Numerical Analysis of Accelerated Stochastic Algorithms near a Critical Temperature

Elbio Dagotto and John B. Kogut
Loomis Laboratory, Department of Physics, University of Illinois at Urbana-Champaign, Urbana, Illinois 61801
(Received 31 October 1986)

We present a numerical study of stochastic differential equations using the XY model in two dimensions as an example. "Accelerated" and "hybrid" algorithms have been implemented and show their advantages over more standard methods. The (partial) evasion of critical slowing down by accelerated algorithms is studied on a sequence of lattices 16^2, 32^2, and 64^2.

PACS numbers: 11.15.Ha, 05.50.+q

The study of stochastic differential equations has attracted considerable interest recently.[1] Studies of models have shown that the Langevin and molecular-dynamics algorithms are competitive with the standard Monte Carlo technique. However, the molecular-dynamics method does not necessarily satisfy the ergodic hypothesis so that its equivalence with the microcanonical ensemble is suspect. On the other hand, the Langevin equation typically explores phase space at the rate of a random walk and this is too slow for many applications. To attack these additional problems, a "hybrid" algorithm has been proposed[2] which interpolates between the Langevin and molecular-dynamics methods. Moreover, Parisi[3] has proposed a method based on stochastic equations to avoid critical slowing down.[4] Recently, this approach has been analyzed by Batrouni *et al.*[5] In the language of the hybrid algorithm the idea is that by our cleverly choosing the (arbitrary) kinetic energy of the dynamical Lagrangean it is possible to "accelerate" the low-momentum components of the fields such that they evolve in "time" as fast as the ultraviolet modes.

The purpose of this Letter is to present a high-statistics study of the planar XY model near the Kosterlitz-Thouless phase transition. We have implemented both the acceleration and the hybrid improvements on lattices of sizes 16^2, 32^2, and 64^2 to see if the accelerated algorithm eliminates critical slowing down at an *interacting* critical point.

First we set up the basic equations of the simulation. For more details see Kogut,[6] where a preliminary study of the XY model on small lattices was presented. Consider the Lagrangean

$$L=\frac{1}{2}\sum_{i,j}\dot{\theta}_i M_{ij}\dot{\theta}_j-\beta\sum_{i,\hat{l}}\cos(\theta_i-\theta_{i+\hat{l}}), \tag{1a}$$

where

$$M_{ij}=[(1-x)-\tfrac{1}{8}x(\partial_1^2+\partial_2^2)]_{ij}. \tag{1b}$$

θ_i is an angular variable on the site $i=(i_1,i_2)$ of a square lattice of size L, $\hat{l}=1,2$ labels directions, x is a free parameter, and $\partial_1^2,\partial_2^2$ are discrete second derivatives. The Euler-Lagrange equation is

$$\theta_i(t+dt)=\theta_i(t)+\left(\frac{\theta_i(t)-\theta_i(t-dt)}{dt}\right)dt-\beta(dt)^2M_{ij}^{-1}\frac{d}{d\theta_j}\sum_{\hat{l}=\pm1,\pm2}\cos(\theta_j-\theta_{j+\hat{l}}). \tag{2}$$

dt is a discrete time step and the second term of the right-hand side represents velocity multiplying dt. In the thermodynamic limit, the equilibrium statistical mechanics of Eqs. (1) and (2) is equal to the canonical ensemble with the usual XY action. The form of the kinetic energy in Eq. (1b) can be justified as follows: For $x=0$ we obtain the standard $\frac{1}{2}\dot{\theta}^2$ term used in molecular-dynamics simulations where the low-momentum modes evolve very slowly. For $x\neq0$ it can be shown (in a free theory, for example) that the time step for the infrared ($p=0$) and ultraviolet ($p=\pi$) modes are related by $dt(0)=dt(\pi)(1-x)^{-1/2}$. So, choosing x close to 1 the algorithm accelerates the low-momentum modes and relatively few iterations of Eq. (2) should be needed to generate a statistically independent field configuration. For an interacting theory, it is assumed that M will still be a good *Ansatz* to attenuate critical slowing down.

Note that the last term of Eq. (2) can be evaluated in momentum space where M is diagonal, i.e.,

$$M_{ij}^{-1}\frac{d}{d\theta_j}\sum_l\cos(\theta_j-\theta_{j+l})=F^{-1}\left[M^{-1}(p)F\left(\frac{d}{d\theta_j}\sum_l\cos(\theta_j-\theta_{j+l})\right)\right], \tag{3}$$

where F is a Fourier-transformation operator that can be efficiently implemented with use of fast-Fourier-transform (FFT) subroutines. Since the entire lattice is updated simultaneously we need only two FFT's per sweep. This is a potential advantage of stochastic differential equations over the Monte Carlo technique.

Above we have described the molecular-dynamics equations with acceleration. Now consider the hybrid technique. The basic idea is that we let the system evolve in time following Eq. (2) during $N_{RF}-1$ sweeps (N_{RF} denotes the "number of refreshments"). In the next sweep, the lattice is updated using the Langevin equation. This process is now repeated during the whole simulation. If $N_{RF}=1$ ($N_{RF}\to\infty$), we recover the usual Langevin (molecular-dynamics) algorithms. In general, there exists an optimal value of N_{RF} ($\neq 1$ or ∞) that improves both limits. Note that Eq. (2) is very similar to a Langevin equation because the velocity should lie in a Boltzmann distribution $\exp(-\frac{1}{2}\sum \dot\theta_i M_{ij}\dot\theta_j)$, which can be associated with the distribution of Langevin noise (see Ref. 5). So, in practice, we only need to replace the velocity by a random variable generated with a Gaussian distribution to convert a molecular-dynamics step into a Langevin step.

In our simulation we have two free parameters: N_{RF} and x. To analyze the advantages of our various algorithms we have measured the relaxation time[4] for the observable A defined as[7]

$$\tau_A=\int_0^\infty dt\left[\frac{\langle A(0)A(t)\rangle-\langle A\rangle^2}{\langle A^2\rangle-\langle A\rangle^2}\right]. \tag{4}$$

We have measured the action S and the susceptibility χ defined as

$$\chi=(1/L^2)\sum_{i,j=1,\ldots,L^2}\cos(\theta_i-\theta_j). \tag{5}$$

χ is also the Fourier transform of the correlation function at zero momentum, so that it will give us information about the behavior of the infrared modes.

To study critical slowing down, we simulated the model on 16^2, 32^2, and 64^2 lattices. We found that acceleration is relatively more effective on large lattices than on small ones.

Now we describe our numerical results.

(i) First we have studied small lattices to find an optimal value for N_{RF} in the unaccelerated algorithm ($x=0$). In Fig. 1 we show τ_χ and τ_S as functions of N_{RF} on a 16^2 lattice with $\beta=0.9$ (near the critical temperature) and $dt=0.1$ Throughout the text τ and N_{RF} will be given in units of number of sweeps. Each point of Fig. 1 has been obtained with 1.2×10^6 sweeps through the lattice (plus 5×10^4 sweeps for thermalization). The errors come from our dividing the sample into a small number of equal groups.

Note that $\tau_{\chi,S}$ are very large in the Langevin and microcanonical limits. In these cases one needs a huge number of sweeps between measurements of observables to get independent numbers. For intermediate values of N_{RF} the situation is clearly better. There is an extended regime ($N_{RF}\sim 10$–100) where τ_χ is reduced by a factor between 8 and 16 with respect to the Langevin limit. A qualitatively similar situation exists for τ_S. In this case

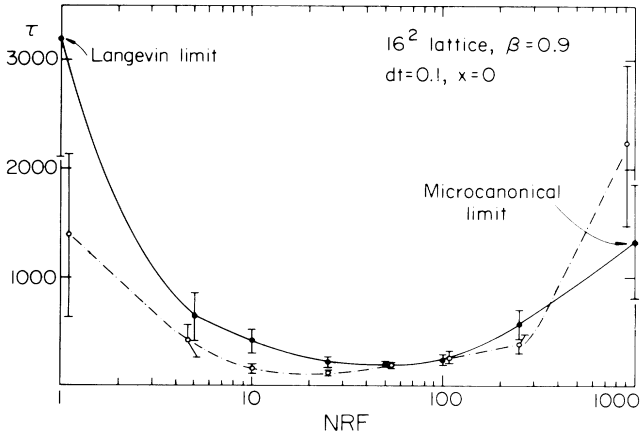

FIG. 1. Relaxation times for the susceptibility (continuous line) and the action (dashed-dotted line) as a function of N_{RF} in the unaccelerated case.

its optimal N_{RF} is shifted to smaller values.[8] It can be shown[9] by development of a generalized Fokker-Planck equation that the systematic errors are independent of N_{RF} so that Fig. 1 represents a fair comparison of the efficiencies of the various schemes.

We have repeated our calculation for $\beta=0.8$ and have also considered a 8^2 lattice. The position of the wide zone where $\tau_{\chi,S}$ have their minima does not change appreciably, and so we conclude that it is reasonable to choose N_{RF} from a small lattice study. We select $N_{RF}=25$ as our optimal value. Here both τ_S and τ_χ are small.

As the next step in our study we accelerate the system using $x\neq 0$ in Eq. (1b). We have done an analysis similar to the one described above for the unaccelerated case, but now choose x between 0 and 1 (at $N_{RF}=25$). Since the relaxation times are not as large as in the Langevin case we did only 360000 iterations. For $x\geq 0.99$ the results are relatively insensitive to x.

In Table I we show $\tau_{\chi,S}$ for some special cases. The combination $N_{RF}=25$, $x\sim 1$ is the optimal one which improves the unaccelerated Langevin result by a factor ~ 40 for the susceptibility. Note that the gain between

TABLE I. Relaxation times for χ and S on a 16^2 lattice, $\beta=0.9$, and $dt=0.1$ for some values of N_{RF} and x. For the unaccelerated case we used 1 200 000 iterations, while for the accelerated 720 000 iterations.

N_{RF}	$1-x$	τ_χ	τ_S
1	1	3200	1400
1000	1	1300	2200
25	1	220	130
1	10^{-8}	410	240
1000	10^{-8}	900	2000
25	10^{-8}	75	85

the unaccelerated and accelerated algorithms is very large for $N_{RF}=1$ but decreases with N_{RF} such that for the optimal $N_{RF}=25$ the factor is only ~ 3. Therefore, the optimized unaccelerated hybrid algorithm is a considerable improvement over the unaccelerated Langevin procedure and there is relatively less to be gained by acceleration. We shall see below, however, that the gain ~ 3 increases with the size of the lattice and is related to the attenuation of critical slowing down by the accelerated algorithm.

What about systematic errors in the mean value of S and χ? At $x=0$, changing N_{RF} led to no changes in S or χ. However, with turning on of the acceleration these mean values become systematically smaller than the unaccelerated results by about $\sim 0.5\%$ ($\sim 3\%$) for S (χ) (the same occurs for bigger lattices). This effect should be explained by construction of the effective action that our algorithm is simulating by use of the Fokker-Planck equation.

(ii) In Fig. 2 we show the results for τ_χ as a function of β for a 32^2 lattice. A clear peak in τ_χ near the Kosterlitz-Thouless transition is present for the unaccelerated algorithm, even with the optimal value of N_{RF}. On the other hand, the calculation with $N_{RF}=25$, $x\sim 1$ reduces the height of the peak by a factor 5. Since this gain seems to increase with the lattice size (a factor of 3 for 16^2 and 5 for 32^2) we believe that accelerated hybrid algorithms do, in fact, eliminate (partially) critical slowing down. Also some results with $N_{RF}=5$, $x=0$ are included in Fig. 2 to show the impressive gain an optimized hybrid algorithm has over algorithms close to the Langevin limit.

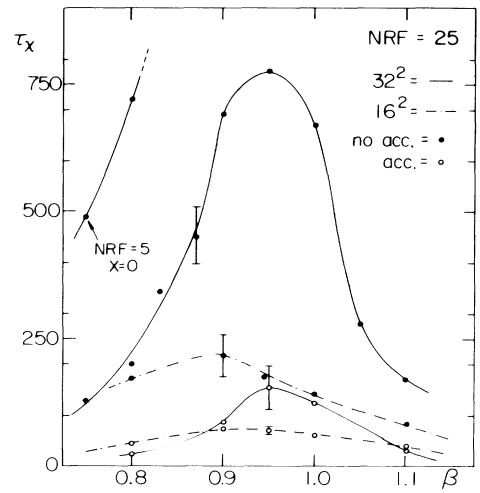

FIG. 2. Relaxation time for the susceptibility in a 32^2 lattice at $N_{RF}=25$, $dt=0.1$. The points without acceleration (filled circles) have been obtained using between 720000 and 240000 iterations, while for the accelerated ones (open circles) we have used 240000 iterations and $1-x=10^{-8}$. We show only some typical error bars (their size increases near the peaks). We have also plotted results for $N_{RF}=5$, $x=0$ (showing the beginning of a big peak) and for a 16^2 lattice. The lines (solid for 32^2, dashed-dotted for 16^2) guide the eye.

A comment about the stability of the algorithm: For $N_{RF}=25$, $x=0$, and $\beta=0.95$, we observe stability problems that were solved with $dt=0.05$. For cases with bigger τ the situation is worse since the divergences are present for a wide range of β near the critical temperature and the value of dt must be reduced further to achieve convergence. However, the hybrid accelerated case has no stability problems. Therefore, it seems to be that the lower the relaxation times the better the stability of the algorithm.

What about central processing unit (CPU) time with and without the improvements described in this paper? Let us take as an example $N_{RF}=25$ in a 32^2 lattice. The ratio in CPU time between the accelerated and unaccelerated cases is around 1.5. And the same ratio between the code with $N_{RF}=1$, $x\sim 1$ and $N_{RF}=25$, $x\sim 1$ is about 2–3 due to the time spent in generating Gaussian random numbers at every sweep. Of course these estimates will change depending on the details of the code, but at least they show us the order of magnitude of the effort. It is clear that the use of FFT subroutines does not involve an appreciable increase in computing time, while the implementation of the hybrid method reduces the CPU time with respect to the Langevin limit.

(iii) In Fig. 3 we have plotted τ_χ vs β on a 64^2 lattice. Each point represents an average over 120000 sweeps through the lattice. The value of N_{RF} was changed to 10 (still in the region where τ is minimum in Fig. 1) because we observe stability problems with $N_{RF}=25$ near the critical temperature (all the other cases of Table I have the same problem for a larger range of β). In Fig. 3 we can see a peak in τ_χ near $\beta\sim 1.0$. Changing the

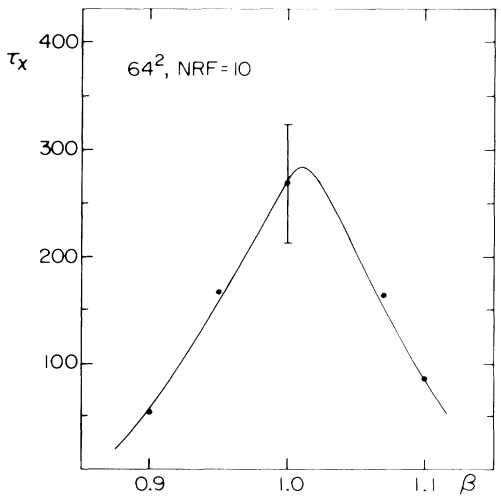

FIG. 3. τ_χ vs β on a 64^2 lattice with $N_{RF}=10$ and $1-x=10^{-8}$.

value of x, we could not avoid the existence of this peak. This shows that acceleration which uses free field theory ideas does *not* completely eliminate the problem of critical slowing down, but it certainly attenuates the divergence by a big factor. It has been argued in Ref. 5 that a more convenient method to accelerate algorithms is to measure the relaxation time for different modes, fit them with a function $\tau(p)$, and then use $M^{-1}(p)=\tau(p)$. In practice we found this method difficult to implement because of the large number of sweeps required to measure relaxation times (even with a limited accuracy) on large lattices. We believe that efforts to improve acceleration should instead be focused on a more intelligent choice of L.

Summarizing, in this Letter we have presented a study of the XY model in two dimensions by use of stochastic differential equations. We have shown that accelerated and hybrid algorithms represent a very important improvement over the more standard techniques, and they require only a small computational effort to be implemented. The combination of both methods has the best performance (even the ultraviolet stability is improved) reducing critical slowing down drastically.

This work was done as part of the Materials Research Program under Contract No. DMR83-16981. It was also supported in part by the National Science Foundation under Grants No. DMR-84-15063 and No. PHY82-01948. The computer simulations were done on a FPS-264 which is supported in part by National Science Foundation Grant No. PHY83-04872.

[1]J. Kogut, University of Illinois at Urbana-Champaign Report No. 75, 1985 (to be published).

[2]S. Duane, Nucl. Phys. **B257 [FS14]**, 652 (1985).

[3]G. Parisi, in *Progress in Gauge Field Theory,* edited by G. 't Hooft *et al.* (Plenum, New York, 1984), p. 531.

[4]*Monte Carlo Methods in Statistical Mechanics,* edited by K. Binder (Springer-Verlag, New York, 1979), Chapts. 1 and 6.

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

[6]J. Kogut, Nucl. Phys. B275 **[FS17]**, 1 (1986).

[7]There are other criteria to obtain the temporal correlation length. For example τ_A can be defined as the time where the time correlation equals 10% of its peak value. However, using this criterion we have obtained huge error bars in τ_A because of the erratic behavior of the tail of the time correlation function.

[8]This is a reasonable result since it has been shown in simple examples (Ref. 2) (harmonic oscillators) that the optimal N_{RF} is inversely proportional to the "characteristic frequency" of the observable. This frequency is (intuitively) bigger for S than for χ.

[9]S. Duane and J. Kogut, University of Illinois at Urbana-Champaign Report No. 15, 1986, Nucl. Phys. (to be published).