurations that have the correct distribution. However, this may not be the best strategy, for the following two reasons. First, the global changes are accepted or rejected using the change to the entire Hamiltonian, including the momenta. Second, at each refresh one is drawing the gauge fields from the exact distribution, rather than one equilibrated with S_{eff}. This will tend to reduce the shift in β that is needed.

It is difficult to calculate this reduced shift because the true measure, Eq. (5.12), depends on the number of steps taken, and does so in a complicated way. One can carry through the acceptance analysis described above for a tuned action, and the results are identical if one substitutes $H_2'=H_2+S_2$ for H_2. It also remains true that the leading nonvanishing change for $n_{\mathrm{MD}}=1$ occurs at order ϵ^6 (Ref. 17). Clearly what one should do is to minimize $\langle \Delta H \rangle$. This will not remove the $O(\epsilon^4)$ correction, but it will reduce it as much as possible. In general, the best choice of S_2 will depend on n_{MD}. A possible choice of S_2 is that which minimizes $\langle H_2' H_2' \rangle$, i.e., the first term on the RHS of Eq. (5.15). This turns out to be equivalent to choosing S_2 such that H_2' vanishes after averaging over momenta. The result is therefore

$$S_2'=(\tfrac{1}{2}a+b)S_iS_i+bS_{ii} \quad (5.22)$$

leading to

$$\beta_h=\beta[1+\epsilon^2(2a+\tfrac{4}{3}b)]\ , \quad (5.23)$$

where a and b are defined in Eq. (5.5). The coefficient of ϵ^2 is $-\frac{4}{9}$ for method 1, and $\frac{7}{18}$ for method 2. The shifts in β are of the same sign as those needed to correct the pure HA. However, as we expected, the magnitude of the shifts is smaller. Notice that for this choice of S_2 method 2 requires a smaller shift than method 1.

In summary, the two methods are expected to require opposite shifts in β. Method 1 should be optimized by shifting to smaller β, method 2 by shifting to larger β. These shifts should depend quadratically on ϵ for acceptances away from the limiting values of 0 and 1.0. The shifts should depend on n_{MD}, but for $\epsilon\times n_{\mathrm{MD}}\sim 1$, the magnitude of the shifts should be in the range suggested by Eqs. (5.21) and (5.23). Thus, for $\epsilon=0.1$ we expect a $\sim 1\%$ shift in β for both methods. This is consistent with the data in Figs. 1 and 2. As for the difference in the magnitude of the shifts between the methods, or for the behavior of the shifts at small n_{MD}, our results are inconclusive.

VI. CONCLUSIONS

We have presented the results of a systematic study of the HMC algorithm. We find that by tuning the parameters we can increase the acceptance of the global Metropolis step. For the pure gauge theory tuning can increase the acceptance by as much as a factor of 2. The importance of tuning is decreased when we include dynamical fermions. Tuning is also needed less when we use method 1, the leapfrog algorithm in which we move the gauge fields forward in time first. Thus, for the simulation of QCD with dynamical fermions, we can expect almost optimal acceptances if we use method 1 with $\beta_h=\beta$ and $m_h=m$. This is encouraging as it makes the algorithm simple to use in practice.

We also find that the acceptance has little dependence on the number of MD steps, n_{MD}. This means that one should determine the optimal n_{MD} by studying decorrelation times. Studies of uncorrected hybrid algorithms suggest using $\epsilon\times n_{\mathrm{MD}}\sim 0.5-1$. It is very satisfying that the introduction of the Metropolis step does not make the optimization of n_{MD} more difficult.

We find that the acceptance drops rapidly with increasing ϵ. For each lattice size and quark mass, there is an optimal choice of ϵ for which the motion through configuration space is maximized. For that choice of ϵ, the acceptance is roughly 50–70%. The optimal ϵ decreases with increasing lattice size, and with decreasing quark mass. We have argued that, asymptotically, $\epsilon\propto V^{-1/4}$ in order to maintain constant acceptance. Our results are roughly consistent with this prediction.

How well does the HMC compare to approximate algorithms for the simulation of dynamical fermions? This depends crucially on the values of ϵ that one must use. For the HMC algorithm one simply increases ϵ until the acceptance drops to roughly 50%. For users of the pure HA the choice of ϵ is less straightforward. One must decide what is an acceptable error in physical quantities and reduce ϵ until this level of systematic error is achieved. As discussed above, if we compare the optimal ϵ for the HMC algorithm on our largest lattice with the

values used in production runs by practitioners of the HA, we find that the HMC can tolerate somewhat larger values. If we reduce the HMC algorithm ϵ by the acceptance, then the values are comparable, so that for a given CPU time the number of independent configurations is similar.

We can scale these results to larger lattices using the asymptotic formula. For example, if we assume equal speeds on an $8^3\times 4$ lattice, then on a $16^3\times 40$ lattice the HMC algorithm will only be three times slower than the HA. There are clearly considerable uncertainties in the estimation of this ratio, and a reasonable range of possible values is 1–10. Since an exact algorithm gives one the considerable advantage of not having to worry about systematic errors, we think it is very worthwhile testing the HMC algorithm on larger lattices. Although all the tests in this paper have been done using staggered fermions, the algorithm is most useful for Wilson fermions, where one is only restricted to multiples of two flavors. In fact, studies with Wilson fermions on 8^4 lattices find that acceptances of 60–70 % can be attained using values of ϵ similar to those used in the HA simulations.[18]

Finally we would like to comment on how one might improve the HMC algorithm. In Sec. III we found evidence that the dominant source of incomplete acceptance was the initial and final half-steps. This is a result that is not explained by the analysis of Sec. V. Nevertheless, it suggests that a simple way to reduce the systematic error of the hybrid algorithm is to increase the accuracy of the computation of the half-steps. This could be done, for example, with a higher-order Runge-Kutta scheme. Such an approach only makes sense for large $n_{\rm MD}$, where the extra work at the ends requires only a small increase in the number of steps. Of course, for this improvement to be applicable to the HMC algorithm, the evolution must be reversible.

Note added. While this work was being written up we received two papers containing related results. Gausterer and Sanielevici[19] test the exact Langevin algorithm (HMC algorithm with $n_{\rm MD}=1$). Bitar *et al.*[20] study the HMC algorithm using method 2. Their definition of step size differs from ours according to ϵ (Bitar)$=\sqrt{2}\epsilon$ (Gupta). For this reason we added the curve with $\epsilon=0.1/\sqrt{2}$ to Fig. 2(c).

ACKNOWLEDGMENTS

We acknowledge the support of the Pittsburgh Supercomputing Center and the National Magnetic Fusion Energy Computer Center for time on their respective Cray X-MP's. We thank G. Guralnik for his constant encouragement and support, Mike Creutz for very useful discussions and for communicating his results prior to publication, and C. Baillie, K. Bitar, T. Kennedy, A. Patel, and P. Rossi for helpful conversations. This work was supported by the Department of Energy, Contract No. DE-AC03-76SF00515. R.G. thanks the J. Robert Oppenheimer Foundation for financial support.

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

[2]R. T. Scalettar, D. J. Scalapino, and R. L. Sugar, Phys. Rev. B **34**, 7911 (1986).

[3]S. Duane, Nucl. Phys. **B257**, 652 (1985); S. Duane and J. Kogut, Phys. Rev. Lett. **55**, 2774 (1985).

[4]S. Duane and J. B. Kogut, Nucl. Phys. **B275**, 398 (1986).

[5]D. Callaway and A. Rahman, Phys. Rev. D **28**, 1506 (1983); J. Polonyi and H. W. Wyld, Phys. Rev. Lett. **51**, 2257 (1983).

[6]G. Parisi and Y. Wu, Sci. Sin. **24**, 483 (1981).

[7]G. G. Batrouni, G. R. Katz, A. S. Kronfeld, G. P. Lepage, B. Svetitsky, and K. G. Wilson, Phys. Rev. D **32**, 2736 (1985).

[8]M. Creutz, in *Field Theory of the Lattice,* proceedings of the International Symposium, Seillac, France, 1987, edited by A. Billoire *et al.* [Nucl. Phys. B, Proc. Suppl. **4** (1988)].

[9]R. Gupta, G. Kilcup, and S. Sharpe, following paper, Phys. Rev. D **38**, 1288 (1988).

[10]S. A. Gottlieb, W. Liu, D. Toussaint, R. L. Renken, and R. L. Sugar, Phys. Rev. D **35**, 2531 (1987).

[11]N. Metropolis, A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller, and E. Teller, J. Chem. Phys. **21**, 1087 (1953).

[12]R. Gupta, G. W. Kilcup, A. Patel, S. R. Sharpe, and P. de Forcrand, Report No. LAUR-88-824, 1988 (unpublished).

[13]S. A. Gottlieb, W. Liu, D. Toussaint, R. L. Renken, and R. L. Sugar, Phys. Rev. D **35**, 2531 (1987).

[14]S. A. Gottlieb, W. Liu, D. Toussaint, R. L. Renken, and R. L. Sugar, Phys. Rev. D **36**, 3797 (1987).

[15]S. A. Gottlieb, W. Liu, D. Toussaint, R. L. Renken, and R. L. Sugar, Phys. Rev. D **35**, 3972 (1987).

[16]Of course, in practice one does not step in this way, since we only need to have ϕ and P evaluated at the same time at the ends of the trajectory. But for the purposes of the present discussion it is simpler to use the equivalent representation in terms of single steps.

[17]This is most easily seen using a general argument to Mike Creutz (private communication); this issue, Phys. Rev. D **38**, 1228 (1988).

[18]C. Baillie, R. Gupta, G. Guralnik, G. Kilcup, A. Patel, and S. R. Sharpe (in preparation).

[19]H. Gausterer and S. Sanielevici, this issue, Phys. Rev. D **38**, 1220 (1988).

[20]K. M. Bitar, R. Horsley, A. D. Kennedy, S. Meyer, and P. Rossi, Report No. SCRI-88-14, 1988 (unpublished).