Adaptive step size for the hybrid Monte Carlo algorithm

Philippe de Forcrand and Tetsuya Takaishi
Swiss Center for Scientific Computing (SCSC), ETH-Zürich, CH-8092 Zürich, Switzerland
(Received 24 September 1996)

We implement an adaptive step-size method for the hybrid Monte Carlo algorithm. The adaptive step size is given by solving a symmetric error equation. An integrator with such an adaptive step size is reversible. Although we observe appreciable variations of the step size, the overhead of the method exceeds its benefits. We propose an explanation for this phenomenon. [S1063-651X(97)09102-2]

PACS number(s): 02.70.Lq, 11.15.Ha, 12.38.Gc, 02.50.Ng

I. INTRODUCTION

Simulations including dynamical fermions remain the most challenging ones for lattice QCD. The standard method to simulate dynamical fermions is, at the moment, the hybrid Monte Carlo (HMC) algorithm [1], although it still requires large amounts of computational time. An alternative method to simulate the dynamical fermions is a local multiboson algorithm based on a polynomial approximation of the fermion matrix, proposed by Lüscher [2]. Much interest has been recently devoted to this algorithm [3] to make it as efficient as the HMC algorithm.

The HMC simulation combines molecular dynamics (MD) evolution with a Metropolis test. In order to obtain the correct equilibrium, the integrator used in the MD evolution must satisfy two conditions; it must be (i) time reversible and (ii) area preserving.

One such integrator satisfying these conditions is the leapfrog integrator, which is normally used in the HMC simulation. Errors of the leapfrog integrator start with $O(\Delta t^3)$, where Δt is the step size of the integrator. These errors cause violation of the conservation of the total energy, which must be corrected by the Metropolis test at the end of the MD trajectory. Let ΔH be the energy violation:

$$\Delta H = H_{end} - H_{begin}, \tag{1}$$

where H_{begin} (H_{end}) is the total energy at the beginning (end) of the MD trajectory. The Metropolis test accepts a new configuration with a probability P_{prob}:

$$P_{prob} \propto \min(1, \exp(-\Delta H)). \tag{2}$$

In order to maximize acceptance of the Metropolis test, it might be preferable to use a higher-order integrator [4]. However, higher-order integrators do not appear practical for lattice QCD since they require more arithmetic operations (force evaluations coming from the fermionic action) than the simplest low-order integrator, and this overhead exceeds the gain in step size.

So far conventional HMC simulations have been performed with a fixed step size Δt during the MD simulation. The local integration error does not remain constant in this case. When the trajectory approaches an energy barrier ($S_{eff} = -\ln \det D$ large), it is repelled and bounces off. The curvature of the trajectory increases, and with it the integration error. This situation becomes more pronounced at small quark mass, since the height of the energy barriers diverges in the presence of zero modes. Therefore we expect a behavior of the MD trajectory similar to Fig. 1. Varying the step size adaptively, keeping the local error constant, may be a good way to obtain a better integrated trajectory and may result in higher acceptance. Naively it would seem that this can be accomplished by calculating a local error at (p,U), where $p=(p_1,p_2,\ldots)$ and $U=(U_1,U_2,\ldots)$ collectively represent momenta and link variables, respectively, and then by keeping this local error constant. However, this naive scheme is not applicable for the HMC simulation because it violates reversibility.

Recently an adaptive step size method with reversible structure was proposed by Stoffer [5]. He constructed a symmetric error estimator which gives the same error value at a reflected point. The step size is then determined at every integration point by demanding that the symmetric error estimator remain constant. Stoffer implemented his method for the Kepler problem and obtained better results than the conventional ones. The possibility to apply this adaptive step size method to the HMC algorithm was stated in Ref. [6]. In this paper, we implement this method for the HMC simulation and examine its cost and its efficiency.

II. CONSTRUCTION OF AN ADAPTIVE STEP SIZE

Here we construct an adaptive step size compatible with a time-reversible integrator. We follow the idea proposed by Stoffer [5].

Let H be the Hamiltonian of our system,

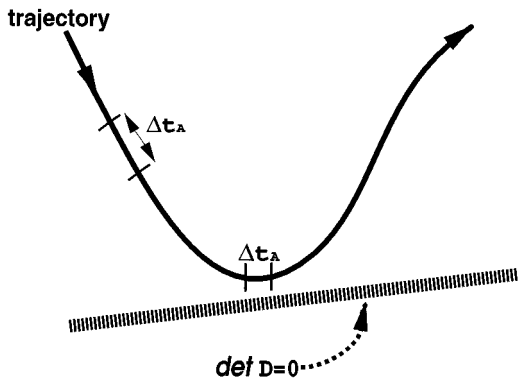

FIG. 1. Schematic behavior of a molecular dynamics trajectory in configuration space. det$D$ is the determinant of the Dirac matrix.

$$H=\frac{1}{2}\sum\ p_i^2+S(U), \tag{3}$$

where p_i are momenta, U are gauge links, and $S(U)$ consists of a gauge part $S_g(U)$ and a fermionic part $S_f(U)$,

$$S(U)=S_g(U)+S_f(U), \tag{4}$$

$$S_g(U)=\beta\sum\ \left(1-\frac{1}{N_c}\mathrm{Re\ Tr}U_{plaq}\right), \tag{5}$$

$$S_f(U)=\phi^\dagger(DD^\dagger)^{-1}\phi, \tag{6}$$

where N_c is the number of colors, ϕ is a pseudofermion vector and $D=1+\kappa M$ is the Wilson fermion matrix, with κ the hopping parameter.

Call $T(\Delta t)$ a one-step integrator. It maps momenta and link variables (p,U) onto (p',U'),

$$T(\Delta t):(p,U)\rightarrow(p',U'). \tag{7}$$

If this one-step integrator is reversible, then it satisfies

$$T(-\Delta t):(p',U')\rightarrow(p,U). \tag{8}$$

In this study, we use the leapfrog integrator as our one-step integrator. In terms of the time evolution operators [4,7], the one-step integrator $T(\Delta t)$ can be written as

$$T(\Delta t)=\exp\left[\frac{\Delta t}{2}L\left(\frac{1}{2}\sum\ p_i^2\right)\right]\ \exp[\Delta tL(S(U))]\times\exp\left[\frac{\Delta t}{2}L\left(\frac{1}{2}\sum\ p_i^2\right)\right], \tag{9}$$

where $L(\cdot)$ is the linear operator which is given by the Poisson bracket [7]. The one-step integrator requires one force evaluation represented by $\exp[\Delta tL(S(U))]$. The fermionic part of the force depends on the solution of a linear equation of the type $Dx=\phi$, which is obtained iteratively at great expense of computer time. Thus force evaluations dominate the computation, and the cost of our algorithm can be measured in units of force evaluation.

Now we define a symmetric error estimator,

$$E_S(p,U:\Delta t)=e(p,U:\Delta t)+e(p',U':-\Delta t), \tag{10}$$

where $e(p,U:\Delta t)$ is a local error at (p,U) when the system is integrated by some integrator with a step size Δt, and the integrator maps (p,U) on (p',U'). The local error is assumed to increase monotonically with Δt. We will define the local error later. If the integrator is reversible, Eq. (10) is obviously symmetric under the exchange:

$$(p,U,\Delta t)\leftrightarrow(p',U',-\Delta t). \tag{11}$$

Namely, this means

$$E_S(p,U:\Delta t)=E_S(p',U':-\Delta t). \tag{12}$$

The adaptive step size is then determined by solving a symmetric error equation,

$$E_S(p,U:\Delta t)=\mathcal{T}. \tag{13}$$

The tolerance $\mathcal{T}$ should be kept constant during the MD simulation. The adaptive step size determined by Eq. (13) takes the same value at the reflected point $(-p',U')$. Therefore we find that an integrator with the adaptive step size determined by Eq. (13) is reversible.

Any local error can be defined provided that it increases monotonically with Δt. Our local error is defined as follows.

First, we integrate (p,U) by the two-step integrator $T^2(\Delta t)$ and the one-step integrator $T(2\Delta t)$.

$$T^2(\Delta t):(p,U)\rightarrow(p',U'), \tag{14}$$

$$T(2\Delta t):(p,U)\rightarrow(\tilde{p}',\tilde{U}'). \tag{15}$$

If Δt is not too large, (p',U') and $(\tilde{p}',\tilde{U}')$ should be close to each other. We define the local error at (p,U) by

$$e(p,U:\Delta t)=\sum_{\mu,x}^{4,V}\left(1-\frac{1}{N_c}\mathrm{Re\ Tr}U'^\dagger_\mu(x)\tilde{U}'_\mu(x)\right)\Big/\ (4V), \tag{16}$$

where V is the volume of the lattice. One could also use the momenta in the definition of the local error. Similarly, we integrate (p',U') by $T^2(-\Delta t)$ and $T(-2\Delta t)$ in the inverse time direction,

$$T^2(-\Delta t):(p',U')\rightarrow(p,U), \tag{17}$$

$$T(-2\Delta t):(p',U')\rightarrow(\tilde{p},\tilde{U}). \tag{18}$$

Since the integrator is reversible, the calculation of Eq. (17) is not needed. The local error at (p',U') is also defined like Eq. (16),

$$e(p',U':-\Delta t)=\sum_{\mu,x}^{4,V}\left(1-\frac{1}{N_c}\mathrm{Re\ Tr}U^\dagger_\mu(x)\tilde{U}_\mu(x)\right)\Big/\ (4V) \tag{19}$$

In the case of the leapfrog integrator of Eq. (9), we need four force evaluations to construct the symmetric error estimator E_S, instead of just two for the evolution $T^2(\Delta t)$.

Equation (13) is a nonlinear equation. In general, it should be solved numerically, e.g., by iterative bisection. With our definition of the symmetric error estimator, however, we can anticipate the scaling behavior of Eq. (13) and use it to accelerate convergence. The vector potentials evolved by the leapfrog integrator have $O(\Delta t^3)$ errors,

$$\tilde{A}'_\alpha=A'_\alpha+O(\Delta t^3). \tag{20}$$

Therefore,

$$U'^\dagger\tilde{U}'=\exp(-iA'_\alpha\lambda_\alpha)\ \exp(i\tilde{A}'_\alpha\lambda_\alpha) \tag{21}$$

$$\approx 1+ic_\alpha\lambda_\alpha O(\Delta t^3)+d_{\alpha\beta}\lambda_\alpha\lambda_\beta O(\Delta t^6), \tag{22}$$

where λ_α are SU(3) generators, and c_α and $d_{\alpha\beta}$ are some real constants whose explicit values are not important here. Taking Re and Tr of Eq. (22) and substituting it into Eqs.(16) and (19), we find that in the leading order the symmetric error estimator starts with $O(\Delta t^6)$. This behavior is verified

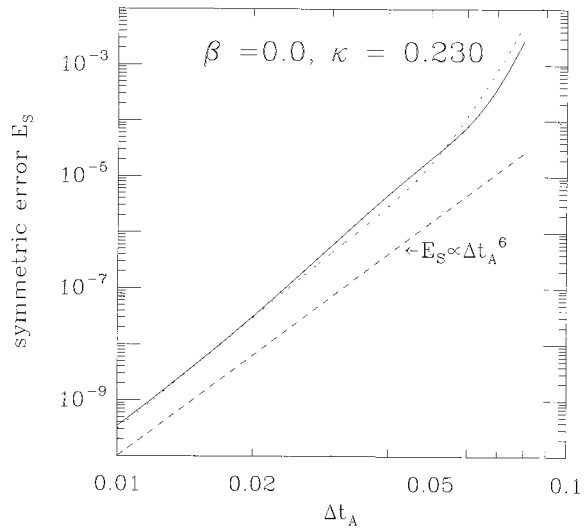

FIG. 2. Symmetric error E_S versus adaptive step size Δt_A, for two configurations of size 4^4 at $\kappa=0.230$. The straight line $E_S \propto \Delta t_A^6$ is shown for comparison.

numerically, as illustrated in Fig. 2: if Δt is not too large, the symmetric error estimator behaves like $E_S \propto \Delta t_A^6$. This property is used for solving Eq. (13). Choose some initial Δt_{A1} for the step size and calculate $E_{S1}=E_S(\Delta t_{A1})$. If Δt_{A1} does not satisfy the symmetric error equation (13) then input the second trial value Δt_{A2}, which is the solution of

$$\ln\left(\frac{\mathcal{T}}{E_{S1}}\right) = 6\,\ln\left(\frac{\Delta t_{A2}}{\Delta t_{A1}}\right). \tag{23}$$

If further trials are necessary, the following approximation can be used [5]:

$$\ln\left(\frac{\Delta t_{A3}}{\Delta t_{A2}}\right) = \frac{\ln(\Delta t_{A2}/\Delta t_{A1})}{\ln(E_{S2}/E_{S1})}\,\ln\left(\frac{\mathcal{T}}{E_{S2}}\right). \tag{24}$$

This recurrence is continued until Eq. (13) is satisfied to sufficient accuracy, and then a new configuration (p',U') is stored. Note that two strategies are available, just like for the stopping criterion of the linear solver in the force calculation: either the initial guess Δt_{A1} is invariant under time-reversal (e.g., it is equal to the average step size), and the accuracy to which Eq. (13) must be satisfied can be set arbitrarily low; or the initial guess makes use of past information (e.g., it is equal to the previous step size), and Eq. (13) must be satisfied exactly. We use the second method, and take the previous result as our initial guess. We then solve Eq. (13) to within 5%. Since we do not solve Eq. (13) exactly, we introduce a tiny, controllable source of irreversibility in the dynamics: the step size under time reversal could be different by about $\frac{1}{6}5\%\sim 1\%$. For this exploratory study, we have not considered this aspect further.

III. EFFICIENCY

We examine the method for full QCD ($N_c=3$) at several parameters (β,κ,volume) listed in Table I. We chose $\beta=0$ in several instances, to eliminate the gauge part of the action and hopefully to be more sensitive to the energy barriers

TABLE I. Run parameters.

Case	β	Size	κ	Traj. length	Tolerance ($\pm 5\%$)
A	0.0	4^4	0.215	0.8	10^{-4}
B	0.0	4^4	0.230	0.4	8×10^{-6}
C	5.4	4^4	0.162	1.0	10^{-6}
D	0.0	8^4	0.215	0.3	10^{-7}

coming from the fermionic part. The effect of changing the quark mass can be obtained by comparing cases *A* and *B*, that of changing the volume by comparing *A* and *D*. The adaptive step size is determined by the symmetric error equation (13), with the tolerance set as per Table I. The adaptive step sizes are summed up from the beginning of the trajectory and when the total trajectory length becomes greater than the trajectory length of Table I, a Metropolis test is performed.

Figure 3 shows histograms of the adaptive step size at $\beta=0.0$, $\kappa=0.215$, and 0.230 (cases *A* and *B* in Table I). The distribution remains strongly peaked. This is also true of cases *C* and *D*. The average step size $\langle \Delta t_A \rangle$ and its relative variance are summarized in Table II, where the relative variance σ is defined by

$$\sigma^2 = \frac{1}{N}\sum_{i=1}^{N}\left(\frac{\Delta t_{Ai}}{\langle \Delta t_A\rangle}-1\right)^2. \tag{25}$$

As the quark mass m_q decreases, the energy barriers in phase space become higher and sharper, so that one would expect the variation of the step size to increase. Indeed this is what happens, and the relative variance in cases *A* and *B* increases roughly like $1/m_q$, where $m_q \propto (\kappa^{-1}-\kappa_c^{-1})$ and $\kappa_c=1/4$ at $\beta=0$. On the other hand, as the volume increases, the relative variance seems to decrease sharply, like $1/\sqrt{V}$ or faster (see cases *A* and *D*). Perhaps this can be explained by considering the relative fluctuations of the effective action $-\ln \det D$: as the volume increases, the relative fluctuations

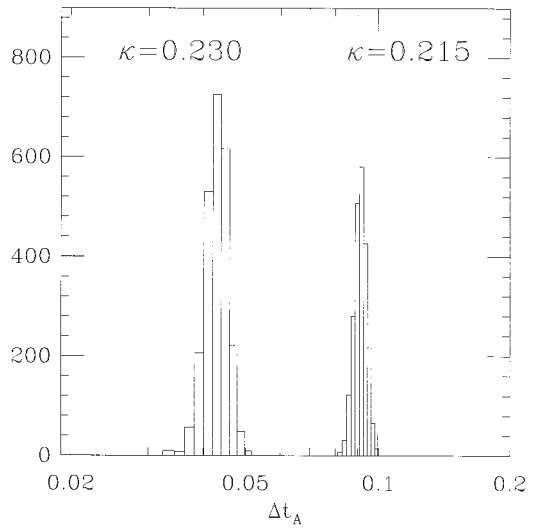

FIG. 3. Histograms of adaptive step size Δt_A, for 4^4 lattices at $\kappa=0.215$ and 0.230. The logarithmic scale shows the increase with κ of the relative variance.

TABLE II. Results of the adaptive step-size method and fixed step size Δt_{HMC} of the HMC algorithm. $\langle$traj. length$\rangle$ stands for the average total trajectory length. The fixed step size Δt_{HMC} is determined so that it gives the same acceptance as the adaptive step size method.

Case	$\langle \Delta t_A \rangle$	σ (%)	$\langle$(Traj. length)$\rangle$	Acceptance (%)	Δt_{HMC}
A	0.0911(3)	3.3	0.91	36(2)	0.0897(13)
B	0.0431(1)	5.6	0.44	57(2)	0.0419(09)
C	0.0673(1)	0.8	1.08	87(2)	0.0688(80)
D	0.03281(2)	0.6	0.33	63(2)	0.0328(10)

decrease, so that the system tends to stay at some average distance from the energy barriers, rather than bouncing off them.

From the schematic picture of Fig. 1, it is expected that the approach of an energy barrier causes a reduction in the adaptive step size, and at the same time an increase in the number of iterations taken by the solver to converge. Figure 4 shows Δt_A versus the number of iterations in the solver: the expected anticorrelation between them can be observed, and becomes more pronounced as the quark mass is reduced.

In order to compare the adaptive method with the conventional HMC algorithm, we define the efficiency of the adaptive method as follows.

First, find the fixed step size Δt_{HMC} of the HMC simulation which gives the same acceptance as the adaptive step-size method. The total trajectory length of the HMC simulation is set to the average total trajectory length $\langle$(trajectory length)$\rangle$ of the corresponding adaptive step-size method's case. We performed the HMC simulations with several step sizes and determined the corresponding step size of the HMC algorithm by interpolating those results. The results of the corresponding step size are summarized in Table II. For the acceptance of the adaptive step-size method, see Table II.

Then, define the gain by

$$g_A = \langle \Delta t_A \rangle / \Delta t_{HMC} . \tag{26}$$

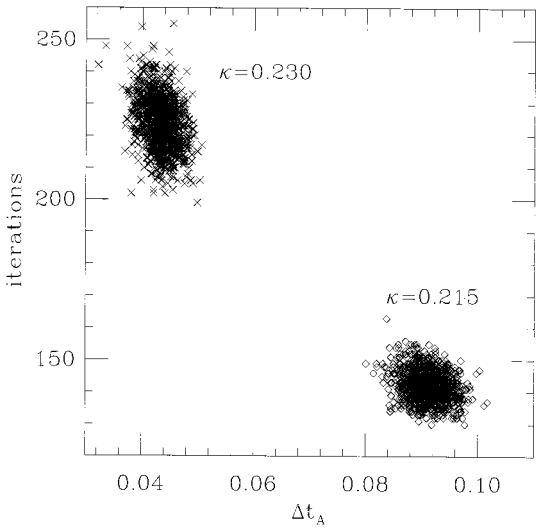

FIG. 4. Number of iterations in the solver (BiCGγ_5) versus adaptive step size.

When $g_A > 1$, the adaptive step size method really takes larger steps on average, without compromising the acceptance. However the real efficiency of the method can only be assessed by taking into account the overhead of determining the adaptive step-size, since additional force evaluations are necessary.

From the definition of our symmetric error estimator Eqs. (14)–(19), we know that one construction of E_S needs four force evaluations. Call R_T the average number of trial steps needed to solve Eq. (13). After $4R_T$ force evaluations, we have determined the step size Δt_A and use the integrator $T^2(\Delta t_A)$ [Eq. (14)] to advance the dynamics by two steps Δt_A. Therefore the cost of the method is $2R_T$ force evaluations per step, compared with 1 force evaluation per step for standard HMC.

Real efficiency will be achieved if the number of force evaluations per unit time decreases, i.e., if $2R_T < g_A$. Results for the gain g_A and the cost $2R_T$ are summarized in Table III. For all cases we studied, real efficiency is not achieved.

It is disappointing to see how small the gains g_A are. The reason for such small gains can be understood by considering the behavior of the Hamiltonian. The dependence of the energy violation at each step with the step size is, in general, nonlinear. Therefore it is not necessary that a small local error correspond to a proportionally small energy violation of the Hamiltonian. Figure 5 shows $|\Delta H|$, the absolute value of the energy violation after one integration step, versus the local error E_S. The two clusters of points correspond to fixed step sizes $\Delta t = 0.04$ and 0.08. No strong correlation between E_S and ΔH can be observed. This is further evidenced by the dashed lines, which are the result of fitting to a scaling law $\sqrt{\langle \Delta H^2 \rangle} = E_S^b$: for the larger step size, b is almost zero. Therefore, it becomes clear that fixing E_S and varying Δt adaptively cannot have a strong effect on the acceptance, which solely depends on ΔH.

Two approaches could be used to improve the efficiency of our scheme.

(i) Decrease the overhead: instead of estimating the error by comparing $T^2(\Delta t)$ with $T(2\Delta t)$, one could replace the

TABLE III. Gain g_A, average number of trial steps R_T, and cost per step.

Case	g_A	R_T	Cost per step ($=2R_T$)
A	1.016(15)	2.25	4.50
B	1.029(22)	2.45	4.90
C	1.0(1)	1.13	2.26
D	1.00(3)	1.15	2.30

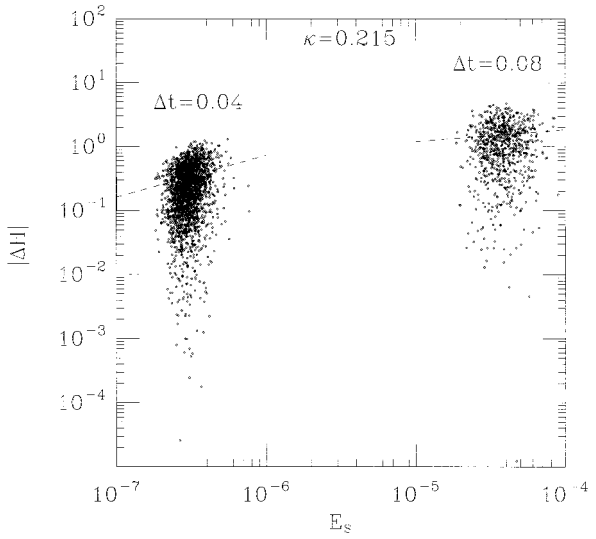

FIG. 5. $|\Delta H|$ versus the local error E_S, on a 4^4 lattice at $\kappa=0.215$. The step size is fixed at $\Delta t=0.04$ and 0.08. ΔH is the change in the total energy after one integration step. The dotted lines result from fitting to the form $\sqrt{\langle \Delta H^2\rangle}=E_S^b$, and show the correlation (or absence of) between the two quantities. $\langle \Delta H^2\rangle$ is obtained by dividing the data in 10 bins and averaging the values ΔH^2 in each bin.

latter by an Euler integrator, which requires no additional force evaluation. Note that the error equation (10) remains symmetric under the exchange $(p,U,\Delta t)\leftrightarrow(p',U',-\Delta t)$ even though the Euler integrator is not time reversible. The problem we found with that approach is that, for the large step sizes used on our small lattices, the error (10) no longer obeyed a simple scaling law (22) as a function of the step size. Then the number of iterations needed to solve Eq. (13) increased, defeating the expected reduction in overhead. On larger lattices with smaller step-sizes, this problem would be milder.

(ii) Change the definition of the error (16), so that it is better correlated with $|\Delta H|$, the energy violation at each step. Note that $|\Delta H|$ itself cannot be chosen, because it does not increase monotonically with the step-size: in that case Eq. (13) admits multiple solutions; the overhead of converging to one of them, and the same one under time-reversal, increases considerably. With our definition, Eq. (16), the error is only weakly correlated with $|\Delta H|$, but the situation again seems to improve with smaller step sizes, on larger lattices (compare the two dashed lines in Fig. 5). Nonetheless, it would be desirable to control the step size with a more relevant quantity than Eq. (16), since all that matters in the end is energy conservation.

Finally, instead of varying the step size, one could vary adaptively the couplings of the Hamiltonian H, Eq. (3), at each step, or even include some new operators in H, trying to tune them so as to best conserve energy. The general difficulty with that approach is to find an error Eq. (16), which varies monotonically with the couplings of H.

IV. CONCLUSION

We have implemented an adaptive step-size method for hybrid Monte Carlo simulations, and tested it at several parameters (β,κ,volume). The relative variance of the step size increases for small quark masses and small volumes. The average step size seems somewhat larger than the corresponding fixed step size at the same acceptance. But this gain is more than offset by the overhead of determining the adaptive step size. It seems very difficult to achieve real gains in efficiency, because conservation of energy, which is necessary for high Metropolis acceptance in the HMC algorithm, is poorly correlated with the conventional error governing the adaptive step size.

A plausible extrapolation from our results would indicate that the relative variance of the step size scales like $m_q^{-1}V^{-1/2}$, i.e., as $(m_\pi L)^{-2}$, where m_π is the pion mass and L the physical size of the lattice. This quantity normally remains constant as the continuum limit $a\to 0$ of the lattice theory is taken, so that the relative fluctuations in the adaptive step size Δt would tend to a constant. Even if this analysis is no more than plausible at this stage, it is clear that the two limits $m_q\to 0$, $V\to\infty$ have opposite effects on the fluctuations of Δt, making it unlikely that such fluctuations become very large on present lattice sizes. This observation is consistent with the limited fluctuations (a factor of 2 or so) in the number of iterations needed by the solver to compute the force in the largest QCD simulations [8].

Thus it appears that QCD is much ''easier'' to simulate than the Kepler problem: in lattice QCD, the force on the gauge links varies little in magnitude, and the curvature of the molecular dynamics trajectory is rather small. One intuitive explanation is that the QCD force is dominated by short-range UV contributions, which drown the IR component sensitive to the energy barrier $\det D\sim 0$.

ACKNOWLEDGMENTS

We thank D. Stoffer for helpful discussions. T.T. is supported in part by the Japan Society for the Promotion of Science. Ph. de F. thanks Hiroshima University and Tsukuba University, especially Professor O. Miyamura and Professor Y. Iwasaki, for hospitality during this project. T.T. thanks Professor M. Gutknecht for hospitality.

[1] S. Duane, A.D. Kennedy, B.J. Pendleton, and D. Roweth, Phys. Lett. B **195**, 216 (1987)
[2] M. Lüscher, Nucl. Phys. B **418**, 637 (1994)
[3] B. Bunk, K. Jansen, B. Jegerlehner, M. Lüscher, H. Simma, and R. Sommer, Nucl. Phys. B (Proc. Suppl.) **42**, 49 (1995); B. Jegerlehner, *ibid.* **42**, 879 (1995); A. Boriçi and Ph. de Forcrand, Nucl. Phys. B **454**, 645 (1995); C. Alexandrou, A. Borrelli, Ph. de Forcrand, A. Galli, and F. Jegerlehner, *ibid.* **456**, 296 (1995); B. Jegerlehner, *ibid.* **465**, 487 (1996); K. Jansen, B. Jegerlehner, and C. Liu, Phys. Lett. B **375**, 255 (1996); A.

Borrelli, Ph. de Forcrand, and A. Galli, Nucl. Phys. B **477**, 809 (1996).
[4] M. Creutz and A. Gocksch, Phys. Rev. Lett. **63**, 9 (1989).
[5] D. Stoffer, Computing **55**, 1 (1995).
[6] Ph. de Forcrand, Nucl. Phys. B (Proc. Suppl.) **47**, 228 (1996).
[7] J.C. Sexton and D.H. Weingarten, Nucl. Phys. B **380**, 665 (1992).
[8] T. Lippert, SESAM collaboration (personal communication).