Collective Monte Carlo Updating for Spin Systems

Ulli Wolff
Institut für Theoretische Physik, Universität Kiel, D-2300 Kiel, West Germany
(Received 13 October 1988)

A Monte Carlo algorithm is presented that updates large clusters of spins simultaneously in systems at and near criticality. We demonstrate its efficiency in the two-dimensional O(n) σ models for $n=1$ (Ising) and $n=2$ (x-y) at their critical temperatures, and for $n=3$ (Heisenberg) with correlation lengths around 10 and 20. On lattices up to 128^2 no sign of critical slowing down is visible with autocorrelation times of 1–2 steps per spin for estimators of long-range quantities.

PACS numbers: 05.50.+q, 11.15.Ha

Recently Swendsen and Wang (SW)[1] have put forward a novel Monte Carlo algorithm for Potts spin models using ideas from percolation theory.[2] Whole clusters of spins are thus enabled to move in one step, and consequently critical slowing down is greatly improved. The information residing in the clusters can also be used to construct reduced variance estimators[3] for physical observables. The combination of both advantages was quite useful in recent studies of the physics of the four-dimensional φ^4 theory in the Ising limit.[4] In Ref. 16, the sign of a variable length φ^4 field in two dimensions has been updated by SW, leading to the expected acceleration. The SW method has been developed further in two directions: In Ref. 5 it is claimed that critical slowing down in the two-dimensional Ising model is eliminated completely by a synthesis of SW with multigrid ideas. Other research[6-8] was motivated by the desire to simulate models other than Potts systems, and, in particular, ones with continuous fields such as the x-y model. There the results on slowing down have been essentially negative so far; improved estimators could be exploited in Ref. 6 for the O(3) σ model albeit only for a variant action. In this Letter we present another percolation-inspired algorithm that allows the simulation of spin models with standard actions. It is tested numerically in the two-dimensional Ising and x-y models at criticality. Preliminary tests are also performed for the O(3) model.

The algorithm.—For simplicity, we only consider O(n) σ models on a cubic lattice Λ of $|\Lambda|$ sites with periodic boundary conditions,

$$Z=\prod_{x\in\Lambda}\int_{S_{n-1}} d\sigma_x \exp\left\{\beta\sum_{\langle xy\rangle}\boldsymbol{\sigma}_x\cdot\boldsymbol{\sigma}_y\right\}, \tag{1}$$

where $\boldsymbol{\sigma}_x$ are unit vectors in $\mathbb{R}^n$, β is the inverse coupling temperature, and the sum runs over all nearest-neighbor pairs. An important concept for the new algorithm is the generalization of the spin-flip operation $\boldsymbol{\sigma}_x\to-\boldsymbol{\sigma}_x$ in the Ising model. For $n\geq 2$ and any $\mathbf{r}\in S_{n-1}$ we define it as the reflection with respect to the hyperplane orthogonal to $\mathbf{r}$

$$R(\mathbf{r})\boldsymbol{\sigma}_x=\boldsymbol{\sigma}_x-2(\boldsymbol{\sigma}_x\cdot\mathbf{r})\mathbf{r}. \tag{2}$$

Clearly this is an idempotent operation[9]

$$R(\mathbf{r})^2=1, \tag{3}$$

and the action in (1) is invariant under global R transformations

$$[R(\mathbf{r})\boldsymbol{\sigma}_x]\cdot[R(\mathbf{r})\boldsymbol{\sigma}_y]=\boldsymbol{\sigma}_x\cdot\boldsymbol{\sigma}_y. \tag{4}$$

An elementary cluster update step now consists of the following sequence of operations:

(a) Choose a random reflection $\mathbf{r}\in S_{n-1}$ and a random lattice site $x\in\Lambda$ as the first point of a cluster $c\in\Lambda$ to be built. (b) Flip $\boldsymbol{\sigma}_x\to R(\mathbf{r})\boldsymbol{\sigma}_x$ and mark x. (c) Visit all links connecting $x\in c$ to its nearest neighbors y. The bond $\langle xy\rangle$ is activated with probability

$$\begin{aligned}P(\boldsymbol{\sigma}_x,\boldsymbol{\sigma}_y)&=1-\exp(\min\{0,\beta\boldsymbol{\sigma}_x\cdot[1-R(\mathbf{r})]\boldsymbol{\sigma}_y\})\\&=1-\exp\{\min[0,2\beta(\mathbf{r}\cdot\boldsymbol{\sigma}_x)(\mathbf{r}\cdot\boldsymbol{\sigma}_y)]\},\end{aligned} \tag{5}$$

and, if this happens, $\boldsymbol{\sigma}_y$ is flipped, and y is marked and adjoined to c. (d) Continue iteratively in the same way for all bonds leading to unmarked neighbors of newly adjoined sites until the process stops. Ergodicity of processes (a) to (d) is guaranteed by the fact that there is always a nonvanishing probability that c consists of only one site, and that there is always a reflection connecting any two spins. Then each configuration may be reached in principle by at most $|\Lambda|$ update steps. Detailed balance is also easily seen to be fulfilled. We consider two configurations $\{\boldsymbol{\sigma}_x\}$ and $\{\boldsymbol{\sigma}'_x\}$ that differ by a flip $R(\mathbf{r})$ on a cluster c. The transition probabilities W obey

$$\frac{W(\{\boldsymbol{\sigma}_x\}\to\{\boldsymbol{\sigma}'_x\})}{W(\{\boldsymbol{\sigma}'_x\}\to\{\boldsymbol{\sigma}_x\})}=\prod_{\langle xy\rangle\in\partial c}\frac{1-P(R(\mathbf{r})\boldsymbol{\sigma}_x,\boldsymbol{\sigma}_y)}{1-P(R(\mathbf{r})\boldsymbol{\sigma}'_x,\boldsymbol{\sigma}'_y)}=\exp\left\{\beta\sum_{\langle xy\rangle\in\partial c}\boldsymbol{\sigma}_x\cdot[R(\mathbf{r})-1]\boldsymbol{\sigma}_y\right\}=\exp\left\{\beta\sum_{\langle xy\rangle}(\boldsymbol{\sigma}'_x\cdot\boldsymbol{\sigma}'_y-\boldsymbol{\sigma}_x\cdot\boldsymbol{\sigma}_y)\right\}, \tag{6}$$

where the surface ∂c of c consists of all links $\langle xy\rangle$ with $x\in c$ and $y\notin c$. All probabilities for activating bonds *within* c are the same starting from $\{\boldsymbol{\sigma}_x\}$ or $\{\boldsymbol{\sigma}'_x\}$ because of (3) and (4). To grow a specific cluster c, its surface bonds must *not*

be activated, and these probabilities supply the noncanceling factors in (6). Note that $R(\mathbf{r})$ appears in the left argument of $P(\ldots,\ldots)$ because σ_x has already been flipped when $\langle xy\rangle \in \partial c$ is probed. The desired action-energy difference in the exponent finally arises in a way similar to the Metropolis algorithm.

The adequacy of discrete flips for continuous spins may seem surprising at first sight. Intuitively, it works for the following reason: If we choose $\mathbf{r}$ and start a cluster in a region where spins are nearly orthogonal to $\mathbf{r}$, then (2) is effectively a small deformation. Bond probabilities (5) typically lead to small clusters in this case. If $\mathbf{r}$ is (anti)parallel to the local spin direction, the cluster tends to grow until it reaches an energetically acceptable surface.

For the Ising case ($n=1$), the bond activation probabilities that we use are the same as in the SW process [$1-\exp(-2\beta)$ if the original spins are parallel and 0 otherwise]. The connection between both algorithms becomes clearer by consideration of a less efficient but equivalent implementation of the new algorithm.[10] It consists of our first activating all bonds of the lattice with the appropriate probabilities and decomposing all sites into clusters. These are the SW clusters c_{SW}. Then we pick a random site x, flip the cluster connected to x, and ignore the remaining ones. The one cluster that we flip is reached from any of its sites, and consequently the probability of our choosing a particular SW cluster is given by the fraction $|c_{\mathrm{SW}}|/|\Lambda|$ of sites it occupies. The mean size of flipped clusters is

$$\langle |c| \rangle = \left\langle \frac{1}{|\Lambda|} \sum_{c_{\mathrm{SW}}} |c_{\mathrm{SW}}|^2 \right\rangle , \tag{7}$$

which for $n=1$ is the improved estimator[6] for the magnetic susceptibility χ. We thus expect (and find)

$$\chi = \frac{1}{|\Lambda|} \left\langle \left(\sum_{x\in\Lambda} \sigma_x \right)^2 \right\rangle = \langle |c| \rangle \tag{8}$$

to hold. The average size of SW clusters in the critical Ising model, on the other hand, is seen to be constant from data in Ref. 6 since their number grows proportional to L^2. We conclude that with our variation of the SW algorithm for $n=1$, we invest a relatively higher fraction of the central processing unit time into large clusters. This should bring about an even better decorrelation behavior. To quote values for dynamical exponents is beyond the scope of the present paper. Here we rather concentrate on the applicability for $n>1$.

Numerical results.—In Table I a subset of our numerical results is listed. The first two columns distinguish the three different $O(n)$ nonlinear σ models that were simulated on L^2 lattices. The exactly known critical $\beta_c = \log(1+\sqrt{2})/2$ was chosen for the Ising model. For the x-y model the value $\beta=1.12$ is expected to be very close to the Kosterlitz-Thouless point[11,12] or possibly somewhat beyond in the critical spin-wave phase.[13] The lower values $\beta=1.04$ and 1.07 were simulated to check if a possible finite-size shift of the transition region leads to rising autocorrelation times there. For $\beta=1.5$, 1.6, and $n=3$ we expect (and confirm) a spatial correlation length around 10 and 20.[14] The numbers of generated clusters c follow together with their average size. To diagnose the algorithm we quote the magnetic susceptibility χ and the autocorrelation time τ_χ manifested in the estimates for this observable. It is derived as follows: Measurements of physical observables are separated by a fixed number of m cluster update steps (in the quoted results, $m=1$ for $n=1,2$ and $m=10$ for $n=3$). In equilibrium the connected autocorrelation function in time is

TABLE I. Results for the magnetic susceptibilities χ and autocorrelation times τ_χ (in units comparable to sweeps) for simulations of $O(n)$ σ models on L^2 lattices. In each run a total of c update steps have been performed involving clusters of an average of $\langle |c| \rangle$ spins. The effective autocorrelation time τ_χ^{eff} is directly relevant for error estimation.

n	L	β	$c\times10^{-6}$	$\langle \lvert c \rvert \rangle L^{-2}$	χL^{-2}	τ_χ	τ_χ^{eff}
1	32	0.4406...	0.50	0.4602(7)	0.4598(7)	2.3(3)	1.4
1	64	0.4406...	0.25	0.3858(11)	0.3852(10)	2.3(3)	1.9
1	128	0.4406...	0.20	0.3225(11)	0.3229(10)	2.7(5)	1.8
2	32	1.12	0.62	0.3582(5)	0.4420(2)	3.6(7)	1.4(2)
2	64	1.12	0.26	0.3043(6)	0.3754(3)	2.4(6)	1.3(2)
2	128	1.12	0.20	0.2582(7)	0.3190(3)	2.1(6)	1.2(2)
2	32	1.07	0.26	0.3247(7)	0.3985(4)		1.5(3)
2	64	1.07	0.13	0.2629(9)	0.3245(5)		1.1(2)
2	128	1.07	0.10	0.2114(9)	0.2608(5)		1.2(2)
2	128	1.04	0.26	0.1638(5)	0.2032(5)		1.4(2)
				$\langle \lvert c \rvert \rangle$	χ		
3	128	1.5	1.54	132.0(4)	174.4(0.9)		0.25(4)
3	128	1.6	2.00	334.5(7)	444.9(1.4)	1.0(6)	0.40(6)

measured from successive estimates for χ (and other quantities). An autocorrelation time $\bar{\tau}_\chi$ is determined from its exponential falloff in time separation. All entries in the table are based on our seeing constant stable logarithmic ratios for successive values of the correlation function at least in a window from $\bar{\tau}_\chi$ to $3\bar{\tau}_\chi$. Unless the statistics is increased enormously, noise then takes over. The errors on $\bar{\tau}_\chi$ are estimated subjectively from its fluctuations in the stability window, and from some repeated experiments. The measured correlation is also summed and a tail, approximated as an exponential (using $\bar{\tau}_\chi$), added to incorporate autocorrelations into error estimates. Finally, τ_χ is related to $\bar{\tau}_\chi$ as

$$\tau_\chi = \bar{\tau}_\chi m\langle |c| \rangle L^{-2}. \tag{9}$$

This is a reasonable quantity to compare to the sweeps that are common and natural in local algorithms: $\tau_\chi = 1$ means that on the average one has to "touch" (here, reflect) each spin once to suppress correlations by a factor of $1/e$. Needless to say the values of τ_χ signal an enormously enhanced efficiency for the simulation of spin models. We also estimated errors from the fluctuations of blocks of 1000 measurements each. The two results always agreed within 10%. The effective autocorrelation time is defined as

$$\bar{\tau}^{\text{eff}} = \tfrac{1}{2}(\epsilon_{\text{block}}/\epsilon_{\text{naive}})^2. \tag{10}$$

Here ϵ_{block} and ϵ_{naive} are the representative errors, and an exactly simply exponential autocorrelation with scale $\bar{\tau}^{\text{eff}}$ would produce the same errors. Where errors for τ^{eff} are quoted, they were estimated by a further division of the data into subsamples. For extreme short-range quantities like the nearest-neighbor correlation and $n = 2,3$, we saw somewhat longer correlation times up to $\tau^{\text{eff}} \simeq 4$ but flat in L. These observables, however, are physically less interesting, and one could always intersperse local update steps if the *short*-wavelength modes evolve too slowly. In Fig. 1 we see the size distribution of clusters for 128^2 runs. While both critical models look similar, the finite physical correlation length in O(3) is clearly visible as a sharp cutoff.

Some reference runs with a local random site updating heat-bath algorithm for $Z(16) \simeq O(2)$ at $\beta = 1.12$ yielded $\tau_\chi \simeq 20(4)$ at $L = 16^2$ and indicated the expected growth $\propto L^2$. Cluster runs were also made for Z(16) (with reflections suitably restricted) and produced results indistinguishable from O(2). The difference in speed is only small for the new algorithm, and there is consequently no need for the discrete spin approximation. The equality (8) of χ and $\langle |c| \rangle$ is clearly borne out by the data for $n = 1$. For $n = 2,3$ the ratios $\langle |c| \rangle/\chi$ are also very nearly β-independent constants in the range that we investigate. The only remarks on the physics of the models we wish to make here is that the growth of χ at criticality is compatible with the expected scaling behavior

$$\chi \propto L^{2-\eta}, \tag{11}$$

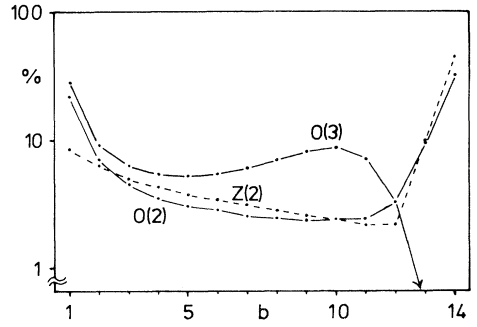

FIG. 1. Distribution of the number of spins updated collectively in $L = 128$ simulations at $\beta = 0.44\ldots$, 1.12, and 1.6 for $n = 1$, 2, and 3, respectively. Bin b comprises clusters of size $2^{b-1} \le c < 2^b - 1$.

and $\eta = 0.25$ in the Ising model. For the x-y case at $\beta = 1.12$ our results give $\eta = 0.235(1)$. We conclude by mentioning that now an accurate and detailed study of the Kosterlitz-Thouless picture and asymptotic freedom scaling behavior for the $n = 2$ and $n \ge 3$ σ models, respectively, which have both recently been challenged,[15] should be feasible.[17] The new algorithm should also be useful in the simulation of four-dimensional O(4) models to study Higgs physics and triviality. The generalization to spin-glasses or neural networks, where β is bond dependent, seems straightforward. Attempts for extensions to lattice gauge theory are under way.

All numerical calculations were carried out on the Cray-XMP/18 at Kiel University. I wish to thank Alan Sokal for exchange, Istvan Montvay for discussions, and the DESY theory group for their hospitality.

[1]R. H. Swendsen and J. S. Wang, Phys. Rev. Lett. **58**, 86 (1987).

[2]For a review, see, e.g., D. Stauffer, Phys. Rep. **54**, 1 (1979).

[3]M. Sweeny, Phys. Rev. B **27**, 4445 (1983).

[4]I. Montvay, G. Münster, and U. Wolff, Nucl. Phys. **B305 [FS23]**, 143 (1988); K. Jansen, J. Jersak, T. Trappenberg, I. Montvay, G. Münster, and U. Wolff, Phys. Lett. B **213**, 203 (1988).

[5]D. Kandel, E. Domany, D. Ron, A. Brandt, and E. Loh, Jr., Phys. Rev. Lett. **60**, 1591 (1988).

[6]U. Wolff, Phys. Rev. Lett. **60**, 1461 (1988), and Nucl. Phys. **B300 [FS22]**, 501 (1988).

[7]R. G. Edwards and A. D. Sokal, Phys. Rev. D **38**, 2009 (1988).

[8]F. Niedermayer, Phys. Rev. Lett. **61**, 2026 (1988). This paper reached me during numerical tests with practically the same algorithm. My results were negative: either the whole lattice ends in one cluster, or large nontrivial clusters rarely move because of their surface energy. This frustation led to the present algorithm, where a cluster, once created, is always flipped.

[9]Note that for the x-y model written with U(1) spins (phases), this is an antilinear operation.

[10]This connection with SW clusters can also be employed to prove (6).

[11]J. M. Kosterlitz and D. J. Thouless, J. Phys. C **6**, 1181 (1973); J. M. Kosterlitz, J. Phys. C **7**, 1046 (1974).

[12]J. Tobochnik and G. V. Chester, Phys. Rev. B **20**, 3761 (1979).

[13]W. Bernreuther and M. Göckeler, Phys. Lett. B **214**, 109 (1988).

[14]I. Bender, W. Wetzel, and B. Berg, Nucl. Phys. **B269**, 389 (1986).

[15]A. Patrascioiu, E. Seiler, and I. O. Stamatescu, Max-Planck-Institut, München, Report No. MPI-PAE/PTh 75/87 (to be published); A. Patrascioiu, E. Seiler, I. O. Stamatescu, and V. Linke, Max-Planck-Institut, München, Report No. MPI-PAE/PTh 76/87 (to be published).

[16]R. Brower, in Lattice '88, Proceedings of the 1988 Symposium on Lattice Field Theory, Fermilab, 22–25 September 1988, edited by W. A. Bardeen *et al.* (North-Holland, Amsterdam, to be published).

[17]U. Wolff, to be published.

PHYSICAL REVIEW

LETTERS

VOLUME 60 | 18 APRIL 1988 | NUMBER 16

Simulations without Critical Slowing Down

Daniel Kandel and Eytan Domany
Department of Electronics, Weizmann Institute of Science, Rehovot 76100, Israel

Dorit Ron and Achi Brandt
Department of Applied Mathematics, Weizmann Institute of Science, Rehovot 76100, Israel

and

Eugene Loh, Jr.
Theoretical Division and Center for Non-Linear Studies, Los Alamos National Laboratory, Los Alamos, New Mexico 87545
(Received 4 February 1988)

We have developed a novel simulation method that combines a multigrid technique with a stochastic blocking procedure. Our algorithm eliminates critical slowing down completely, as demonstrated by simulations of the two-dimensional Ising model at criticality.

PACS numbers: 05.50.+q, 64.60.Ht, 75.40.Mg

Considerable effort has been devoted in recent years to simulations of various problems, that range from calculating hadron masses,[1] through equilibrium phase transitions,[2] to a variety of nonequilibrium, complex time-dependent phenomena.[2,3] Of the many problems that hinder large-scale simulations we address that of *critical slowing down* (CSD). This phenomenon gives rise to a divergent relaxation time τ as the critical point is approached.[4] Thus, the times needed to equilibrate the system and to generate statistically independent configurations (at equilibrium) become exceedingly large. At criticality, τ grows as

$$\tau \sim L^z, \tag{1}$$

with the linear size L of the system. Here τ is measured in units that scale with the total number of sites (i.e., L^d), and z is the dynamic critical exponent.

In this Letter we present a method that overcomes this difficulty. While a number of recently reported simulation techniques[5,6] have partially eliminated CSD, this is the first simulation that evades CSD completely for a nontrivial model. We illustrate the method by simulating the $d=2$ Ising model at the critical temperature of the infinite lattice. The equilibration time for the energy is found to be constant, $\tau \approx 3.0$, for lattices of linear sizes $8 \leq L \leq 128$ (see Fig. 1). The method used is a combination of a *multigrid* method[8-10] with a *stochastic blocking* (coarsening) technique.[6,11,12] General features of our method (applicable to discrete or continuous-state

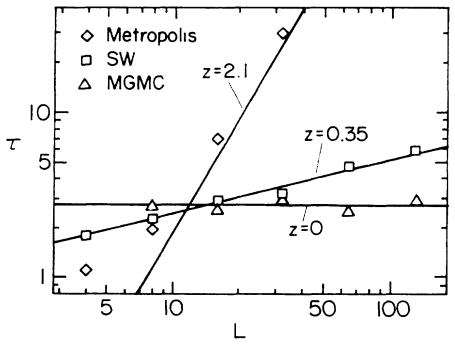

FIG. 1. Relaxation time τ vs (linear) system size L. Three methods are compared: Metropolis algorithm (Ref. 7), Swendsen and Wang's method (Ref. 6) (SW), and the multigrid Monte Carlo technique (MGMC).

Hamiltonians, and a wide class of coarse-to-fine interpolations) were first presented recently.[8]

The structure of this Letter is as follows: First, we provide an elaborate description of our algorithm and show that it satisfies detailed balance. Details of the simulations mentioned above are given next. Finally, we explain why our procedure does eliminate CSD completely, whereas the Swendsen-Wang method[6] (SW), which uses stochastic blocking without incorporating multigrid ideas, does not.

Description of the algorithm.—In our coarsening procedure, simulation of the full Hamiltonian is replaced by simulation of a stochastically generated, simplified Hamiltonian, over a restricted space with fewer degrees of freedom. Write $\mathcal{H}=\mathcal{H}_0+V$, where factors of $1/k_BT$ have been absorbed into $\mathcal{H}$ and where $\mathcal{H}_0$ is somehow easier to simulate than the original $\mathcal{H}$. Assume that the system is in some state Q. We "kill" the interaction V either by "deleting" it with probability $p_d=C_V\times\exp[V(Q)]$ or by "freezing" it with probability $p_f=1-p_d$. If the interaction is frozen, only states Q' with $V(Q')=V(Q)$ are considered in the ensuing simulation. If the interaction is deleted, no such restriction is placed on the states. In either case, the thermodynamics is subsequently governed only by the simplified Hamiltonian $\mathcal{H}_0$. One must choose C_V so that $p_d,p_f\in[0,1]$. The largest choice of C_V produces the best statistics.

In practice, $\mathcal{H}_0$ is still nontrivial to simulate (without CSD) and so additional terms of the Hamiltonian must be killed. After killing all the interactions in $\mathcal{H}$, one arrives at a system which is completely decoupled—and so is trivial to simulate—but is subject to an arbitrary set of restrictions on its states.

For the two-dimensional ferromagnetic Ising model $\mathcal{H}=-\sum_{\langle ij\rangle}K_{ij}s_is_j$, we kill the interactions $K_{ij}s_is_j$ one at a time. The optimal probability for deletion, then, is $p_d=\exp[-K_{ij}(1+s_is_j)]$. Interactions between antiparallel spins will always be deleted; only parallel spins can be frozen together. We may kill the interactions, generate a new configuration for the decoupled system, and then return to the full Hamiltonian to continue the simulation. This is the procedure followed by SW; it still suffers from CSD, although with a considerably reduced dynamical exponent.

To eliminate CSD completely, we combine this stochastic blocking procedure with multigrid ideas. As an example, consider Fig. 2(a). Choose the boxed sites as our coarse lattice, which has $(L/b)^2$ spins; here the length rescaling factor is $b=2$. We coarsen by considering all pairs of coupled sites i,j. If *both* i and j have already been frozen to coarse spins, we move on to the next pair; otherwise kill $K_{ij}s_is_j$. Figure 2(a) is a possible coarsening. Single (double) lines between sites indicate living (frozen) bonds; deleted bonds are not marked. Each cluster of fine spins connected by frozen bonds constitutes an irregular, stochastically generated block. A

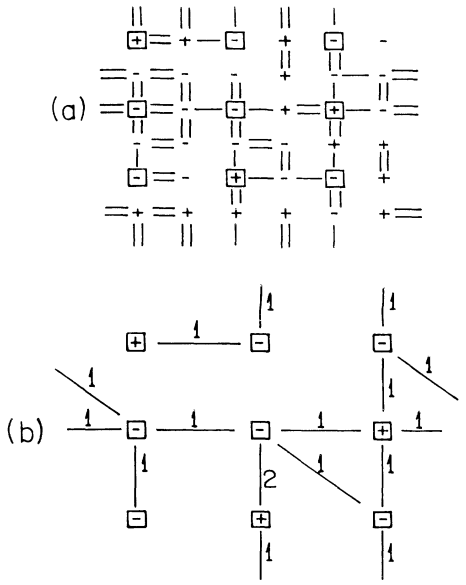

FIG. 2. Coarsening of a fine-spin configuration. Boxed spins constitute the coarse lattice. (a) Outcome of a possible coarsening. Single (double) lines denote living (frozen) bonds; deleted bonds are not marked. (b) Coarse lattice. Numerals near bonds denote the number of fine couplings added together to form the associated coarse bond.

block is either frozen to (and viewed as) a single coarse spin or it is completely decoupled.

Finally, the coarse Hamiltonian is constructed: The coupling between coarse spins s_i and s_j is the sum of the living couplings that connect fine spins associated with coarse spins s_i and s_j. This is illustrated in Fig. 2(b): Numerals on the lines between sites indicate the number of fine couplings that contribute to the coarse bond. The new Hamiltonian is inhomogeneous, and may contain long-range interactions.

In addition to coarsening the system, we need "uncoarsening" and Metropolis procedures as well. To "uncoarsen" the system, decoupled blocks are set to some arbitrary value of spin, all fine-lattice spins take the value of the block spin to which they were frozen, and the fine-lattice couplings are restored. Metropolis updates may be performed at any length scale by use of a standard Metropolis algorithm[13] on the block spins at that scale, with the corresponding Hamiltonian.

Our dynamic procedure "cycles" through all length scales, starting from the finest. At each intermediate length scale, the system is coarsened γ times before it is uncoarsened (see Fig. 3). At the coarsest level all blocks are decoupled, and each time this level is reached, it is immediately uncoarsened. The cycle ends when the finest level is reached. A few Metropolis sweeps are performed at each level. These sweeps, however, are not essential to elimination of CSD. The SW procedure[6]

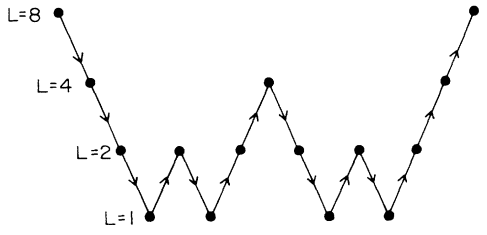

FIG. 3. An example of a cycle on an 8×8 lattice, with length rescaling factor $b=2$ and $\gamma=2$.

corresponds to $\gamma=1$ with $b=L$, going directly from the finest to the coarsest level and back, without Metropolis sweeps at the finest level.

Proof of detailed balance.—Clearly, our procedure is strongly ergodic since there is always a nonzero probability that no restriction will be placed on the simulation, allowing nonzero transition probabilities between all states. Consider a transition generated by coarsenings and a Metropolis step, which satisfies detailed balance with respect to the coarse Hamiltonian. It can be viewed as taking place between the two corresponding fine-spin states, and we want to show that detailed balance with respect to the fine-spin Hamiltonian is satisfied. Once proven, this implies that the transitions satisfy detailed balance with respect to the *finest* Hamiltonian.

Start from some level in state Q with a Hamiltonian $\mathcal{H}$, and "kill" an interaction V to get the coarse Hamiltonian $\mathcal{H}_0=\mathcal{H}-V$. Then a transition from Q to another state Q', with $V(Q)\neq V(Q')$, can take place only if V has been deleted:

$$T(Q\to Q')=C_V e^{V(Q)}T_0(Q\to Q'), \tag{2}$$

where T_0 are transition probabilities with $\mathcal{H}_0$. Hence,

$$\frac{T(Q\to Q')}{T(Q'\to Q)}=\frac{e^{V(Q)}e^{-\mathcal{H}_0(Q')}}{e^{V(Q')}e^{-\mathcal{H}_0(Q)}}=\frac{e^{-\mathcal{H}(Q')}}{e^{-\mathcal{H}(Q)}}. \tag{3}$$

Alternatively, if $V(Q)=V(Q')$, the interaction V may either be deleted for frozen:

$$\begin{aligned}T(Q\to Q')&=C_V e^{V(Q)}T_0(Q\to Q')+(1-C_V e^{V(Q)})T_0^f(Q\to Q')\\&=[C_V e^{V(Q)}T_0(Q'\to Q)+(1-C_V e^{V(Q)})T_0^f(Q'\to Q)]e^{-\mathcal{H}_0(Q')}/e^{-\mathcal{H}_0(Q)},\end{aligned} \tag{4}$$

where T_0^f are transition probabilities with the reduced Hamiltonian over the restricted space. Remembering that $V(Q')=V(Q)$ yields

$$T(Q\to Q')=T(Q'\to Q)e^{-\mathcal{H}(Q')}/e^{-\mathcal{H}(Q)}, \tag{5}$$

completing the proof of detailed balance.

Details of simulations.—We simulated the $d=2$ Ising model on square lattices of linear size $8\le L\le 128$ with periodic boundary conditions. A cycle of $\gamma=2$ with rescaling factor $b=2$ was used. At each level, one Monte Carlo sweep was performed. Starting from a fully magnetized state we measured the decay of the energy to its equilibrium value,[14] averaged over an ensemble of up to 100 systems. From such data we extracted the relaxation times $[\tau(L)]$, and found $\tau\approx 3.0$ *for all lattice sizes.* Figure 1 is a log-log plot of $\tau(L)$ measured with three different algorithms: standard Metropolis procedure[7] ($z\approx 2.1$), the SW algorithm[6] ($z\approx 0.35$), and our method ($z=0$).

Why does it work?—To eliminate CSD, one must allow flips of large domains with high acceptance ratios. Had our coarsening procedure created lattices with higher connectivity or stronger bonds than those of the fine lattices, the acceptance ratios for the large-scale moves would have become prohibitively small. The production of lower connectivities or weaker couplings would have generated decoupled blocks at short length scales, and large-scale flips would not have been possible. Hence, since our algorithm does produce $z=0$, we expected and indeed confirmed that our coarsening procedure yields similar distributions of bonds and connectivities in lattices at different length scales.

The SW algorithm[6] does not eliminate CSD completely. The reason for this is that coarsening introduces spin-spin correlations on an average length scale of b (the rescaling factor of the coarsening transformation) by freezing spins on that scale. Equilibration of the coarse lattice has no effect on these correlations. One has to repeat the process of coarsening, equilibration, and uncoarsening several times, in order to decorrelate and equilibrate the fine lattice. Note, however, that in each such step equilibration involves further coarsening, for which the same considerations hold. Hence a nested sequence of coarsening-equilibration-uncoarsening steps is needed. Denote by $\tau_0(b)$ the number of coarsening-equilibration-uncoarsening steps needed to eliminate correlations on length scale b. If we assume scale invariance at criticality, i.e., that $\tau_0(b)$ is the same for all levels in the hierarchy,[15] the equilibration process described above is precisely the cycle shown in Fig. 3, with $\gamma=\tau_0$. Therefore, if such a cycle is used, the configuration of the fine lattice becomes independent of the initial configuration after τ_0 cycles. To estimate τ_0, recall that the SW procedure is one coarsening step with $b=L$, yielding $\tau_0(L)\sim L^{z_{\rm sw}}$. Combining scale invariance with the SW result, we get $\tau_0(b)\approx b^{z_{\rm sw}}$. From the same reasoning it follows that, if one uses a cycle with $\gamma<\tau_0$, the number of cycles needed to decorrelate diverges as $L^{z_{\rm sw}-\log_b\gamma}$. Hence, in order to equilibrate in a finite number of cycles, γ must satisfy the inequality $\gamma\ge b^{z_{\rm sw}}$. Since the time needed to complete a cycle (measured in

steps per site) diverges (as $L \to \infty$) if $\gamma > b^d$, CSD is eliminated completely only if $b^{z_{sw}} \leq \gamma < b^d$. For the Ising model, this condition translates to $b^{0.35} \leq \gamma < b^2$ for $d=2$, and $b^{0.75} \leq \gamma < b^3$ for $d=3$. For the three-state Potts model in two dimensions we must satisfy $b^{0.6} \leq \gamma < b^2$. For values of b and γ that do not satisfy these bounds, we expect CSD. In particular, for the three-state Potts model in $d=2$, with $\gamma=2$ and $b=4$, we expect to get $\tau \sim L^{z_{sw} - \log_b \gamma} \approx L^{0.1}$.

A more complete account of this work as well as detailed tests of the scaling arguments given above will be reported elsewhere.[16] We are also working on optimal coarsening procedures that reduce statistical fluctuations and are starting simulations on the 2D XY model.

One of us (D.K.) thanks R. Benav for stimulating discussions. This research was supported by the U.S. Air Force Office of Scientific Research, under Grant No. AFOSR-86-0127 (A.B.), by the U.S.-Israel Binational Science Foundation (E.D.), and by the U.S. Department of Energy (E.L.).

[1]I. Montvay, Rev. Mod. Phys. **59**, 263 (1987).

[2]See *Applications of Monte Carlo Methods in Statistical Physics,* edited by K. Binder (Springer-Verlag, Berlin, 1987), 2nd ed.

[3]P. S. Sahni, G. S. Grest, and S. A. Safran, Phys. Rev. Lett. **50**, 60 (1983); D. A. Huse, Phys. Rev. B **34**, 7845 (1986); E. Kremer and K. Binder, J. Chem. Phys. **81**, 6381 (1984).

[4]See, for example, P. C. Hohenberg and B. I. Halperin, Rev. Mod. Phys. **49**, 435 (1977).

[5]K. E. Schmidt, Phys. Rev. Lett. **51**, 2175 (1983); J. Goodman and A. Sokal, Phys. Rev. Lett. **56**, 1015 (1986); E. Dagotto and J. Kogut, Phys. Rev. Lett. **58**, 299 (1987); R. G. Edwards and A. D. Sokal, unpublished.

[6]R. H. Swendsen and J. S. Wang, Phys. Rev. Lett. **58**, 86 (1987).

[7]J. K. Williams, J. Phys. A **18**, 49 (1985).

[8]A. Brandt, in Preliminary Proceedings of the Third Copper Mountain Conference on Multigrid Methods, Copper Mountain, Colorado, 1987 (unpublished), and in Proceedings of the Third Copper Mountain Conference on Multigrid Methods, Copper Mountain, Colorado, 1987, edited by S. McCormick (Marcel Dekker, New York, to be published).

[9]A. Brandt, D. Ron, and D. J. Amit, in *Multigrid Methods II; Proceedings, Cologne,* edited by W. Hackbusch, Lecture Notes in Mathematics, Vol. 1228 (Springer-Verlag, Berlin, 1985), p. 66.

[10]W. L. Briggs, *A Multigrid Tutorial* (Society for Industrial and Applied Mathematics, Philadelphia, PA, 1987).

[11]P. W. Kasteleyn and C. M. Fortuin, J. Phys. Soc. Jpn. Suppl. **26s**, 11 (1969); C. M. Fortuin and P. W. Kasteleyn, Physica (Utrecht) **57**, 536 (1972).

[12]M. Sweeny, Phys. Rev. B **27**, 4445 (1983).

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

[14]A. E. Ferdinand and M. E. Fisher, Phys. Rev. **185**, 832 (1969).

[15]This assumption is reasonable since we found that lattices of different levels have the same properties. The effect of performing intermediate Monte Carlo steps on τ_0 has not yet been studied in detail.

[16]D. Kandel *et al.*, unpublished.

Multicanonical Ensemble: A New Approach to Simulate First-Order Phase Transitions

Bernd A. Berg[(1),(2),(a)] and Thomas Neuhaus [(1)]

[(1)]*Fakultät für Physik, Universität Bielefeld, D-4800 Bielefeld, Federal Republic of Germany*

[(2)]*Supercomputer Computations Research Institute, Tallahassee, Florida 32306*

(Received 19 July 1991)

Relying on the recently proposed multicanonical algorithm, we present a numerical simulation of the first-order phase transition in the 2D 10-state Potts model on lattices up to sizes 100×100. It is demonstrated that the new algorithm *lacks* an exponentially fast increase of the tunneling time between metastable states as a function of the linear size L of the system. Instead, the tunneling time diverges approximately proportional to $L^{2.65}$. On our largest lattice we gain more than 2 orders of magnitude as compared to a standard heat-bath algorithm. As a first physical application we report a high-precision computation of the interfacial free energy per unit area.

PACS numbers: 05.50.+q, 11.15.Ha, 64.60.Fr, 75.10.Hk

Critical slowing down is of crucial importance to computer simulations of phase transitions. For second-order phase transitions long autocorrelation times at criticality cause severe restrictions on the maximum lattice size for which one can obtain good statistics of thermodynamic quantities. For a number of spin systems this critical slowing down was overcome by the nonlocal cluster algorithm of Swendsen and Wang [1]; for a recent review see [2]. However, for first-order transition one encounters an even worse and different problem of critical slowing down. The interfacial free energy between disordered and ordered states has a finite value on the critical point for the infinite volume system. Configurations dominated by the presence of the interface will be exponentially suppressed by the Boltzmann factor in the canonical ensemble. On finite lattices this leads then to an exponentially fast suppression of the tunneling between metastable states of the system with increasing lattice size. To overcome this critical slowing down effect for first-order transitions, we recently proposed a *multicanonical* Monte Carlo (MC) algorithm [3]. The multicanonical MC algorithm is designed to enhance configurations which are dominated by the presence of the interface and therefore exponentially suppressed. In this way it is possible to avoid the exponentially fast growing slowing down at the first-order phase transition. In this paper we demonstrate this in the case of our example: the 2D 10-state Potts model.

The 2D 10-state Potts model [4] is defined by the partition function

$$Z(\beta)=\sum_{\text{configurations}}\exp(\beta S)\,, \tag{1}$$

$$S=\sum_{\langle i,j\rangle}\delta_{q_i,q_j}\,, \tag{2}$$

$$q_i,q_j=0,\ldots,9\,. \tag{3}$$

Recently there has been renewed interest in this model [5–7]. It serves as an excellent laboratory for finite-size-scaling (FSS) studies of temperature-driven strong first-order phase transitions. To calculate the interfacial free energy F^s between the disordered and the (ten) ordered states has remained the hardest problem. The reason is the pronounced double-peak structure of the sampled action density $P_L(S)$ in the canonical ensemble near the critical point. The pseudocritical point β_L^t is defined such that both maxima are of equal height:

$$P_L^{1,\max}=P_L(S_L^{1,\max})=P_L(S_L^{2,\max})=P_L^{2,\max}\,. \tag{4}$$

In addition we have imposed the normalization condition $1=P_L^{1,\max}=P_L^{2,\max}$. Figure 1 depicts the action densities for lattices with $L=16$, 24, 34, 50, 70, and 100 on a logarithmic scale and we see that 4 orders of magnitude are involved: $P_L^{\min}/P_L^{\max}\simeq5.1\times10^{-5}$ for $L=100$. With our conventions for $P_L(S)$ the interfacial free energy per unit area $F^s=F^s_\infty$ can now be defined [8] as the $L\to\infty$ limit of the quantity

$$F_L^s=-\frac{1}{L}\ln P_L^{\min}\,. \tag{5}$$

For a numerical calculation of F_L^s it is now clear that any algorithm which samples configurations with a probability $\sim P_L(S)$ would slow down proportional to $1/P_L^{\min}$. As for large lattices $P_L^{\min}\sim\exp(-F^sL^{d-1})$, it is expected that an appropriately defined tunneling time τ_L^t will be-

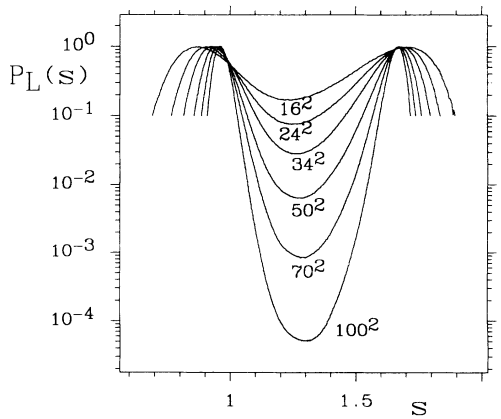

FIG. 1. Action density distributions $P_L(s)$ for lattices of size $L=16$, 24, 34, 50, 70, and 100 on a logarithmic scale. The values of the maxima have been normalized to 1.

have as

$$\tau_L^f = A_\tau L^a e^{+F^s L^{d-1}}. \quad (6)$$

The parameters A_τ and a can, in principle, be determined by a fit to the measured tunneling times.

The multicanonical MC algorithm samples configurations with the weight

$$\mathcal{P}_L^{\mathrm{MC}}(S) \sim e^{(\alpha_L^k + \beta_L^k S)} \text{ for } S_L^k < S \le S_L^{k+1} \quad (7)$$

instead of sampling with the usual Boltzmann factor $P_L^B(S) \sim \exp(\beta_L^c S)$ corresponding to the canonical ensemble. Here we partitioned the total action interval $0 \le S \le 2V$ into $k = 0, \ldots, N$ ($N+1$ odd) intervals $I_k = (S_L^k, S_L^{k+1}]$. The idea of the multicanonical MC algorithm is to choose intervals I_k and values of β_L^k and α_L^k at the pseudocritical point β_L^c in such a way that the resulting multicanonical action density $\mathcal{P}_L(S)$ has an approximately flat behavior for values of the action in the interval $[S_L^{1,\max}, S_L^{2,\max}]$; that is to say, configurations dominated by the interface are no longer exponentially suppressed as they are in the metastable-unstable region of the canonical ensemble. Physically this can be achieved by choosing the β parameters β_L^k such that the system gets heated when it is in the ordered state of the metastable region, cooled when it is in the disordered state, and neither of these if it is in the unstable region. The parameters β_L^k hereby take the form $\beta_L^k = \beta_L^c + \delta\beta_L^k$, where the coupling constant difference $\delta\beta_L^k$ changes sign as a function of S and is responsible for the altered dynamics of the model. The parameters α_L^k are adjusted in such a way that $\mathcal{P}_L(S)$ is a steady function of S.

In [3] we demonstrated, that when the double-peak distribution $P_L(S)$ can be approximated by a double Gaussian, the multicanonical action density $\mathcal{P}_L(S)$ can be made arbitrarily flat by driving a control parameter $r > 1$ towards 1. In this case we choose action values S_L^k with $S_L^0 = 0$, $S_L^{N+1} = 2V$, $S_L^1 = S_L^{1,\max}$, $S_L^N = S_L^{2,\max}$ and in the interval $[S_L^{1,\max}, S_L^{\min})$ action values defined by the equation $P_L(S_L^k) = r^{1-k} S_L^{1,\max}$ for $k = 1, \ldots, N/2$. An analog procedure is adopted in the interval $(S_L^{\min}, S_L^{2,\max}]$. Having defined the action values S_k and corresponding intervals I_k the setting

$$\beta_L^k = \begin{cases} \beta_L^c \text{ for } k = 0, N/2, N, \\ \beta_L^c + \ln(r)/(S_L^{k+1} - S_L^k) \text{ for } k = 1, \ldots, N/2 - 1, \\ \beta_L^c - \ln(r)/(S_L^{k+1} - S_L^k) \text{ for } k = N/2 + 1, \ldots, N - 1, \end{cases} \quad (8)$$

and the recursion

$$\alpha_L^{k+1} = \alpha_L^k + (\beta_L^k - \beta_L^{k+1}) S_L^{k+1}, \quad \alpha_L^0 = 0 \quad (9)$$

defines the multicanonical ensemble. In accordance with detailed balance, standard Metropolis and heat-bath updating algorithms have been generalized to the multicanonical situation [3,9]. Finally we obtain the canonical action density distribution $P_L(S)$ through a reweighting step similar to [10,11] from the multicanonical distribution $\mathcal{P}_L(S)$:

$$P_L(S) = e^{(\beta_L^c - \beta_L^k S - \alpha_L^k)} \mathcal{P}_L(S) \text{ for } S_L^k < S \le S_L^{k+1}. \quad (10)$$

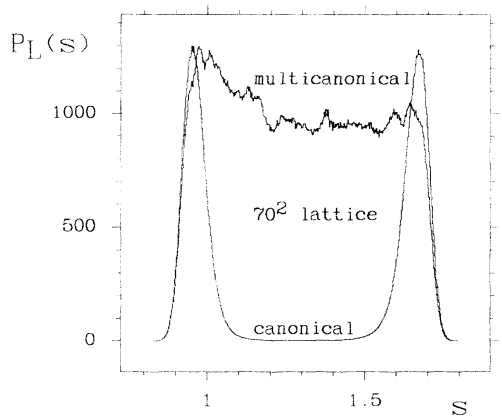

FIG. 2. Multicanonical action density distribution $\mathcal{P}_{70}(s)$ together with its reweighted distribution $P_{70}(s)$.

As an example we show for our $L = 70$ system in Fig. 2 the multicanonical action density distribution $\mathcal{P}_{70}(S)$ together with the reweighted distribution $P_{70}(S)$. In practice the appropriate choice of the parameters in Eq. (7) is obtained by making from the given systems a FSS prediction of the density distribution $P_L(S)$ for the next larger system. A second run may then be performed with optimized parameters. It is our experience that the guess works normally so well that the second run is barely an improvement as compared with the first. On the smallest systems standard MC simulation provides initial data. Our statistics for this investigation was 4×10^6 heat-bath sweeps per run and lattice size. One sweep updates each spin of the lattice once.

We define the tunneling time τ_L^f as the average number of sweeps needed to get from a configuration with action $S = S_L^{1,\max}$ to a configuration with $S = S_L^{2,\max}$ *and* back. With our statistics of 4×10^6 sweeps per run the system tunnels then in total $8 \times 10^6/\tau_L^f$ times. Table I collects

TABLE I. The tunneling times τ_L^f as a function of the lattice size L for the multicanonical MC algorithm (second row) and the heat-bath algorithm (third row). For some lattice sizes we display the results of several simulations, whose difference lies in slightly different coupling parameters.

L	τ_L^f (multicanonical)	τ_L^f (heat bath)
12	542(4)	793(7)
12	...	776(9)
16	1147(10)	1988(23)
24	3354(57)	9634(408)
34	8375(245)	43923(3151)
50	23763(1321)	270565(63222)
50	24932(1064)	...
70	69492(6383)	...
70	62218(5560)	...
100	160334(16252)	...

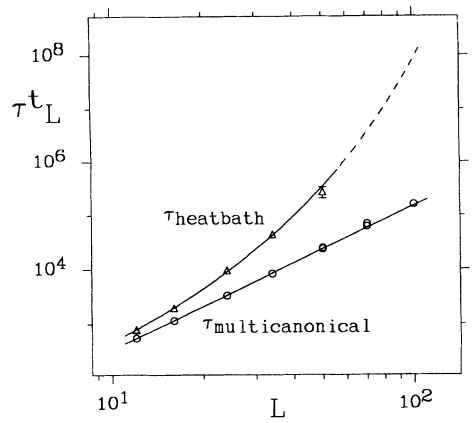

FIG. 3. Tunneling times for the multicanonical MC algorithm and the heat-bath algorithm in a double-log scale. The curves correspond to the fits in Eqs. (11) and (12). The dashed part of the curve indicates the extrapolation to the $L=100$ lattice for the heat-bath algorithm. On the 100 lattice the system still tunnels 50 times between the metastable states during 4×10^6 sweeps, when the multicanonical simulation is used.

the measured tunneling times. For our smaller systems we have also carried out standard heat-bath MC runs at β^c_L and the associated tunneling times are also reported in Table I. For the larger systems standard MC runs would not tunnel often enough to allow for a reliable direct calculation of their tunneling times. This is of course due to the exponential slowing down of the standard MC simulation. In Fig. 3 we display on a log-to-log scale the divergence of the tunneling times τ^t_L for the multicanonical MC algorithm (circles) and the heat-bath algorithm (triangles). There is clearly a different behavior of the two algorithms involved. While for the multicanonical MC algorithm the increase of the tunneling time is consistent with a power law, the heat-bath algorithm displays an exponentially fast growing tunneling time. Performing a χ^2 fit we obtain the following fits:

$$\tau^t_L(\text{multicanonical}) = 0.73(3) L^{2.65(2)} \quad (11)$$

with $\chi^2/n_{\text{d.o.f.}} = 1.09$,

$$\tau^t_L(\text{heat bath}) = 1.46 L^{2.15} e^{+0.080L} \quad (12)$$

with $\chi^2/n_{\text{d.o.f.}} = 2.2$.

The quality of the fits as indicated by the χ^2 values ($n_{\text{d.o.f.}}$ denotes the number of degrees of freedom) is reasonable. In case of the heat-bath algorithm we could not reliably determine the errors from a three-parameter fit. The ratio $R = \tau^t_L(\text{heat bath})/\tau^t_L(\text{multicanonical})$ is a direct measure for the relative efficiency of the two algorithms. Using the fits we extrapolate its value to the $L=100$ system and estimate a factor $R \approx 500$ for this case. The multicanonical algorithm approximately slows down like $\sim V^{2.325}$ with respect to the number of updates per degree of freedom. This is only slightly worse than the optimal performance $\sim V^2$ which was estimated in [3] based on a random-walk picture. For the heat-bath algorithm the inefficiency of the algorithm prohibits a very accurate estimate of F^s from the behavior of the tunneling time according to Eq. (6). The fitted value in Eq. (11) is, however, close to the determination of the next paragraph [Eq. (14)].

Our multicanonical data allow the so far most precise determination of the interfacial free energy per unit area for the 2D 10-state Potts model. For this purpose we determine maxima and minima of the $P_L(S)$ distributions by self-consistent straight line fits over suitable S ranges. Together with the central values of their associated ranges, our F^s_L values are collected in Table II. Performing the FSS fit according to [8]

$$F^s_L = F^s + c/L, \quad (13)$$

we obtain consistent results for lattices of size $L=16$, 24, 34, 50, 70, and 100, as displayed in Fig. 4 (we have $\chi^2/n_{\text{d.o.f.}} = 0.54$). We estimate the infinite volume value of the interfacial free energy per unit area to be

$$F^s = 0.09781 \pm 0.00075. \quad (14)$$

This value may, however, depend weakly on the analytical form of the FSS fit [8] and even with our large lattices we may still face additional systematic errors of a similar order as the quoted statistical error. Future simulations on even larger lattice sizes might therefore be of interest.

In summary, we have introduced a multicanonical ensemble for the numerical simulation of first-order phase

TABLE II. The pseudocritical couplings β^c_L, locations of the maxima and minima of the action density distribution and interfacial free energies, as determined from the multicanonical distributions.

L	β^c_L	$S^{1,\max}_L$	$S^{\min}_L$	$S^{2,\max}_L$	F^s_L
12	1.40738(09)	116	169	243	0.1071(06)
16	1.41534(12)	216	309	429	0.1086(07)
24	1.42100(08)	523	723	978	0.1058(08)
34	1.42338(09)	1072	1466	1945	0.1039(13)
50	1.42481(06)	2358	3162	4192	0.1027(11)
50	1.42469(06)	2357	3186	4190	0.1006(10)
70	1.42536(06)	4661	6257	8190	0.0983(20)
70	1.42541(05)	4660	6250	8178	0.1007(12)
100	1.42576(04)	9602	13060	16686	0.0986(18)
100	1.42577(04)	9577	13093	16711	0.0994(15)

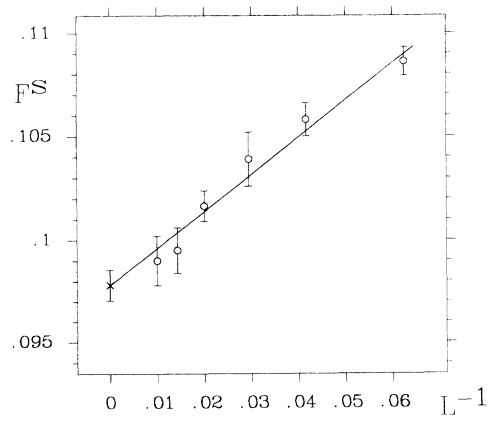

FIG. 4. FSS estimate of the interfacial free energy F^s. Averages are used for those lattices for which we have two data sets.

transitions, which eliminates an exponentially fast increase of the tunneling time between the ordered and disordered states in the critical region of the system. This finding is achieved by replacing the usual equilibrium dynamics of the canonical ensemble, through a new equilibrium dynamics, where the ordered and disordered states of the system get heated and cooled in a well controlled way. Thus configurations dominated by the presence of the interface are enhanced during the simulation.

The multicanonical MC algorithm gives a general framework for the numerical studies of first-order phase transitions in statistical mechanics as well as for field theoretic models. From the numerical point of view the interesting question will be what improvement factors can be achieved as compared with standard algorithms for certain models on certain lattice sizes. We expect the answer to this question to be determined by the value of the quantity $Q = F_L^s \times L^{d-1}$, where the strength of the first-order phase transition is indicated by the magnitude of F_L^s and the d is the dimensionality of the system. In the case of the 2D 10-state Potts model we find at values of $Q \sim 10$ approximately an improvement of 2 to 3 orders of magnitude, while at values of $Q \sim 1$ the improvement is marginal.

An implementation of the multicanonical MC algorithm for non-Abelian gauge theories is straightforward and we think that future investigations of the QCD deconfining phase transition will benefit from this. Beyond first-order phase transition, it may well be that multicanonical algorithms could be of use for other numerical calculations in statistical mechanics, such as estimates of the free energy or spin-glass simulations.

Our simulations were performed on the SCRI cluster of fast RISC work stations and the Convex C240 at the University of Bielefeld. We would like to thank Nelson Alves for collaboration in the early phase of this work. One of us (T.N.) appreciates discussions with A. M. Ferrenberg and D. P. Landau. This research project was partially funded by the National Science Foundation under Grant No. INT-8922411 and by the Department of Energy under Contract No. DE-FG05-87ER40319.

[a]On leave of absence from Department of Physics, The Florida State University, Tallahassee, FL 32306.

[1] R. H. Swendsen and J.-S. Wang, Phys. Rev. Lett. **58**, 86 (1987).

[2] R. H. Swendsen, J.-S. Wang, and A. M. Ferrenberg, "New Monte Carlo Methods for Improved Efficiency of Computer Simulations in Statistical Mechanics" (unpublished).

[3] B. Berg and T. Neuhaus, University of Bielefeld Report No. BI-TH 91/08 (to be published).

[4] R. B. Potts, Proc. Cambridge Philos. Soc. **48**, 106 (1952); F. Y. Wu, Rev. Mod. Phys. **54**, 235 (1982).

[5] A. Billoire, R. Lacaze, A. Morel, S. Gupta, A. Irbäck, and B. Peterson, Nucl. Phys. **B358**, 231 (1991).

[6] J. Lee and J. M. Kosterlitz, Phys. Rev. Lett. **65**, 137 (1990); Phys. Rev. B **43**, 3265 (1991).

[7] C. Borgs, R. Kotecký, and S. Miracle-Solé, J. Stat. Phys. **62**, 529 (1991); C. Borgs and W. Janke, Freie Universität Berlin Report No. HEP 6/91 (to be published).

[8] K. Binder, Phys. Rev. A **25**, 1699 (1982); Z. Phys. B **43**, 119 (1981).

[9] B. Baumann, Nucl. Phys. **B285**, 391 (1987).

[10] A. M. Ferrenberg and R. H. Swendsen, Phys. Rev. Lett. **61**, 2635 (1988); **63**, 1658(E) (1989), and references given therein.

[11] B. Baumann and B. Berg, Phys. Lett. **164B**, 131 (1985).

Efficient, Multiple-Range Random Walk Algorithm to Calculate the Density of States

Fugao Wang and D. P. Landau
Center for Simulational Physics, The University of Georgia, Athens, Georgia 30602
(Received 25 October 2000)

We present a new Monte Carlo algorithm that produces results of high accuracy with reduced simulational effort. Independent random walks are performed (concurrently or serially) in different, restricted ranges of energy, and the resultant density of states is modified continuously to produce locally flat histograms. This method permits us to directly access the free energy and entropy, is independent of temperature, and is efficient for the study of both 1st order and 2nd order phase transitions. It should also be useful for the study of complex systems with a rough energy landscape.

DOI: 10.1103/PhysRevLett.86.2050 PACS numbers: 64.60.Cn, 02.70.Rr, 05.50.+q

Computer simulation has become an essential tool in condensed matter physics [1], particularly for the study of phase transitions and critical phenomena. The workhorse for the past half-century has been the Metropolis importance sampling algorithm, but more recently new, efficient algorithms have begun to play a role in allowing simulation to achieve the resolution which is needed to accurately locate and characterize phase transitions. For example, cluster flip algorithms, beginning with the seminal work of Swendsen and Wang [2], have been used to reduce critical slowing down near 2nd order transitions. Similarly, the multicanonical ensemble method [3] was introduced to overcome the tunneling barrier between coexisting phases at 1st order transitions, and this approach also has utility for systems with a rough energy landscape [4–6]. In both situations, histogram reweighting techniques [7] can be applied in the analysis to increase the amount of information that can be gleaned from simulational data, but the applicability of reweighting is severely limited in large systems by the statistical quality of the "wings" of the histogram. This latter effect is quite important in systems with competing interactions for which short range order effects might occur over very broad temperature ranges or even give rise to frustration that produces a very complicated energy landscape and limit the efficiency of other methods.

In this paper, we introduce a new, general, efficient Monte Carlo (MC) algorithm that offers substantial advantages over existing approaches. Unlike conventional Monte Carlo methods that directly generate a canonical distribution at a given temperature $g(E)e^{-E/K_{\mathrm{B}}T}$, our approach is to estimate the density of states $g(E)$ accurately via a random walk which produces a flat histogram in energy space. The method can be further enhanced by performing multiple random walks, each for a different range of energy, either serially or in parallel fashion (each walk moves in a bounded range of energy so that moves that go outside the range are rejected). The resultant pieces of the density of states can be joined together and used to produce canonical averages for the calculation of thermodynamic quantities at essentially any temperature. We will apply our algorithm to the 2D ten state Potts model and Ising model which have 1st- and 2nd-order phase transitions, respectively, to demonstrate the efficiency and accuracy of the method.

Our algorithm is based on the observation that if we perform a random walk in energy space with a probability proportional to the reciprocal of the density of states $\frac{1}{g(E)}$, then a flat histogram is generated for the energy distribution. This is accomplished by modifying the estimated density of states in a systematic way to produce a "flat" histogram over the allowed range of energy and simultaneously making the density of states converge to the true value. At the very beginning of the random walk, the density of states is *a priori* unknown, so we simply set all densities of states $g(E)$ for all energies E to $g(E) = 1$. Then we begin our random walk in energy space by flipping spins randomly. In general, if E_1 and E_2 are energies before and after a spin is flipped, the transition probability from energy level E_1 to E_2 is

$$p(E_1 \to E_2) = \min\left[\frac{g(E_1)}{g(E_2)}, 1\right]. \tag{1}$$

This is also the probability to flip the spin. Each time an energy level E is visited, we update the corresponding density of states by multiplying the existing value by a modification factor $f > 1$, i.e., $g(E) \to g(E)f$. The initial modification factor can be as large as $f = f_0 = e^1 \simeq 2.71828\ldots$ which allows us to reach all possible energy levels very quickly, even for large systems. We keep walking randomly in energy space and modifying the density of states until the accumulated histogram $H(E)$ is "flat." At this point, the density of states converges to the true value with an accuracy proportion to $\ln(f)$. We then reduce the modification factor to a finer one according to some recipe such as $f_1 = \sqrt{f_0}$ (any function that monotonically decreases to 1 will do) and reset the histogram $H(E) = 0$. Then we begin the next level random walk with a finer modification factor $f = f_1$, continuing until the histogram is again "flat" after which we stop and reduce the modification factor as before, i.e., $f_{i+1} = \sqrt{f_i}$. The MC steps needed for a given f generally increase as we refine the modification factor. We stop the simulation process when

 0031-9007/01/86(10)/2050(4)$15.00

the modification factor is smaller than some predefined final value [such as $f_{\text{final}} = \exp(10^{-8}) \simeq 1.000\,000\,01$]. It is very clear that the modification factor f in our random walk acts as a control parameter for the accuracy of the density of states during the simulation and also determines how many MC sweeps are necessary for the whole simulation. It is impossible to obtain a perfectly flat histogram and the phrase "flat histogram" in this paper means that histogram $H(E)$ for all possible E is not less than 80% of the average histogram $\langle H(E)\rangle$. Since the density of states is modified every time the state is visited, we obtain a relative density of states only at the end of the simulation. To calculate the absolute values, we use the condition that the number of ground states for the Ising model is 2 (all spins are up or down) to rescale the density of states; and if multiple walks are performed within different energy ranges, they must be matched up at the boundaries in energy.

Because of the exponential growth of the density of states in energy space, it is not efficient to simply update the density of states until enough histogram entries are accumulated. All methods based on the accumulation of entries, such as the histogram method [7], Lee's version of the multicanonical method (entropic sampling) [3], the broad histogram method [8], and the flat histogram method [9,10] have the problem of scalability for large systems. These methods suffer from systematic errors and substantial deviations which increase rapidly for large system size. The algorithm proposed in this paper is of both high efficiency and accuracy over wide ranges of temperature for sizes that are beyond those that are tractable by other approaches.

We should point out here that during the random walk (especially for the early stage of iteration), the algorithm does not exactly satisfy the detailed balance condition, since the density of states is modified constantly during the random walk in energy space; however, after many iterations, the density of states converge to the true value very quickly as the modification factor approaches 1. From Eq. (1), we have

$$\frac{1}{g(E_1)}\, p(E_1 \to E_2) = \frac{1}{g(E_2)}\, p(E_2 \to E_1)\,, \qquad (2)$$

where $\frac{1}{g(E_1)}$ is the probability at the energy level E_1 and $p(E_1 \to E_2)$ is the transition probability from E_1 to E_2 for the random walk. We can thus conclude that the detailed balance condition is satisfied to within the accuracy proportion to $\ln(f)$.

The convergence and accuracy of our algorithm may be tested for a system with a 2nd order transition, the $L \times L$ Ising square lattice with nearest neighbor coupling which is generally perceived as an ideal benchmark for new theories [11] and simulation algorithms [7,12]. We simulated both small lattices for which exact results are available as well as $L = 256$ for which exact enumeration is impossible. In Fig. 1, the densities of states estimated by our algorithm are shown along with the exact results obtained by the method proposed by Beale [13]. We only show the density for systems up to $L = 50$ which is the maximum size we can calculate with the Mathematica program used in Ref. [13]. Since no difference is visible, we show the relative error $\varepsilon(\log(g(E)))$, which is defined as $\varepsilon(X) \equiv |(X_{\text{sim}} - X_{\text{exact}})/X_{\text{exact}}|$ for a general quantity X in this paper. With our algorithm we obtain an average error as small as 0.035% on the 32×32 lattice with 7×10^5 sweeps. It is possible to estimate the density of states for small systems with the broad histogram method [8]. Recent broad histogram simulational data [14] for the 2D Ising model on a 32×32 lattice with 10^6 MC sweeps yielded an average deviation of the microcanonical entropy from about 0.08% from the exact solution [13].

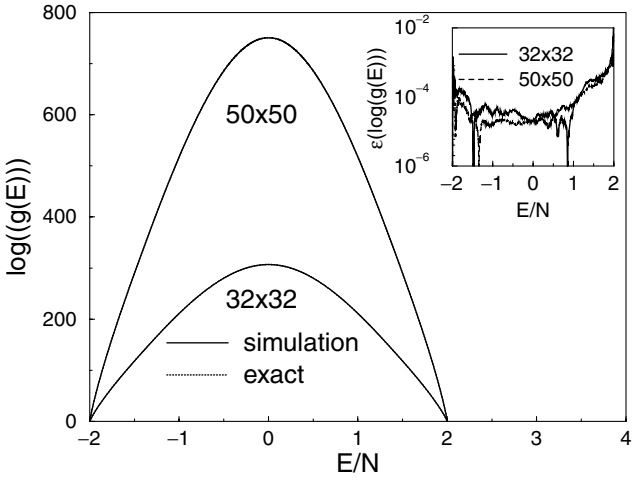

FIG. 1. Comparison of the density of states obtained by our algorithm for 2D Ising model and the exact results calculated by the method in Ref. [13]. Relative errors $\varepsilon(\log(g(E)))$ are shown in the inset.

With the Monte Carlo algorithm proposed in this paper, we can estimate the density of states efficiently even for large systems. Because of the symmetry of the density of states for Ising model $g(E) = g(-E)$, we need only to estimate the density of states in the region $E/N \in [-2, 0]$, where N is total lattice sites. To speed up the simulation for $L = 256$, we perform 15 independent random walks, each for a different region of energy from $E/N = -2$ to $E/N = 0.2$ using $f_{\text{final}} = \exp(10^{-8})$. To reduce the "boundary effect," random walks over adjacent energy regions overlap by $\Delta E/N = 0.06$. The density of states for $E/N \in [-2, 0.2]$ is obtained by joining 15 densities of states from random walks on different energy regions using a total simulational effort of only 6.1×10^6 MC sweeps.

One advantage of our algorithm is that we can readily calculate the Gibbs free energy and the entropy, quantities which are not directly available in conventional Monte Carlo simulations. With the density of states, the Gibbs free energy can be calculated by

$$F(T) = -k_{\text{B}} T \ln(Z) = -k_{\text{B}} T \ln\left(\sum_E g(E) e^{-\beta E}\right). \qquad (3)$$

Although it is impossible to calculate the exact density of states of Ising model on a lattice as large as $L = 256$ with the method proposed by Beale [13], the free energy and specific heat were calculated exactly by Ferdinand and Fisher [15] on finite-size lattices. In Fig. 2, we compare simulational data and exact solutions for the Gibbs free energy as a function of temperature. The agreement is excellent and a more stringent test of the accuracy shows that the relative error $\varepsilon(F)$ is smaller than 0.0008% for temperature region $T \in [0, 8]$.

The entropy is a very important thermodynamic quantity that cannot be calculated directly in conventional Monte Carlo simulations. It can be estimated by integrating over other thermodynamic quantities, such as specific heat, but such calculations are not so reliable since the specific heat itself is not so easy to estimate accurately. With an accurate density of states estimated by our method, the entropy can be calculated easily by $S(T) = \frac{U(T)-F(T)}{T}$, where $U(T) = \langle E \rangle_T \equiv \sum_E Eg(E)e^{-\beta E} / \sum_E g(E)e^{-\beta E}$ is the internal energy. According to our calculation, the errors for $L = 256$ are smaller than 1.2% in all temperature regions $T \in [0, 8]$ [16]. Very recently, with the flat histogram method [9] and the broad histogram method [8], the entropy was estimated with 10^7 MC sweeps for the same model on a 32×32 lattice; however, the errors in Ref. [10] are even much bigger than our errors for 256×256.

A more stringent test of the accuracy of the density of states is calculation of the specific heat defined by the fluctuation expression

$$C(T) = \frac{\langle E^2 \rangle_T - \langle E \rangle_T^2}{T^2}. \quad (4)$$

Our simulational data on the finite-size lattice are compared with the exact solution obtained by Ferdinand and Fisher [15] in Fig. 3. A stringent test of the accuracy is provided by the inset which shows the relative error $\varepsilon(C)$. The average error over the entire range $T \in [0.4, 8]$ only used a total of 6.1×10^6 MC sweeps is 0.39%. The relative errors are not bigger than 4.5% even with fine scale near T_c. Recently, Wang *et al.* [12] estimated the specific heat of the same model on a 64×64 lattice by the transition matrix Monte Carlo reweighting method [17], and with 2.5×10^7 MC sweeps, the maximum error in temperature region $T \in [0, 8]$ was about 1%. When we apply our algorithm to the same model on the 64×64 lattice, with a final modification factor of 1.000 000 001 and a total of 2×10^7 MC sweeps on single processor, the errors of the specific heat are reduced below 0.7% for all temperature [16]. The relatively large errors at low temperature reflect the small values for the specific heat at low temperature. The errors in specific heat estimated from the density of states with a broad histogram method are obviously visible even for systems as small as 32×32 [8], whereas with our method, such differences are invisible even for a system as large as 256×256.

With our algorithm, we not only dramatically reduce the computational effort by avoiding multiple simulations for different temperatures close to the transition, but also overcome the slow kinetics at low temperature or near T_c for both 1st and 2nd-order phase transitions since the random walk does not depend on the temperature. To show how our simulation method overcomes the tunneling barrier between order and disorder phases at a 1st-order phase transition, we perform random walks to calculate the density of states for the 2D ten state Potts model [18] with nearest neighbor interactions on square lattices of size $60 \leq L \leq 200$. In Fig. 4, we show the canonical distributions at the temperatures at which the peaks are of equal height. Because of the double peak structure of strongly 1st-order phase transitions [19], conventional Monte Carlo simulations are not efficient since it takes an extremely long time

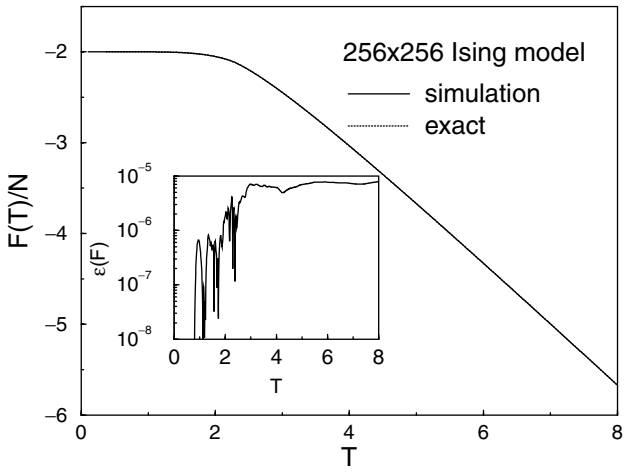

FIG. 2. Comparison of the Gibbs free energy per lattice site calculated directly from the density of states from our simulation for the $L = 256$ Ising model and the exact solutions from Ref. [15]. The relative errors $\varepsilon(F)$ are shown in the inset. The density of states was obtained by random walks with only 6.1×10^6 MC sweeps totally.

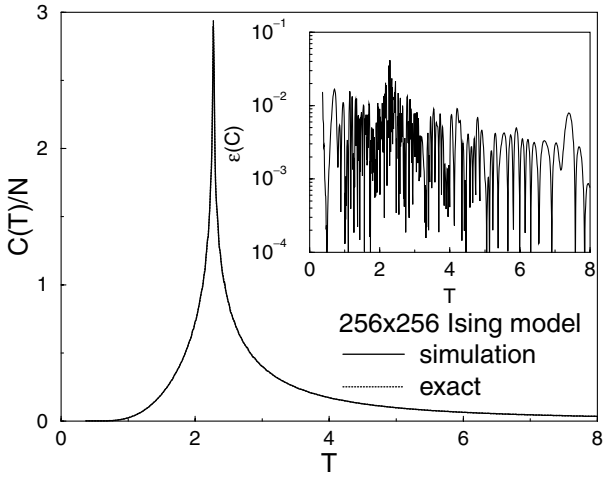

FIG. 3. Specific heat for the 2D Ising model on a 256×256 lattice in a wide temperature region. The relative error $\epsilon(C)$ is shown in the inset in the figure.

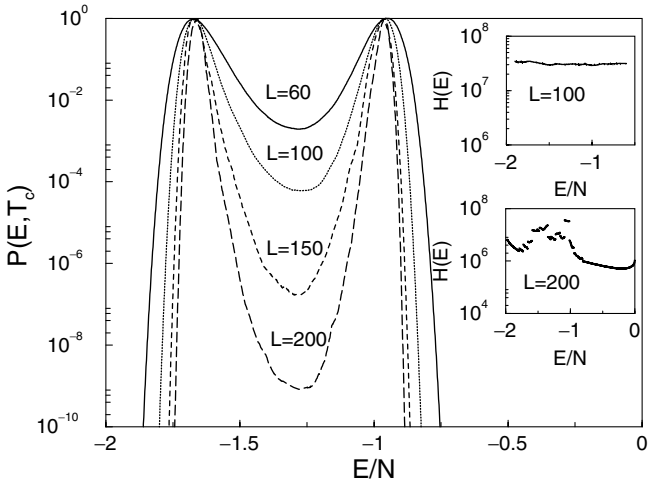

FIG. 4. Canonical distributions for the 2D ten state Potts model on $L \times L$ lattice at T_c $P(E, T_c) \equiv g(E)e^{-E/K_B T_c}$. For $L = 150$ and 200, multiple random walks were performed in different energy regions with locally flat histograms. The peaks are normalized to 1. $T_c(L)$ is 0.701 243 for $L = 200$; $T_c(\infty) = 0.701\,232\ldots$ (exact solution). In the inset, we show the overall histograms for $L = 100$ and 200. 3.1×10^7 visits per energy level (6.2×10^7 MC sweeps) were used for $L = 100$ with a single random walk. With multiple random walks, the density of states for $L = 200$ was obtained with only 9.8×10^6 visits per energy level.

to tunnel from one peak to the other. Considering the valley which we find for $L = 200$ is as deep as 9×10^{-10}, it is impossible for conventional Monte Carlo algorithms to overcome such a tunneling barrier with available computational resources.

All thermodynamic quantities we discussed so far are directly related to energy. It is also possible to calculate any quantities which may not directly relate to energy [16]. The random walk is not restricted to energy space, and our algorithm can be applied to any other parameter space. To apply our algorithm to a new system, the only thing we need to know is the Hamiltonian, and the algorithm can then be optimized to estimate the relevant density of states to the property and temperature range of interest.

We thank C. K. Hu, N. Hatano, P. D. Beale, S. P. Lewis, and H-B Schuttler for helpful comments and suggestions. The project was supported by the National Science Foundation under Grant No. DMR-9727714.

[1] D. P. Landau and K. Binder, *A Guide to Monte Carlo Methods in Statistical Physics* (Cambridge University Press, Cambridge, England, 2000).
[2] R. H. Swendsen and J.-S. Wang, Phys. Rev. Lett. **58**, 86 (1987); U. Wolff, Phys. Rev. Lett. **62**, 361 (1989).
[3] B. A. Berg and T. Neuhaus, Phys. Rev. Lett. **68**, 9 (1992); Phys. Lett. B **267**, 249 (1991); J. Lee, Phys. Rev. Lett. **71**, 211 (1993); B. A. Berg, J. Stat. Phys. **82**, 323 (1996); Nucl. Phys. **B63**, 982 (1998).
[4] W. Janke and S. Kappler, Phys. Rev. Lett. **74**, 212 (1995).
[5] B. A. Berg and T. Celik, Phys. Rev. Lett. **69**, 2292 (1992); B. A. Berg and W. Janke, Phys. Rev. Lett. **80**, 4771 (1998); N. Hatano and J. E. Gubernatis, cond-mat/0008115.
[6] N. A. Alves and U. Hansmann, Phys. Rev. Lett. **84**, 1836 (2000).
[7] A. M. Ferrenberg and R. H. Swendsen, Phys. Rev. Lett. **61**, 2635 (1988); **63**, 1195 (1989).
[8] P. M. C. de Oliveira, T. J. P. Penna, and H. J. Herrmann, Eur. Phys. J. B **1**, 205 (1998); P. M. C. de Oliveira, Eur. Phys. J. B **6**, 111 (1998).
[9] Jian-Sheng Wang, Eur. Phys. J. B **8**, 287 (1998); cond-mat/9909177.
[10] Jian-Sheng Wang and Lik Wee Lee, Comput. Phys. Commun. **127**, 131 (2000).
[11] D. P. Landau, Phys. Rev. B **13**, 2997 (1976).
[12] Jian-Sheng Wang, T. K. Tay, and R. H. Swendsen, Phys. Rev. Lett. **82**, 476 (1999).
[13] P. D. Beale, Phys. Rev. Lett. **76**, 78 (1996).
[14] A. R. Lima, P. M. C. de Oliveira, and T. J. P. Penna, J. Stat. Phys. **99**, 691 (2000).
[15] A. E. Ferdinand and M. E. Fisher, Phys. Rev. **185**, 832 (1969).
[16] Fugao Wang and D. P. Landau (unpublished).
[17] R. H. Swendsen, J. S. Wang, S. T. Li, B. Diggs, C. Genovese, and J. B. Kadane, cond-mat/9908461.
[18] F. Y. Wu, Rev. Mod. Phys. **54**, 235 (1982).
[19] M. S. S. Challa, D. P. Landau, and K. Binder, Phys. Rev. B **34**, 1841 (1986).

Hybrid Monte Carlo method for condensed-matter systems

B. Mehlig, D. W. Heermann, and B. M. Forrest
Institut für Theoretische Physik, Universität Heidelberg, Philosophenweg 19, and Interdisziplinäres Zentrum für Wissenschaftliches Rechnen der Universität Heidelberg, 6900 Heidelberg, Germany
(Received 7 June 1991)

In this paper the static properties of the hybrid Monte Carlo algorithm are studied in the context of condensed-matter systems. The algorithm is used to simulate a Lennard-Jones liquid near the coexistence region. The hybrid Monte Carlo algorithm generates a canonical distribution in configuration space, permitting the application of the Ferrenberg-Swendsen extrapolation scheme. Moreover, it is an exact algorithm: The configurational averages prove to be independent of the step size, and the algorithm does not suffer from numerical instabilities due to finite step size.

I. INTRODUCTION

It has become widely accepted that the hybrid Monte Carlo (HMC) algorithm proposed by Duane *et al.* in their seminal paper[1] is a promising alternative to both microcanonical and Monte Carlo simulations in the context of lattice gauge theories.[2–6] HMC is a global algorithm, it has been shown to reduce critical slowing down for free-field theories.[7] Moreover, it is an exact algorithm; i.e., the ensemble averages do not depend on the step size chosen and the algorithm does not suffer from numerical instabilities due to finite step size as molecular-dynamics (MD) algorithms do.

It has been suggested to apply the HMC algorithm to condensed-matter systems.[8,9] This is done in the present paper. It is organized into five sections. In Sec. II the Monte Carlo method for condensed-matter systems is reviewed in order to state the terminology. The algorithm is described in Sec. III. Following the outline given in Ref. 1, it is shown that the discretization scheme of Hamilton's equations has to be *area preserving* and *time reversible* so as to guarantee detailed balance. In Sec. IV, the simulations are presented. The results and prospects for future work are summarized in Sec. V.

II. THE MONTE CARLO METHOD FOR CONDENSED-MATTER SYSTEMS

A large class of many-particle systems can be described by a classical Hamiltonian of the form

$$\mathcal{H}(x,p)=\mathcal{T}(p)+\mathcal{U}(x)=\sum_{i=1}^{N}\frac{p_i^2}{2m}+\sum_{i<j}u_{ij}\ ,$$

where N is the number of particles and u_{ij} is a rotationally invariant pair potential

$$u_{ij}=u(|x_i-x_j|)\ .$$

The canonical ensemble average of an observable $\mathcal{O}(x,p)$ is given by

$$\langle\mathcal{O}(x,p)\rangle=\frac{1}{\mathcal{Q}}\frac{1}{(2\pi\hbar)^{3N}}\frac{1}{N!}\int dx\,dp\,e^{-\beta\mathcal{H}(x,p)}\mathcal{O}(x,p)\ ,$$

where the partition function is denoted by

$$\mathcal{Q}=\frac{1}{(2\pi\hbar)^{3N}}\frac{1}{N!}\int dx\,dp\,e^{-\beta\mathcal{H}(x,p)}\ .$$

Performing the Gaussian integral over the momenta yields

$$\mathcal{Q}=\frac{1}{\lambda^{3N}}\frac{1}{N!}\int dx\,e^{-\beta\mathcal{U}(x)}\equiv\frac{1}{\lambda^{3N}}\frac{1}{N!}\mathcal{Z}\ ,$$

where λ is the thermal de Broglie wavelength and $\mathcal{Z}$ is called the configurational integral. If an observable is a function of the coordinates alone, its ensemble average is given by

$$\langle\mathcal{O}(x)\rangle=\frac{1}{\mathcal{Z}}\int dx\,e^{-\beta\mathcal{U}(x)}\mathcal{O}(x)\ . \tag{1}$$

Equation (1) can be integrated numerically using importance sampling. Given a sequence of configurations

$$x^{(i)},\quad i=1,\ldots,n \tag{2}$$

distributed according to the probability

$$p(x)dx=\frac{1}{\mathcal{Z}}e^{-\beta\mathcal{U}(x)}dx\ , \tag{3}$$

the law of large numbers implies

$$\langle\mathcal{O}(x)\rangle=\lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^{n}\mathcal{O}(x^{(i)})\ .$$

In a Monte Carlo calculation[10,11] the sequence (2) is generated as a Markov chain of configurations defined by the conditional probability densities $p_M(x^{(i)}\to x^{(i+1)})$. Provided the Markov chain is irreducible and aperiodic, the detailed balance condition

$$p(x)p_M(x\to x')dx\,dx'=p(x')p_M(x'\to x)dx\,dx' \tag{4}$$

ensures that the Markov chain converges to the unique

stationary probability distribution (3). Note that Eq. (4) is a sufficient but not a necessary condition. In practice, one step $x^{(i)}\to x^{(i+1)}$ in the Markov chain of configurations is realized by suggesting $x^{(i+1)}$ according to a probability density $p_S(x^{(i)}\to x^{(i+1)})$ and accepting it with the probability

$$P_A(x\to x')=\min\left\{1,\frac{p(x')p_S(x'\to x)dx\,dx'}{p(x)p_S(x\to x')dx\,dx'}\right\} \tag{5}$$

(Metropolis function). The conditional probability densities $p_M(x^{(i)}\to x^{(i+1)})$ are then given by

$$p_M(x\to x')=p_S(x\to x')P_A(x\to x') . \tag{6}$$

In conventional Monte Carlo calculations $p_S(x^{(i)}\to x^{(i+1)})$ allows for local moves only (single-particle updates). Updating more than one particle at a time typically results in a prohibitively low average acceptance probability $\langle P_A\rangle$. This implies large relaxation times and high autocorrelations. In a MD simulation, on the other hand, global moves are made. The MD scheme, however, is prone to errors and instabilities due to finite step size. In order to introduce temperature in the microcanonical context, so-called isokinetic MD schemes are often used.[12] However, they do not yield the canonical probability distribution (3), unlike Monte Carlo calculations. The HMC scheme, which is described in the next section, combines the advantages of MD and Monte Carlo methods: while allowing for global moves (which essentially consist in integrating the system through *phase* space), HMC is an exact algorithm, i.e., it does not suffer from step-size errors.

III. HYBRID MONTE CARLO

In the HMC scheme, global moves can be made while keeping the average acceptance probability $\langle P_A\rangle$ high. One global move in *configuration* space consists in integrating the system through *phase* space for a fixed time t using some discretization scheme $g^{\delta t}$ (δt denotes the step size) of Hamilton's equations,

$$\frac{dx}{dt}=\frac{\partial\mathcal{H}}{\partial p} ,\qquad \frac{dp}{dt}=-\frac{\partial\mathcal{H}}{\partial x} , \tag{7}$$

viz.,

$$g^{\delta t}:\ \mathbb{R}^{6N}\to\mathbb{R}^{6N} ,\qquad (x,p)\to g^{\delta t}(x,p)\equiv(x',p') ,$$

where $(x,p)\in\mathbb{R}^{6N}$ and $g^{\delta t}(x,p)\in\mathbb{R}^{6N}$. Since the system is moved deterministically through phase space, the conditional probability of suggesting configuration x' starting at x is given by

$$p_S(x\to x')dx'=p_S(p)dp . \tag{8}$$

The initial momenta are drawn from a Gaussian distribution at inverse temperature β:

$$p_S(p)\propto\exp\left[-\beta\sum_{j=1}^{N}\frac{p_j^2}{2m}\right] . \tag{9}$$

One has thus from (5), (8), and (9),

$$P_A((x,p)\to g^{\delta t}(x,p))=\min\{1,e^{-\beta\delta\mathcal{H}}\} , \tag{10}$$

where

$$\delta\mathcal{H}=\mathcal{H}(g^{\delta t}(x,p))-\mathcal{H}(x,p)$$

is the discretization error associated with $g^{\delta t}$. Using the algebraic identity

$$e^{-\mathcal{H}(x,p)}\min\{1,e^{-\delta\mathcal{H}}\}=e^{-\mathcal{H}(g^{\delta t}(x,p))}\min\{e^{\delta\mathcal{H}},1\} , \tag{11}$$

it can be shown that, for a discretization scheme which is *time reversible,*

$$g^{-\delta t}\circ g^{\delta t}=id , \tag{12}$$

and *area preserving,*

$$\det\frac{\partial g^{\delta t}(x,p)}{\partial(x,p)}=1 , \tag{13}$$

detailed balance is satisfied:

$$\begin{aligned}
p(x)p_M(x\to x')dx\,dp&=p(x)p_S(p)P_A((x,p)\to g^{\delta t}(x,p))dx\,dp\\
&=p(x')p_S(p')P_A(g^{\delta t}(x,p)\to(x,p))dx\,dp\\
&=p(x')p_S(p')P_A((x',p')\to g^{-\delta t}(x',p'))dx\,dp\\
&=p(x')p_S(p')P_A((x',p')\to g^{-\delta t}(x',p'))dx'dp'\\
&=p(x')p_M(x'\to x)dx'dp' .
\end{aligned}$$

Thus, provided the discretization scheme used is *time reversible* and *area preserving,* the HMC algorithm generates a Markov chain with the stationary probability distribution $p(x)$ defined in (3). The probability distribution (3) is entirely determined by the detailed balance condition (4). Therefore, neither $p(x)$ nor any ensemble averages depend on the step size δt chosen.

However, the average acceptance probability $\langle P_A\rangle$, because of (10), depends on the average discretization error $\langle\delta\mathcal{H}\rangle$ and hence does depend on δt. Gupta *et al.*

have shown that for $(\rho,T)\neq(\rho_c,T_c)$,

$$\langle P_A\rangle=\mathrm{erfc}(\tfrac{1}{2}\sqrt{\beta\langle\delta\mathcal{H}\rangle})$$

is a good approximation for sufficiently large systems ($N\to\infty$) and small step sizes ($\delta t\to 0$).[5] In the following an outline of their argument shall be given. From normalization and the area-preserving property one has

$$\langle e^{-\beta\delta\mathcal{H}}\rangle=1\ . \tag{14}$$

Equation (14) can be expanded into cumulants

$$\langle\delta\mathcal{H}\rangle=\frac{\beta}{2}\langle(\delta\mathcal{H}-\langle\delta\mathcal{H}\rangle)^2\rangle+\cdots .$$

In order to obtain a nonzero average acceptance probability $\langle P_A\rangle$ in the limit $N\to\infty$, one has to let $\delta t\to 0$, keeping $\langle(\delta\mathcal{H}-\langle\delta\mathcal{H}\rangle)^2\rangle$ fixed. In this limit, higher-order cumulants will vanish. The resulting distribution of the discretization error will thus be Gaussian with mean and width related through

$$\langle\delta\mathcal{H}\rangle=\frac{\beta}{2}\langle(\delta\mathcal{H}-\langle\delta\mathcal{H}\rangle)^2\rangle\ . \tag{15}$$

From (10) and (15) one has, in this case,

$$\begin{aligned}\langle P_A\rangle&=\frac{1}{\sqrt{2\pi\langle(\delta\mathcal{H}-\langle\delta\mathcal{H}\rangle)^2\rangle}}\\&\quad\times\int_{-\infty}^{\infty}dt\,\min\{1,e^{-\beta t}\}\\&\qquad\times\exp\left[-\frac{(t-\langle\delta\mathcal{H}\rangle)^2}{2\langle(\delta\mathcal{H}-\langle\delta\mathcal{H}\rangle)^2\rangle}\right]\\&=\mathrm{erfc}(\tfrac{1}{2}\sqrt{\beta\langle\delta\mathcal{H}\rangle})\ .\end{aligned} \tag{16}$$

The square root in (16) is always well defined since (14) implies

$$\langle\delta\mathcal{H}\rangle\geq 0\ .$$

Equality holds in the limit $\delta t\to 0$, where energy is conserved exactly and $\langle P_A\rangle=1$. Increasing the step size will result in a lower average acceptance probability $\langle P_A\rangle$. Varying δt, the average acceptance probability $\langle P_A\rangle$ can thus be adjusted to minimize autocorrelations. This will be done in the next section. The simulations presented below were carried out using the HMC scheme as described above. Before discussing the results of the simulations, however, several remarks concerning possible generalizations of this scheme are due.

Firstly, the momenta do not necessarily have to be drawn from the Gaussian distribution (9). A particularly simple and computationally efficient alternative to (9) would be a uniform momentum distribution. This choice, however, did not prove successful, since a cutoff has to be introduced for computational reasons. This cutoff must be taken into account in P_A, leading to a very low average acceptance probability $\langle P_A\rangle$.

Secondly, it is clear that instead of choosing a discretization scheme of Hamilton's equations (7), any time-reversible and area-preserving discrete mapping can be used to propagate the system through phase space.[2] It remains to be seen whether or not using a discretization of Hamilton's equations is the optimal choice. Concluding this section, it should be remarked that it may be possible to formulate more general algorithms, which do not satisfy detailed balance (4) but the relation

$$\int dx\,p(x)p_M(x\to x')=p(x')\ ,$$

which is a necessary and sufficient condition for the Markov chain (2) to have the stationary probability distribution (3).

IV. SIMULATIONS

Using a given discretization scheme of Hamilton's equations, one is left with two parameters to be varied: the step size δt and the number N_{MD} of integration steps per Monte Carlo step (or, alternatively, the trajectory length $t=N_{\mathrm{MD}}\delta t$). In this section the dependence of ensemble averages, autocorrelations, and relaxation on these parameters is discussed.

A. Testing the algorithm

The algorithm was tested on a three-dimensional Lennard-Jones system with $N=256$ particles. A pair potential with a cutoff at 2.5σ was used:

$$u_{ij}^*=\begin{cases}4\left[\dfrac{1}{r_{ij}^{*12}}-\dfrac{1}{r_{ij}^{*6}}\right] & \text{if } r_{ij}^*\leq 2.5\\ 0 \quad \text{otherwise}\ ,\end{cases}$$

where $r_{ij}^*=|x_i^*-x_j^*|$ and the asterisk denotes scaled units. The cutoff was *not* corrected for. The parallel code[13] was written in OCCAM. Most of the simulations were carried out at temperature $T^*=0.72$ and density $\rho^*=0.83$, i.e., just above the coexistence curve. In Fig. 1 the approximate location of this point in the (ρ^*,T^*) plane is shown.

Since higher-order *time-reversible* and *area-preserving* discretization schemes do not seem to have been particularly successful in the context of lattice gauge theories, the *leap frog* scheme was chosen for the present purpose.[15]

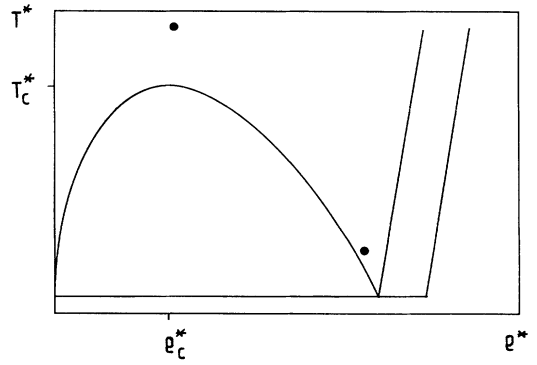

FIG. 1. Approximate location of simulated phase points (dots). The phase diagram is taken from Ref. 14; T^* and ρ^* are given in scaled units as in Ref. 14.

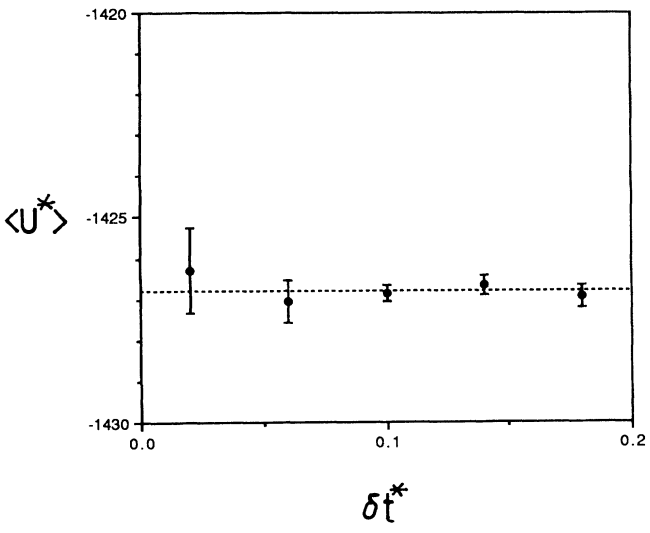

FIG. 2. The ensemble average of the potential energy $\mathcal{U}^*$ as a function of the step size δt^* for $N_{\mathrm{MD}}=10$ (scaled units).

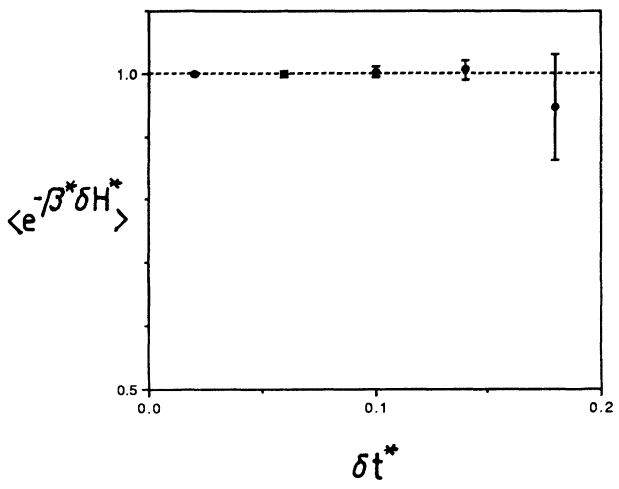

FIG. 4. The normalization condition: $\langle e^{-\beta^*\delta\mathcal{H}^*}\rangle$ is expected to be equal to one for all step sizes δt^* (scaled units).

One of the major advantages of HMC is the fact that it is an exact algorithm, i.e., the ensemble averages should not depend on the step size δt^* chosen. In Fig. 2, the step-size dependence of $\langle \mathcal{U}^* \rangle$ is shown. The errors are given by $2[(2\tau/n)\sigma^2(\mathcal{U}^*)]^{1/2}$, where n is the number of Monte Carlo steps and τ is the autocorrelation time which was determined using a standard statistical analysis as described in Ref. 16. Clearly the ensemble average $\langle \mathcal{U}^* \rangle$ does not depend on the step size chosen. In Fig. 3 the corresponding average acceptance probabilities $\langle P_A \rangle$ are shown. For $\delta t^* \to \infty$ one has $\langle P_A \rangle \to 0$, as expected from (16). For $\delta t^* \to 0$ one has $\langle P_A \rangle \to 1$. In this limit energy $\mathcal{H}^*$ is conserved exactly. It should be emphasized that, for a step size three times as large as the conventional molecular-dynamics step size at the same temperature and density, the average acceptance probability is still 0.3.

In order to test the normalization condition (14), $\langle e^{-\beta^*\delta\mathcal{H}^*}\rangle$ is plotted in Fig. 4 for different step sizes δt^*. It is clear that the normalization condition is satisfied.

In Fig. 5, the average acceptance probability $\langle P_A \rangle$ is plotted against the average discretization error $\langle \delta\mathcal{H}^* \rangle$. The $N=256$ particle system is clearly large enough for the approximation (16) to hold.

B. Autocorrelation and relaxation times

In order to ascertain the performance of the HMC algorithm, autocorrelation times τ for the potential energy

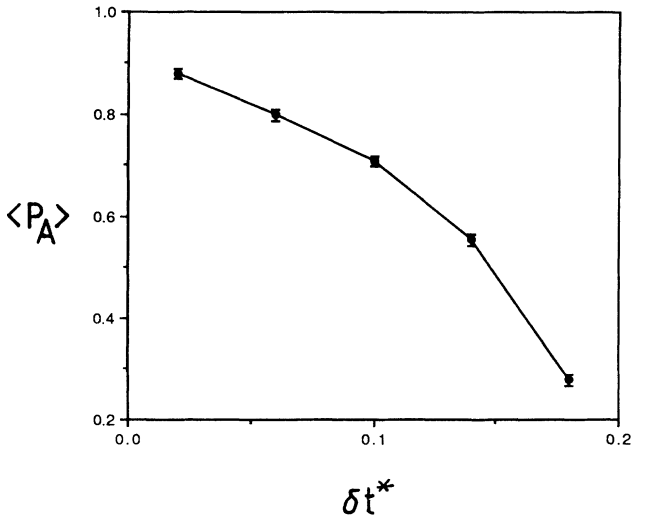

FIG. 3. The average acceptance probability $\langle P_A \rangle$ as a function of the step size δt^* (scaled units) for $N_{\mathrm{MD}}=10$.

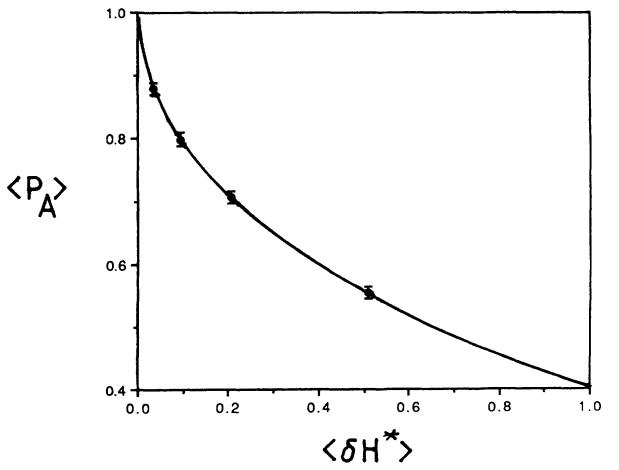

FIG. 5. The average acceptance probability $\langle P_A \rangle$ as a function of the average discretization error $\langle \delta\mathcal{H}^* \rangle$ for $N_{\mathrm{MD}}=10$ (dots). The solid curve represents the analytic estimate of Gupta *et al.* (Ref. 5). All quantities are given in scaled units.

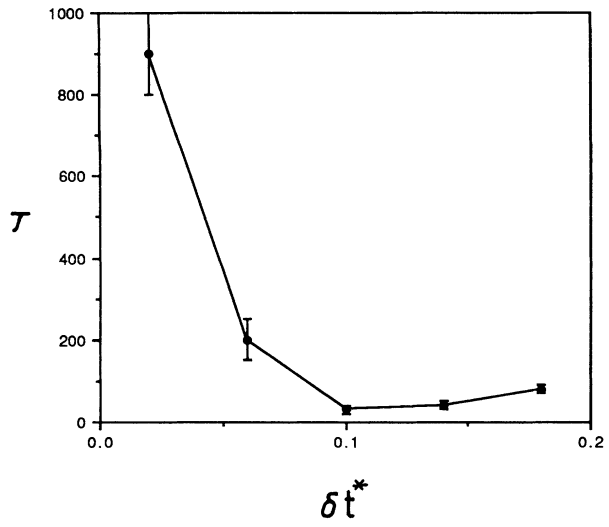

FIG. 6. The autocorrelation time τ of the potential energy $\mathcal{U}^*$ as a function of the step size δt^* for $N_{\mathrm{MD}}=10$, τ is given in units of integration steps.

$\mathcal{U}^*$ were measured for different values of δt^* and N_{MD}. The optimal trajectory length was determined by plotting τ against δt^*. This is shown in Fig. 6 for $N_{\mathrm{MD}}=10$. (In order to facilitate comparison with molecular-dynamics data, τ is given in units of integration steps.) The autocorrelation time τ is minimal for $t^*=N_{\mathrm{MD}}\delta t^*\approx 1$ and rises substantially for $\delta t^*\to 0$, while slightly increasing for large step sizes. This is as expected, since, for small step sizes, consecutive configurations will be highly correlated whereas large step sizes lead to large $\langle \delta\mathcal{H}\rangle$ and thus small $\langle P_A\rangle$, implying high correlations. It seems

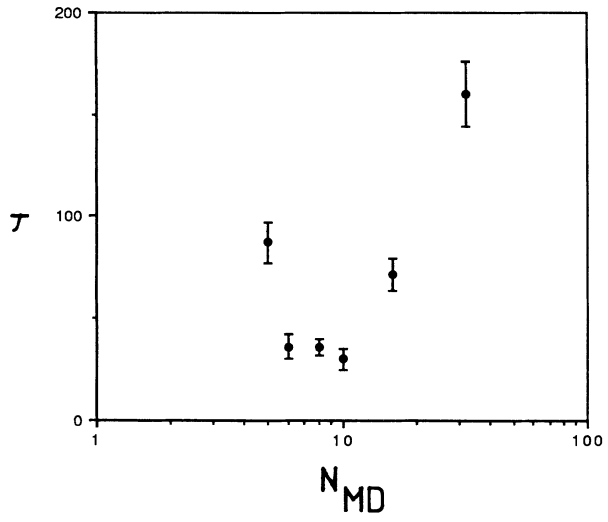

FIG. 7. The autocorrelation time τ of the potential energy $\mathcal{U}^*$ as a function of the number N_{MD} of molecular dynamics steps per Monte Carlo step for $t^*=N_{\mathrm{MD}}\delta t^*=1$, τ is given in units of integration steps.

TABLE I. Autocorrelation times for the potential energy $\mathcal{U}^*$ in units of integration steps for hybrid Monte Carlo (HMC) and molecular dynamics (MD).

		τ
HMC	$t^*=1$, $N_{\mathrm{MD}}=6$	36
MD	$\delta t^*=0.01$	40
	$\delta t^*=0.06$	9
	$\delta t^*=0.10$	5

interesting to note that the optimal trajectory length $t^*=1$ corresponds to $\langle P_A\rangle\approx 0.7$ and not $\frac{1}{2}$.

In a second set of runs N_{MD} was varied keeping the trajectory length constant at $t^*=1$. Figure 7 shows that the autocorrelation time rises substantially for small N_{MD}. The minimum in Fig. 7 therefore approximately corresponds to minimum cost. In Table I, this autocorrelation time is compared with those obtained from a MD simulation at the same temperature and density. For small MD step sizes, HMC and MD autocorrelation times are of the same order of magnitude. For larger MD step sizes, the HMC autocorrelation time is longer.

The relaxation from a fcc lattice to equilibrium at temperature $T^*=0.72$ and density $\rho^*=0.83$ was also compared with (isokinetic) MD. Figure 8 shows that the relaxation time is of the same order of magnitude for HMC and for isokinetic MD.

C. Ferrenberg-Swendsen extrapolation

Another major advantage of the HMC scheme is that the ensemble averages are performed in the canonical ensemble. One can thus make use of an extrapolation scheme, which has been proposed by Ferrenberg *et al.*[17,18] for the Ising model and by Falcioni *et al.*[19] in the context of lattice gauge theories. In this scheme, the fact that the ensemble simulated is canonical,

$$P_{\beta^*}(\mathcal{U}^*)\propto e^{-\beta^*\mathcal{U}^*},$$

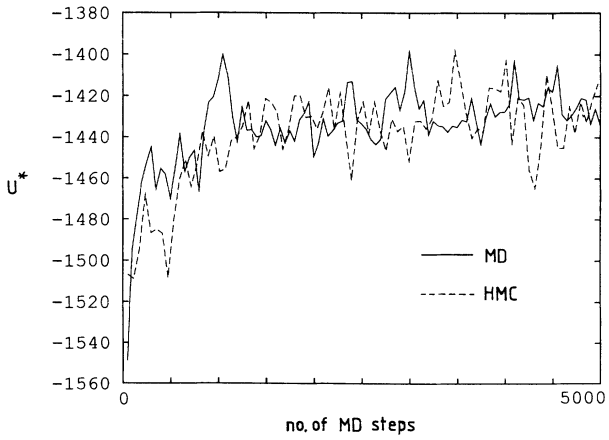

FIG. 8. The relaxation from a fcc lattice to equilibrium for hybrid Monte Carlo ($t^*=1$ and $N_{\mathrm{MD}}=6$) and for isokinetic molecular dynamics ($\delta t^*=0.02$).

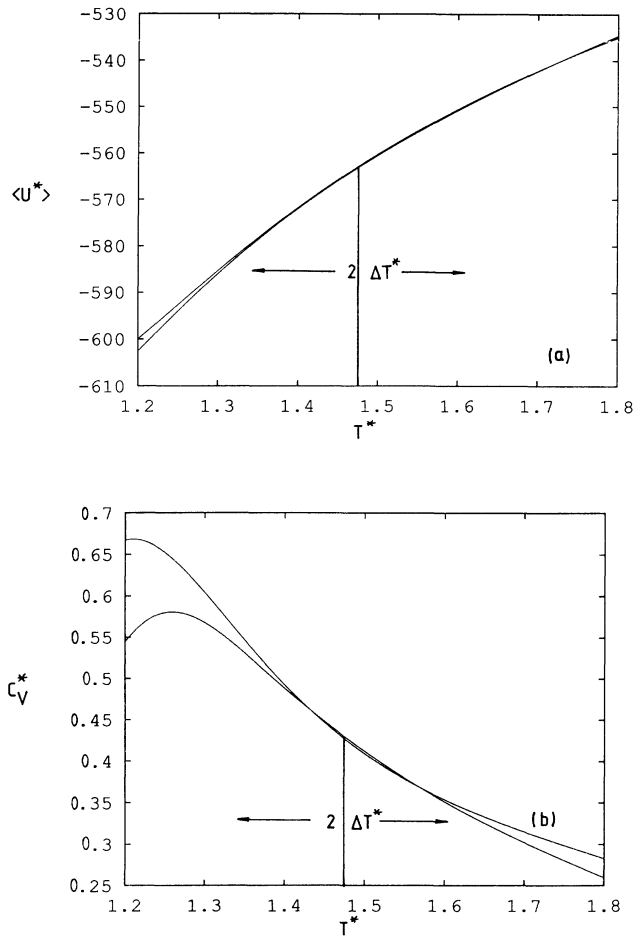

FIG. 9. Ferrenberg-Swendsen extrapolation from two statistically independent runs for (a) $\langle \mathcal{U}^* \rangle$ and for (b) C_V^* (bottom). The estimate (18) for the extrapolation regime $2\Delta T^*$ is also shown. All quantities are given in scales units.

is used to extrapolate ensemble averages measured at a temperature β^* to other temperatures $\beta^{*\prime}$. The probability distribution $P_{\beta^{*\prime}}(\mathcal{U}^*)$ of the potential energy at temperature $\beta^{*\prime}$ is extrapolated from $P_{\beta^*}(\mathcal{U}^*)$ according to

$$P_{\beta^{*\prime}}(\mathcal{U}^*)=\frac{P_{\beta^*}(\mathcal{U}^*)e^{-(\beta^{*\prime}-\beta^*)\mathcal{U}^*}}{\sum_{\mathcal{U}^*}P_{\beta^*}(\mathcal{U}^*)e^{-(\beta^{*\prime}-\beta^*)\mathcal{U}^*}} . \tag{17}$$

In a HMC simulation, the probability distribution $P_{\beta^*}(\mathcal{U}^*)$ is obtained in the form of a histogram. Using $P_{\beta^{*\prime}}(\mathcal{U}^*)$, the mean $\langle \mathcal{U}^* \rangle$ and higher moments can be calculated for other temperatures $\beta^{*\prime}$. One has to be careful, however, not to extrapolate to temperatures too far above or below the simulated temperature β^*, since the tails of the distribution $P_{\beta^*}(\mathcal{U}^*)$ cannot be sampled with sufficient statistical accuracy.[20] An estimate for a sensible extrapolation regime $\Delta T^*=|T^*-T^{*\prime}|$ should be given by

$$\Delta T^* \approx \frac{1}{N}\frac{\sqrt{\sigma^2(\mathcal{U}^*)}}{c_V^*} = \frac{T^*}{\sqrt{Nc_V^*}} , \tag{18}$$

where c_V^* denotes the specific heat:

$$c_V^*=\frac{\langle \mathcal{U}^{*2} \rangle - \langle \mathcal{U}^* \rangle^2}{T^{*2}N} .$$

In three dimensions, at the critical point, c_V^* scales as $c_V^* \sim N^{\alpha/3\nu}$ for large N. Using the scaling relation $3\nu=2-\alpha$,[21] one has from (18)

$$\Delta T^* \sim \sqrt{N^{-(1+\alpha/3\nu)}} \sim N^{-1/3\nu} .$$

The extrapolation regime thus decreases with increasing system size N.[22]

In the present case the mean $\langle \mathcal{U}^* \rangle$ and the specific heat c_V^* were extrapolated from simulations at temperature $T^*=1.475$ and density $\rho^*=0.35$. In order to test the reliability of the Ferrenberg-Swendsen scheme for condensed-matter systems, extrapolations from statistically independent runs were performed. The extrapolations from two runs are shown in Fig. 9. The simulation temperature is indicated by a vertical line. As can be seen in Fig. 9, the extrapolation works very well for the mean $\langle \mathcal{U}^* \rangle$: within $2\Delta T^*$ the two extrapolations coincide. Apparently the necessity of discretizing $P_{\beta^*}(\mathcal{U}^*)$ does not seem to affect the extrapolation. For the specific heat c_V^* the extrapolation works less satisfactorily. Better statistics are needed to improve the specific-heat data.

V. CONCLUSIONS

In the present paper, the HMC algorithm has been successfully used to simulate a condensed-matter system. It has been shown that the HMC scheme combines the advantages of molecular-dynamics and Monte Carlo algorithms: while the autocorrelation times are of the same order of magnitude as in MD calculations, HMC does not suffer from errors or instabilities due to finite step size. Moreover, it generates a canonical distribution in configuration space, so that the Ferrenberg-Swendsen extrapolation scheme can be applied. The latter has been shown to yield reasonable results in a comparatively small extrapolation range which can be estimated by (18).

The dynamical properties of the HMC algorithm remain to be investigated for condensed-matter systems. It is not clear, in particular, whether or not HMC reduces critical slowing down for the system discussed in this paper.

At present, work is in progress concerning the application of HMC to polymer systems. It is hoped that one may thus ascertain the performance of HMC when simulating a system with a spectrum of relaxation times.

Using the Andersen-Parinello-Rahman method,[12] the HMC scheme could be generalized to simulate a constant-pressure ensemble. The Ferrenberg-Swendsen extrapolation scheme may then permit the location of the critical point of the Lennard-Jones system.

ACKNOWLEDGMENTS

The computations presented here were performed on the Supercluster at the Interdisziplinäres Zentrum für Wissenschaftliches Rechnen, Heidelberg. B.M. would like to thank Bernhard Przywara for his invaluable assistance in using the OCCAM programming environment. B.M.F. and B.M. would like to thank Brian Pendleton for discussions, in particular, for pointing out that time reversibility does not guarantee detailed balance. D.W.H. would like to thank P. Nielaba for fruitful discussions. Partial support by the SFB 123 and the BMFT Project Nos. 0326657D and 031240284 is gratefully acknowledged.

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

[2]M. Creutz, Phys. Rev. D **38**, 1228 (1988).

[3]A. D. Kennedy, Nucl. Phys. (Proc. Suppl.) **4**, 576 (1988).

[4]A. D. Kennedy, Nucl. Phys. (Proc. Suppl.) **9**, 457 (1989).

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

[6]R. Gupta, G. W. Kilcup, and S. R. Sharpe, Phys. Rev. D **38**, 1278 (1988).

[7]A. D. Kennedy and B. J. Pendleton, Nucl. Phys. **B 20**, Proc. Suppl., 118 (1991).

[8]D. W. Heermann, P. Nielaba, and M. Rovere, Comput. Phys. Commun. **60**, 311 (1990).

[9]B. J. Pendleton (private communication).

[10]D. W. Heermann, *Computer Simulation Methods in Theoretical Physics* (Springer-Verlag, Heidelberg, 1990).

[11]K. Binder and D. W. Heermann, *Monte Carlo Simulation in Statistical Physics* (Springer-Verlag, Heidelberg, 1988).

[12]K. Kremer, *Computersimulation in der Physik* (KFA Jülich, Jülich, 1989).

[13]D. W. Heermann and A. N. Burkitt, *Parallel Algorithms in Computational Science* (Springer-Verlag, Heidelberg, 1991).

[14]J. J. Nicolas, K. E. Gubbins, W. B. Streett, and D. J. Tildesley, Mol. Phys. A **5**, 1429 (1978).

[15]M. Creutz and A. Goksch, Phys. Rev. Lett. **63**, 9 (1989).

[16]M. P. Allen and D. J. Tildesley, *Computer Simulations of Liquids* (Clarendon, Oxford, 1987).

[17]A. M. Ferrenberg and R. H. Swendsen, Phys. Rev. Lett. **61**, 2635 (1988).

[18]A. M. Ferrenberg and R. H. Swendsen, Phys. Rev. Lett. **63**, 1195 (1989).

[19]M. Falcioni, E. Marinari, M. L. Paciello, G. Parisi, and B. Taglienti, Phys. Lett. **108B**, 331 (1982).

[20]P. Peczak, A. M. Ferrenberg, and D. P. Landau (unpublished).

[21]E. Brézin, J. C. Le Guillou, and J. Zinn-Justin, *Phase Transitions and Critical Phenomena* (Academic, London, 1976), Vol. 6.

[22]Off criticality ΔT^* obviously scales as $\Delta T^* \sim N^{-1/2}$.