General Cluster Updating Method for Monte Carlo Simulations

Ferenc Niedermayer
Institute for Theoretical Physics, Eötvös University, Budapest
(Received 2 May 1988)

A general cluster updating algorithm for Monte Carlo simulations is presented. The method contains arbitrary probability functions which can be used to minimize the relaxation time. It is applicable to systems where the interaction energy has a global (discrete or continuous) symmetry. Special cases cover standard site-by-site updating as well as the cluster updating method proposed by Swendsen and Wang for Ising and Potts models.

PACS numbers: 05.50.+q, 11.15.Ha, 75.40.Cx

Critical slowing down at phase transitions causes major problems for computer simulations. The new configuration obtained after a Monte Carlo (MC) sweep through the lattice is not independent from the previous one. At a critical point, not only is the correlation length ξ divergent but also the relaxation (or autocorrelation) time τ increases as ξ^z, where z is the dynamical critical exponent. The statistical error is determined by the number of independent configurations, $N_{\text{eff}}=N_{\text{sweep}}/\tau$. Since for local dynamics z is about 2 or larger,[1-7] this presents a major limitation on the size of systems. Intuitively, the reason behind such behavior is that—similarly to the problem of random walk—for local stochastic algorithms a local change needs of order R^2 steps to traverse a distance R. Consequently, the time needed to obtain an independent new configuration is proportional to ξ^2. (The reader should consult Ref. 7 for further arguments.) Clearly, critical slowing down could be beaten if one would be able to change larger regions in one step. Ideas in this direction have been suggested in the literature.[8-15]

In this paper we generalize a MC algorithm proposed by Swendsen and Wang (SW).[16] Their method has been formulated for Ising and Potts models, and is based on an equivalence between the configurations of these models and configurations of clusters in a percolation problem which has been suggested by Fortuin and Kasteleyn.[17] For an Ising model in a particular configuration "virtual bonds" between parallel spins are introduced with the probability $p=1-\exp(-2J/k_BT)$, J being the spin-spin coupling constant. No bonds are introduced between antiparallel spins. Then clusters are formed from spins connected by virtual bonds. (The smallest possible cluster contains just a single site.) The next configuration of the system is obtained by assignment of a new orientation independently to each cluster (the same to each spin in a given cluster). The new orientations are chosen randomly with uniform weight. SW have shown that detailed balance holds for this procedure, and they obtained an essential reduction of critical slowing down: The dynamical critical exponent in the energy-energy correlation function for the two-dimensional Ising model was $z=0.35$ instead of $z=2.125$ expected for single-spin-flip dynamics.[2,3] Earlier, Sweeny presented a different MC method[18] based on the same correspondence between Potts models and percolation clusters. Although his method has some advantages for Potts models in two dimensions, it is not clear how to generalize it. Below we describe the generalization of the method of Ref. 16.

We start by considering a general Hamiltonian which is a sum of pair interaction terms. Let us use the notation $S=\{s_i\}$ for the configuration of the local (discrete or continuous) variables s_i, and $E_l(S)\equiv E_{ij}(s_i,s_j)$ for the term in the total energy corresponding to a given pair of sites. Absorbing the factor $1/k_BT$ into the energy, the partition function is given by

$$Z=\mathrm{Tr}_S e^{-H}=\sum_S e^{-E(S)}=\sum_S\prod_l e^{-E_l(S)}. \quad (1)$$

We shall assume that the Hamiltonian is invariant under a global symmetry transformation $g\in G$:

$$s_i\to s_i'=gs_i,\quad E_{ij}(s_i,s_j)=E_{ij}(s_i',s_j'). \quad (2)$$

Now we introduce virtual bonds on the links: We place a bond on the link l with the probability $p_l[E_l(S)]$ where the probability functions $p_l(E_l)$ are arbitrary. The sites connected by virtual bonds form clusters as defined above. The probability to obtain a particular configuration of clusters C is given by

$$w(S\to C)=\sum_{B(C)}\prod_{l\in B}p_l[E_l(S)]\prod_{l\notin B}\{1-p_l[E_l(S)]\}, \quad (3)$$

where products are taken over links with and without bonds, respectively. The sum is over different bond configurations associated with the given cluster configuration C. [Note that for links connecting different clusters one always has a factor $1-p_l$.] Now we imagine the clusters as separate subsystems, and make an independent "global" transformation on each cluster. Denote this configuration by $S'=gS$, where $g\in G_c\equiv\prod_k G_k$ is an element from the direct product over clusters of the symmetry groups G_k acting on the kth cluster.

In order that the system relax to thermodynamic equilibrium we require that the condition of detailed

balance be satisfied by the transition probability $w(S,C\rightarrow S')$:

$$P_0(S)w(S\rightarrow C)w(S,C\rightarrow S') = P_0(S')w(S'\rightarrow C)w(S',C\rightarrow S), \quad (4)$$

where $P_0(S)\propto\exp[-E(S)]$ is the equilibrium distribution.

Since $E_l(S)=E_l(S')$ when both end points of the link belong to the same cluster, one can rewrite this condition as

$$A(S,C)w(S,C\rightarrow S')=A(S',C)w(S',C\rightarrow S), \quad (5)$$

where

$$A(S,C)\equiv\prod_{\partial C}(e^{-E_l(S)}\{1-p_l[E_l(S)]\}). \quad (6)$$

Here the product is taken over links on the boundaries of clusters, i.e., links connecting different clusters.

It is useful to consider the set of all configurations $S'=gS$ obtained from a given S and C with all possible transformations $g\in G_C$. We can view these as configurations on a hyperlattice: The clusters in C are the sites, while the group elements $g_k\in G_k$ acting on the kth cluster are the dynamical variables at the corresponding sites of the hyperlattice. The energy $E_{S,C}(g)$ of the configuration g is defined by

$$A(S',C)=A(gS,C)\equiv e^{-E_{S,C}(g)}. \quad (7)$$

Now we let this system evolve according to some transition probability $w_{S,C}(g\rightarrow g')$, satisfying detailed balance:

$$e^{-E_{S,C}(g)}w_{S,C}(g\rightarrow g')=e^{-E_{S,C}(g')}w_{S,C}(g'\rightarrow g). \quad (8)$$

Making one or a few sweeps on the hyperlattice, one obtains a hyperlattice configuration g and, consequently, a new configuration $S'=gS$ on the original lattice. Detailed balance is clearly satisfied. Equations (6)-(8) present the general result of this paper.

Before proceeding, let us make a special choice of the bond probabilities:

$$p_l(E_l)=p(E_l)=\begin{cases}1-e^{E_l-E_0}, & \text{if } E_l<E_0;\\ 0, & \text{if } E_l\geq E_0,\end{cases} \quad (9)$$

with E_0 as a free parameter. Since $p_l(E_l)$ decreases with E_l, the smaller the energy of the link, the larger the probability that the two corresponding sites belong to the same cluster. Thus the structure of the clusters are expected to be similar to that of the real clusters in the system.

For $E_0\leq E_{\min}\equiv\min_S E_l(S)$, the bond probability is zero, and each cluster consists only of a single site. This is just the standard local updating procedure.

With increasing E_0 the average size of the clusters grows. For $E_0\geq E_{\max}\equiv\max_S E_l(S)$ all factors cancel in Eq. (6). It means that in this case there is no interaction between the clusters (or the hyperlattice temperature is infinite), $E_{S,C}(g)=0$, and one can choose the new configuration g at random. [Note that for the Ising model the choice $E_0=J=E_{\max}$ gives $p=1-\exp(-2J)$ for parallel spins and $p=0$ for antiparallel ones, which are the probabilities used by SW.] For $E_0\rightarrow+\infty$ the whole lattice becomes one single cluster, and the relaxation time goes to infinity. Clearly, there exists an optimal intermediate value of E_0, minimizing the relaxation time. As mentioned before, in the case of Ising and Potts models the choice $E_0=E_{\max}$ leads to a strong reduction of the dynamical critical exponent.[16] The reason that this simple choice works there is the following. At $J=J_c$ for these models the value $E_0=E_{\max}$ is the percolation transition point, where infinite clusters start to appear, as will be explained below.

In general, one should perhaps choose E_0 such that the corresponding size of the clusters matches the correlation length in the problem. For most systems this requirement yields a value $E_0<E_{\max}$. At $E_0=E_{\max}$ in general the largest cluster occupies almost the whole lattice, resulting in a large relaxation time. We have seen this in simulations with the Z_N (or vector Potts) models, but it also follows from more general considerations. Indeed, making an arbitrary rotation on a large cluster in an equilibrium configuration in the case of a generic system one obtains a nontypical configuration having a large energy density on the surface of the cluster. The large surface energy is avoided only when the largest cluster contains almost all sites of the lattice. Consequently, the optimal value of E_0 will be smaller, and the clusters will interact, unlike in the case of $E_0=E_{\max}$. In this situation one can use a standard MC method on the hyperlattice with acceptance probabilities $w_{S,C}(g\rightarrow g')$ satisfying Eq. (8). Since large correlation length is incorporated in the average cluster size, we expect that the system on the superlattice is not critical, and a few sweeps will suffice. Actually, even one update of a single extended cluster takes one far away from the previous configuration on the original lattice, into a new configuration which takes a long time to reach by local updates. Our simulations done with the Metropolis method have shown that even for large hit size the corresponding acceptance rates are not small. The reason for this is that the cluster configuration is governed by the "spin" configuration itself, rather than given *a priori*. Hence we expect in general a considerable decrease of relaxation times for large lattices near criticality.

We note here that the choice (9) is not suited for heat-bath updating of clusters. If we choose $p(E_l)=1-\exp[\alpha(E_l-E_{\max})]$, $0\leq\alpha\leq 1$, however, the effective interaction between the clusters becomes $E_{S,C}(g)=(1-\alpha)\sum_{l\in\partial C}E_l(gS)$. For the XY model this is of the same form as the original energy $E(S)$ and it is no more difficult to use heat-bath updating for clusters than for the standard case.

What really matters is not the average size of clusters but the cluster size distribution. To reach the shortest relaxation time, the system might have to be updated on many different scales. Although a given cluster configuration will have clusters of different sizes, a given bond probability function may not produce the optimal distribution. However, one is allowed to vary the form of this function from sweep to sweep—e.g., one can choose a new value (perhaps at random) for the parameter E_0 in each sweep. It is even possible to take different random values for different links.

The cluster updating method provides another advantage when correlations are measured, since the large background of accidental coincidences can be eliminated completely.[18] For the average of some functional $f(S)$ we have

$$\langle f\rangle = \sum_{S'} P_0(S') f(S') = \sum_{S,C} P_0(S,C)\langle f\rangle_{S,C}\,, \quad (10)$$

where

$$P_0(S,C) \equiv P_0(S) w(S \to C) \quad (11)$$

is the joint equilibrium distribution of S and C. Further, the corresponding conditional average of f is denoted by

$$\langle f\rangle_{S,C} \equiv \sum_{S'} w(S,C \to S') f(S')\,. \quad (12)$$

Here the sum is taken over all possible configurations $S' = gS$ which can be reached from a given S and C. The improved estimator $\langle f\rangle_{S,C}$ is feasible to compute only when the clusters are decoupled, i.e., for $E_0 = E_{\max}$ which is, in general, a bad value for the relaxation time. However, one is allowed to update with the optimal E_0 and then measure with $E_0 = E_{\max}$ using the improved estimator. In some cases the smaller statistical error may compensate the time needed for an extra clusterization. As an example, for the spin-spin correlation function in the Ising model for the special choice $E_0 = J$, we have

$$\langle s_i s_j\rangle = \sum_S P_0(S) s_i s_j = \sum_C P_0(C)\delta_{ij}(C)\,, \quad (13)$$

where $P_0(C)$ is the cluster distribution probability, and $\delta_{ij}(C)$ is unity if the two sites are in the same cluster and zero otherwise. [For the Ising and Potts models, the possible set of S' configurations depends only on C, but not on S, and all spins in a given cluster are in the same state, $\langle s_i' s_j'\rangle_{S,C} = \delta_{ij}(C)$.] Hence the correlation function can be measured simply by counting how many times the two sites are in the same cluster. With $\langle s_i s_j\rangle \equiv \rho \ll 1$ one has for the dispersion in the standard case $D(s_i s_j) = 1 - \rho^2$, while with the cluster updating method $D(s_i s_j) = \rho - \rho^2$. Therefore, in our case one needs $N = O(1/\epsilon^2\rho)$ measurements to reach a relative accuracy ϵ, while with the local algorithm a larger number $N = O(1/\epsilon^2\rho^2)$ is necessary.

For the susceptibility one obtains a simple expression (with $E_0 = J$)

$$\chi = \frac{1}{V}\sum_{ij}\langle s_i s_j\rangle = \left\langle \frac{1}{V}\sum_k v_k^2 \right\rangle \equiv \bar{v}\,, \quad (14)$$

where v_k denotes the size of the kth cluster, $V = L^d$ is the total number of sites, and $\bar{v}$ is the weighted average cluster size which can vary from 1 to V. This quantity is proportional to different powers of L in different regions: for $J < J_c$ ($T < T_c$) one has $\bar{v} \propto L^0$, for $J > J_c$: $\bar{v} \propto L^d$, while for $J = J_c$ the correlation function is $\langle s_i s_j\rangle \propto 1/r^{d-2+\eta}$, hence $\bar{v} \propto L^{2-\eta}$, i.e., the Hausdorff dimension of the clusters is $D = 2 - \eta$ at the critical point (for $E_0 = J = J_c$). Since $\bar{v}$ grows with E_0 for fixed J, Eq. (14) shows that the percolation transition defined by the scaling behavior of $\bar{v}$ is indeed at $E_0 = J_c$ for the Ising model. (Obviously, this statement is true in any dimension. Note that our clusters are smaller than clusters formed by all parallel spins which for $d > 2$ start to percolate already in the disordered phase.)

To see how this method works for models with continuous variables, we made some tests on the two-dimensional XY model. At the critical point $J_c \approx 1.12$ the system undergoes a Kosterlitz-Thouless transition.[19]

The results are shown in Table I. Here χ is the susceptibility per lattice site, and τ_χ is its relaxation time defined through the integral of the corresponding normalized autocorrelation function. The standard form $E_l = -J \times \cos(\theta_i - \theta_j)$ has been used. The hit size in the Metropolis updating method was $|\Delta\theta| < \pi/3$. The acceptance rates were measured separately for different cluster sizes. Even for large clusters, containing about 100 sites, reasonable (≥ 0.3) acceptance rates were found. The average cluster size $\bar{v}$ [cf. Eq. (14)] is also given in Table I.

It should be noted here that not only was the number of sweeps needed to generate an independent configuration less but also that the computer time for one sweep was shorter. This happens in spite of the extra complexity of the algorithm, since the number of degrees of freedom to be updated in one sweep is much less than in the standard case. A natural alternative definition of the relaxation time is to measure it in units

TABLE I. The susceptibility per lattice site χ and the corresponding relaxation time τ_χ measured in sweeps, in the two-dimensional XY model for 16^2 and 32^2 lattices. $\bar{v}$ denotes the average cluster size [cf. Eq. (14)] while $\gamma = \langle n_{\rm cl}\rangle/V$ is the number of degrees of freedom updated in one sweep relative to the standard algorithm.

L	J	E_0	χ	τ_χ	$\bar{v}$	γ
16	0.9	local	69(3)	24(2)	1	1
		−0.05	66(2)	14(1)	28.8	0.26
		−0.10	67(2)	8(1)	19.1	0.30
		−0.15	66(2)	12(1)	13.1	0.34
32	1.0	local	327(9)	120(9)	1	1
		−0.10	327(8)	22(2)	81.7	0.20
		−0.15	310(6)	21(1)	43.3	0.23
		−0.20	311(8)	23(2)	25.3	0.27

where the same number of degrees of freedom is updated as in one sweep with the local algorithm (i.e., $V=L^d$). This is obtained by multiplication of τ_χ in Table I by $\langle n_{cl}\rangle/V$ where $\langle n_{cl}\rangle$ is the average number of clusters. This factor is also given in Table I. The results show a considerable decrease in relaxation time, even with the conservative definition. Clearly, much more data would be necessary to determine the dynamical critical exponent.

Finally, a few comments are in order.

(a) The method we introduced is quite flexible and its potential is not exhausted by the type of calculation presented above. In principle, one should be able to find a best strategy within the given framework.

(b) Although not directly applicable in this form, the method can be extended to lattice gauge theories as well. This work is in progress.

(c) Several papers on generalizations of the SW algorithm have appeared in recent months.[20-22] The present method has the advantage that the average cluster size can be tuned to minimize the relaxation time. Although the method proposed by Wolff[21] has such a freedom, the original action is modified there.

The author thanks Z. Rácz for patient explanation of some questions in critical dynamics and extensive conversations, and A. Patkós for a discussion on the planar model.

[1]H. Yahata and M. Suzuki, J. Phys. Soc. Jpn. **27**, 1421 (1969).

[2]Z. Rácz and M. F. Collins, Phys. Rev. B **13**, 3074 (1976).

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

[4]G. F. Mazenko, M. J. Nolan, and O. T. Valls, Phys. Rev. Lett. **41**, 500 (1978).

[5]C. Kalle, J. Phys. A **17**, L801 (1984).

[6]M. Aydin and M. C. Yalabik, J. Phys. A **17**, 2531 (1984).

[7]*Monte Carlo Studies in Statistical Physics,* edited by K. Binder (Springer-Verlag, Berlin,1979).

[8]M. H. Kalos, Brookhaven Conference on Monte Carlo Methods and Future Computer Architectures, May 1983 (unpublished).

[9]K. E. Schmidt, Phys. Rev. Lett. **51**, 2175 (1983).

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

[11]G. G. Batrouni, G. Katz, A. Kronfeld, G. Lepage, S. Svetitsky, and K. Wilson, Phys. Rev. D **32**, 2736 (1986).

[12]J. Kogut, Nucl. Phys. **B257 [FS17]**, 1 (1986).

[13]E. Dagotto and J. Kogut, Nucl. Phys. **B290 [FS20]**, 451 (1987).

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

[15]J. Goodman and A. D. Sokal, Phys. Rev. Lett. **56**, 1015 (1986).

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

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

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

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

[20]R. E. Edwards and A. D. Sokal, to be published.

[21]U. Wolff, Phys. Rev. Lett. **60**, 1461 (1988).

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