Worm Algorithms for Classical Statistical Models

Nikolay Prokof'ev[1] and Boris Svistunov[1,2]

[1]*Department of Physics, University of Massachusetts, Amherst, Massachusetts 01003*
[2]*Russian Research Center "Kurchatov Institute", 123182 Moscow, Russia*

(Received 6 March 2001; published 27 September 2001)

We show that high-temperature expansions provide a basis for the novel approach to efficient Monte Carlo simulations. "Worm" algorithms utilize the idea of updating closed-path configurations (produced by high-temperature expansions) through the motion of end points of a disconnected path. An amazing result is that local, Metropolis-type schemes using this approach appear to have dynamical critical exponents close to zero (i.e., their efficiency is comparable to the best cluster methods) as proved by finite-size scaling of the autocorrelation time for various universality classes.

DOI: 10.1103/PhysRevLett.87.160601 PACS numbers: 05.70.Jk, 64.60.Ht, 75.10.Hk, 75.40.Mg

The Metropolis scheme [1] is usually the most universal and easy to program approach to Monte Carlo (MC) simulations. However, its advantages are much reduced near critical points. It is believed that local [2] Metropolis updates connecting configurations by a Markovian chain are inefficient at the critical point because their autocorrelation time, τ, scales as L^z, where L is the system linear dimension and z is the dynamical critical exponent which is close to 2 in most systems [3,4].

An enormous acceleration of simulations at the critical point has been achieved with the introduction of cluster algorithms by Swendsen and Wang [5]. However, the original method and its developments (both classical and quantum) [6–8] are essentially nonlocal schemes, and we are not aware of any exception to this rule.

In this Letter we propose a method which essentially eliminates critical slowing down and yet remains local. The cornerstone of our approach is to introduce a configuration space of closed paths to represent statistics of the model. Closed-path (CP) configurations may then be sampled efficiently using the worm algorithm (WA) introduced in Ref. [9] for quantum statistical models in which closed trajectories arise from imaginary-time evolution of world lines. In classical models the CP representation derives from high-temperature expansions for a broad class of lattice models (see, e.g., Ref. [10]). In two dimensions (2D), another family of WA may be introduced by considering domain-wall boundaries as closed paths.

We note that our approach is based on principles which differ radically from cluster methods and, most probably, has a different range of applicability. For example, the CP representation is most suitable for the study of superfluid models by having direct Monte Carlo estimators for the superfluid stiffness (through the histogram of winding numbers [11]) which are not available in the standard site representation.

In what follows we first recall how high-temperature expansions work by employing the Ising model as an example (while keeping notation as general as possible). We then explain how WA is used to update the CP configuration space. Next, we discuss specific implementations of WA for $|\psi|^4$, XY, and $q = 3$ Potts models, and comment on the special property of 2D models which allows an alternative CP parametrization of the configuration space. The efficiency of the new method is studied by looking at energy autocorrelation times for six different universality classes. It is found that for 2D and 3D Ising models, 2D and 3D XY models, and Gaussian model the $\tau(L)$ scaling is consistent with the relation $\tau(L) = \tau_0 + c\ln(L)$, i.e., the dynamic critical exponent is close to zero. For the two-dimensional $q = 3$ Potts model our data are best fit to a power law with $z \approx 0.55$, i.e., in all cases the value of z formally satisfies the Li-Sokal [12] bound $z > \alpha/\nu$ (α and ν are the specific heat and correlation length critical exponents) derived for the Swendsen-Wang [5] algorithm.

Since high-temperature expansions for various models can be found in standard texts (see, e.g., Ref. [10]), we briefly outline the procedure for the Ising model

$$-\frac{H}{T} = \sum_{b=\langle ij\rangle} H_b, \qquad H_b = K s_i s_j, \tag{1}$$

where $K = J/T$ is the dimensionless coupling parameter between spin variables $s_i = \pm 1$ and index $b = \langle ij\rangle$ refers to the simple cubic/square lattice bonds between nearest-neighbor sites i and j (we understand site indices as vectors). We will also use another (identical) notation to label bonds by specifying the lattice site and the direction towards its nearest neighbor, ν, i.e., $b = (i, \nu)$. Since H is a sum of bond Hamiltonians, the corresponding Gibbs weight factorizes in terms of exponentials for each bond. By expanding each exponential in a Taylor series, we arrive at an expansion in powers of K (or inverse temperature), $e^{-H/T} = \prod_b \left(\sum_{N_b=0}^{\infty} K^{N_b} (s_i s_j)^{N_b}/N_b!\right)$. In what follows we will consider summation indices $\{N_b\}$ as integer variables $N_b = 0, 1, 2, \ldots$ living on bonds, and call them the "bond state," or the "bond configuration." For each configuration of bonds, $\{N_b\}$, the summation of $e^{-H/T}$ over spin variables factorizes, $\prod_i \left(\sum_{s_i=\pm 1} s_i^{k_i}\right) = \prod_i Q(k_i)$, where $k_i = \sum_\nu N_{i,\nu}$ is the sum over bond states incident on site i. This allows us to sum out the spin variables and

reformulate the problem in terms of bond states [10]. The partition function, for example, takes the form

$$Z \equiv \sum_{\{s_i\}} \sum_{\{N_b\}} \prod_b \frac{K^{N_b}}{N_b!} (s_i s_j)^{N_b} = \sum_{\{N_b\} \in \mathrm{CP}} W_Z(\{N_b\}) . \quad (2)$$

Graphically, each bond state can be represented by lines ascribed to bonds, the number of lines on the bond being equal to N_b. The closed-path constraint $\{N_b\} \in \mathrm{CP}$ is imposed by symmetry: $Q(k)$ is nonzero only if the number of bond lines connected to the site i is even. The summation includes all possible CPs, both connected and disconnected, with self-intersections and overlaps. The bond configuration weight is given by

$$W_Z(\{N_b\}) = \left(\prod_b \frac{K^{N_b}}{N_b!} \right) \left(\prod_i Q(k_i) \right), \quad (3)$$

and for the Ising model we simply have $Q(k) \equiv 2$.

The spin-spin correlation function $G(i_1 - i_2) = \langle s_{i_1} s_{i_2} \rangle$ by definition equals $g(i_1 - i_2)/Z$, where $g(i_1 - i_2) = \sum_{\{s_i\}} s_{i_1} s_{i_2} e^{-H/T}$. Using the same trick with the expansion of $e^{-H/T}$ and summation over spin variables, we arrive at bond configurations for g which differ from those contributing to the partition function in only one respect: there are two special sites i_1 and i_2 with an odd number of bond lines connected to them (if $i_1 \neq i_2$); see Fig. 1. These points are the only places where the path may start and end. We will abbreviate the configuration space for the g function as P_g: it includes the set of all bond configurations that *either* have an even number of bonds incident on every site *or* have an even number of bonds incident on all but two sites. The P_g-configuration weights W_g are given by the same expression (3), provided $k_i = \delta_{i,i_1} + \delta_{i,i_2} + \sum_\nu N_{i,\nu}$.

The above consideration is a perfect setup for the Monte Carlo simulation of a finite-size cluster [13]. The key point is that configurations for Z and $g(0)$ have identical bond elements since CP is an $i_1 = i_2$ subset of P_g. Therefore, the (non-normalized) statistics for g and Z can be accumulated *within the same MC process,* which means that the (unknown) normalizing factor is *the same* for both g and Z, and thus irrelevant to their ratio $G = g/Z$.

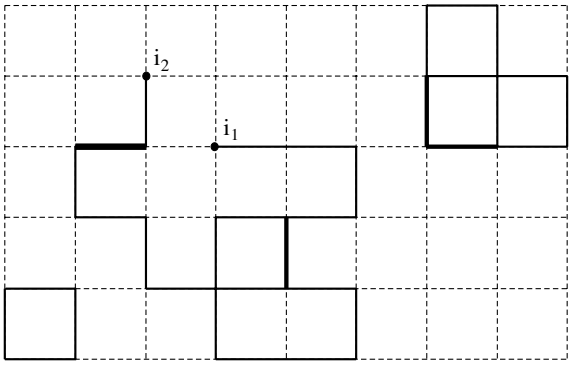

FIG. 1. A typical bond configuration for the correlation function $g(i_1 - i_2)$. The two circles correspond to the points i_1 and i_2, and the solid line width is proportional to the bond state number N_b (number of elementary lines); the $N_b = 0$ terms are shown by dashed lines.

Suppose that bond configurations $\{N_b\}$ (or diagrams, Fig. 1) are generated by some MC process with probabilities proportional to their weights. Then, by standard Metropolis rules, the diagram with $i_1 - i_2 = i$ contributes unity to the stochastic MC sum for $g(i)$. If $i_2 = i_1$ then the closed-path P_g diagram contributing unity to $g(0)$ has a weight $W_g = [Q(k_{i_1})/Q(k_{i_1} - 2)]W_Z$, i.e., it can be easily weighted analytically against the CP diagram contributing to the partition function [the $g(0)$ correlator has two extra powers of spin operators on site i_1, and this is accounted for by the ratio of Q factors]. Thus, this diagram also adds $Q(k_{i_1} - 2)/Q(k_{i_1})$ to the stochastic sum for Z. It means, that MC estimators for $g(i)$ and Z defined on P_g are

$$g(i \,|\, P_g) = \delta_{i_2 - i_1, i}, \quad Z(P_g) = \frac{Q(k_{i_1} - 2)}{Q(k_{i_1})} \delta_{i_2, i_1} . \quad (4)$$

The ratio $g(i)/Z$ yields the physical result for $G(i)$. The spin susceptibility $\chi L^d = \langle (\sum_i s_i)^2 \rangle = \langle \sum_{ij} s_i s_j \rangle$ is given by $\chi = \sum_i g(i)/Z$. The bond energy may be computed in two ways: (i) as the nearest-neighbor sum $E = (L^d/2) K \sum_{|i|=1} g(i)/Z$, and (ii) using direct estimator, $E = \mathcal{E}/Z$, which is nonzero only for the CP configurations contributing to Z:

$$\mathcal{E}(P_g) = \frac{Q(k_{i_1} - 2)}{Q(k_{i_1})} \delta_{i_1 i_2} \sum_b N_b , \quad (5)$$

where the first factor accounts for reweighting between P_g and CP, and the last factor follows from $\sum_{\{s_i\}} \sum_b H_b e^{-H/T} \equiv dZ/d \ln(K)$. Our results for τ were obtained using this estimator.

WA emerges as an idea to update P_g configurations shown in Fig. 1 through the space motion of the end points i_1 and i_2 only [9]. One may consider the end point i_1 as if it is the tip of a "magic pen": when i_1 is shifted to the neighboring site, the pen "draws" a new or "erases" an existing bond line (increases or decreases the bond number by 1; when $i_1 = i_2$, one is free to start drawing/erasing bond lines from any site. The algorithm thus employs two elementary updates which are proposed in the context of a given configuration: if $i_1 = i_2$, then with probability p_0 we propose to "move" both end points to site i_0 selected at random among all L^d lattice sites, and with probability $p_1 = 1 - p_0$ to "shift" the end point i_1 to one of the neighboring sites by selecting the direction of the shift at random; if $i_1 \neq i_2$ we always propose to shift i_1. Formally, $p_0 \in (0, 1)$ and arbitrary otherwise; our calculations were done with $p_0 = 0.5$.

The move-update satisfies detailed balance and its acceptance ratio is unity for the Ising model; in a general case, it involves the ratio of Q functions,

$$P_{\mathrm{mv}}(i \to j) = \frac{Q(k_j + 2)}{Q(k_j)} \frac{Q(k_i - 2)}{Q(k_i)} . \quad (6)$$

We recall that the end point is associated with the extra s_i variable on this site (before summation over spins), and

moving two end points from i_1 to i_0 changes k_i variables on these sites.

Shift-updates change the position of the end point i_1 and simultaneously increase/decrease the bond state by 1. We select which way to proceed with probability 1/2. The shifts $i \to j$ with $N_{b=\langle ij\rangle} \to N_b + 1$ are balanced by shifts $j \to i$ with $N_b \to N_b - 1$, and the solution of the standard balance equation [1] is

$$P_{\rm sh}(i \to j, N_b \to N_b + 1) = r\,\frac{K}{(N_b + 1)}\,\frac{Q(k_j + 2)}{Q(k_j)}, \tag{7}$$

$$P_{\rm sh}(i \to j, N_b \to N_b - 1) = r\,\frac{N_b}{K}\,\frac{Q(k_i - 2)}{Q(k_i)}, \tag{8}$$

where

$$r = \begin{cases} 1/(2p_1) & \text{if } i_1 = i_2 = i\,, \\ 2p_1 & \text{if } i_1 = i,\ i_2 = j\,, \\ 1 & \text{otherwise.} \end{cases} \tag{9}$$

The practical implementation of the algorithm is as follows. One starts from an arbitrary bond configuration, e.g., from $N_b = 0$ for each b, and then performs a set of procedures which forms a standard local Metropolis update repeated in a cycle: (*i*) Select the update. If $i_1 \neq i_2$ then shift i_1; if $i_1 = i_2$ then either move i_1 and i_2 or shift i_1 as described above. (*ii*) Calculate acceptance probabilities using Eqs. (6)–(9); if the update is accepted, then change the configuration accordingly. (*iii*) Collect statistics for various quantities using MC estimators (4) and (5) and go to step (*i*). It takes, on average, 5×10^{-7} s of CPU time on a PIII-600 processor to perform the update.

The Ising model has a special property which can be used to enhance the efficiency by truncating the configuration space. Namely, by using identity $e^{Ks_is_j} \equiv \cosh(K) \cdot [1 + \tanh(K) s_i s_j]$, one can restrict the bond summation to $N_b = 0, 1$ only. The corresponding modifications of the scheme are straightforward.

We now turn to the $|\psi|^4$ model (ψ_i is a complex variable):

$$-\frac{H}{T} = \sum_{\langle ij\rangle} (\psi_i \psi_j^* + \text{c.c.}) + \sum_i [\mu |\psi_i|^2 - U |\psi_i|^4], \tag{10}$$

the two limiting cases of which are the Gaussian model, $U = 0$, and the *XY* model, $U = \infty$. The procedure of factorizing the partition function expansion in terms of the bond Hamiltonian, $\sum_{b=\langle ij\rangle} (\psi_i \psi_j^* + \text{c.c.})$, is exactly as before [the second term in (10) is kept in the exponent]. The only new ingredient is that now for each bond we have two *different* terms, $\psi_i \psi_j^*$ and $\psi_i^* \psi_j$, and therefore the expansion has to be performed for each of them separately. Graphically, this can be captured by drawing lines with arrows, thus specifying each term by the arrow direction. Correspondingly, the bond state is defined by two numbers $(N_b^{(1)}, N_b^{(2)})$ which tell how many lines go along this bond in each of the two directions. All integrals $\int d\psi\, e^{\mu|\psi|^2 - U|\psi|^4} \psi^m (\psi^*)^{m'} = \delta_{m,m'} Q(k = 2m)$ are easily tabulated prior to the simulation. Further discussion of the algorithm and its application to $|\psi|^4$ and Potts models can be found in Ref. [14]. Here we simply mention an estimator for the winding number of the closed-path P_g configuration (which has exactly the same meaning as in the world line quantum Monte Carlo [11])

$$M_\alpha = \frac{Q(k_{i_1} - 2)}{Q(k_{i_1}) L}\, \delta_{i_2, i_1} \sum_{b=(i,\alpha)} (N_b^{(1)} - N_b^{(2)}), \tag{11}$$

where $\alpha = x, y, z, \ldots$.

We performed extensive tests of our algorithm by comparing to known results. In particular, we reproduced the exact answer for $\chi(L = 4)$ for the 3D Ising model [15] to five significant digits, critical points of the 3D *XY* model [16] (to five significant digits using finite-size scaling of magnetic susceptibility up to $L = 160$), $|\psi|^4$ models at several couplings [17] [to 4–5 digits (see Ref. [18])], E and χ for various system sizes for the $q = 3$ Potts model in 2D [12] (to four digits).

In Fig. 2 we present data for the integrated energy autocorrelation time measured in MC sweeps (one sweep = dL^d updates). In all cases, except the $q = 3$ Potts model in 2D, they are best fit to the linear in $\ln L$ behavior. We may not exclude that logarithmic scaling at some larger L crosses over to the power law, $\tau(L) \sim L^z$, with small z. To get an estimate (upper bound) for the possible dynamic exponent we mention in the figure caption, the slope $z(L) = d\ln(\tau)/d\ln(L)$ at the largest simulated L. For the $q = 3$ Potts model our data scale as $\tau \sim L^{0.55}$ for $L > 64$, although the dynamic exponent is showing a systematic decrease with L. This has to be compared with the value $z = 0.515$ for the Swendsen-Wang algorithm [12]. Our data for z do not contradict the Li-Sokal bound $z > \alpha/\nu$ because for all models discussed here α/ν is very small, with the exception of the $q = 3$ Potts model where $\alpha/\nu = 0.4$. Unlike other local schemes, the autocorrelation time for magnetic susceptibility and winding numbers was found to be shorter than τ for energy.

High-temperature expansions are *not* the only procedures to arrive at the CP configuration space and WA. For example, in lattice models with discrete site variables, one can unambiguously specify the state (up to symmetry transformations) by drawing domain boundaries, which, in 2D, may be considered a CP configuration. For the Ising model the configuration weight is simply $W(\text{CP}) = \prod_b e^{K(2N_b - 1)}$, where the bond state takes values $N_{b=\langle ij\rangle} = 0, 1$ ($N_{b=\langle ij\rangle} = 0$ if sites i and j belong to the same domain, and $N_b = 1$ otherwise). To implement WA for efficient sampling of domain boundaries we formally enlarge the configuration space to include nonphysical g-type configurations of domain walls with two end points and proceed exactly as discussed above. The only difference is that now "open" boundaries with $i_1 \neq i_2$ have *no* physical meaning and serve just for updating purposes. Our results for the 2D Ising model in Fig. 2 were obtained using domain-wall representation.

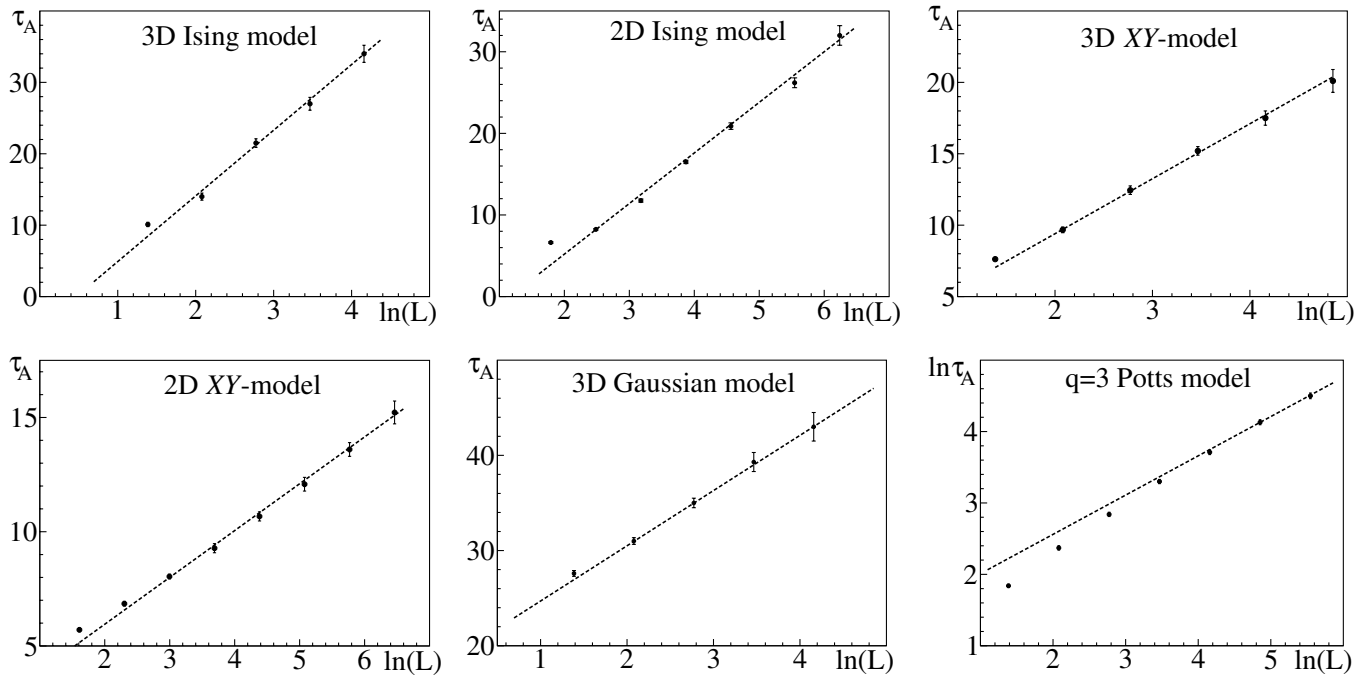

FIG. 2. Autocorrelation times for various universality classes. The 3D Ising model is fitted to $\tau = -4.3 + 9.2\ln(L)$, and $z(L = 64) = 0.18$ (see the text). The 2D Ising model is fitted to $\tau = -7.2 + 6.2\ln(L)$, and $z(L = 512) = 0.25$. The 3D XY model is fitted to $\tau = 1.7 + 3.85\ln(L)$, and $z(L = 128) = 0.2$. The 2D XY model is fitted to $\tau = 1.85 + 2.05\ln(L)$, and $z(L = 640) = 0.16$. The 3D Gaussian model is fitted to $\tau = 18.9 + 5.8\ln(L)$, and $z(L = 64) = 0.17$. The $q = 3$ Potts model in 2D is fitted to the power law $\tau = 4.3L^{0.55}$.

We thank J. Machta and A. Sokal for valuable discussions. This work was supported by the National Science Foundation under Grant No. DMR-0071767. B. S. acknowledges support from the Russian Foundation for Basic Research under Grant No. 01-02-16508.

[1] N. Metropolis *et al.,* J. Chem. Phys. **21**, 1087 (1953).
[2] Updates are local if they address configuration elements and change them without knowing about other elements outside the immediate neighborhood of the updated part.
[3] H. Yahata and M. Suzuki, J. Phys. Soc. Jpn. **27**, 1421 (1969).
[4] K. Kawasaki, in *Phase Transitions and Critical Phenomena,* edited by C. Domb and M. Green (Academic, New York, 1972), Vol. 2; S. K. Ma, *Modern Theory of Critical Phenomena* (Benjamin, Reading, MA, 1976).
[5] R. H. Swendsen and J.-S. Wang, Phys. Rev. Lett. **58**, 86 (1987).
[6] U. Wolff, Phys. Rev. Lett. **62**, 361 (1989).
[7] H. G. Evertz, G. Lana, and M. Marcu, Phys. Rev. Lett. **70**, 875 (1993).
[8] J. Machta *et al.,* Phys. Rev. Lett. **75**, 2792 (1995).
[9] N. V. Prokof'ev, B. V. Svistunov, and I. S. Tupitsyn, Phys. Lett. A **238**, 253 (1998).
[10] G. Parisi, in *Statistical Field Theory,* Frontiers in Physics (Addison-Wesley, Reading, MA, 1988).
[11] E. I. Pollock and D. M. Ceperley, Phys. Rev. B **36**, 8343 (1987).
[12] X.-J. Li and A. D. Sokal, Phys. Rev. Lett. **63**, 827 (1989); J. Salas and A. D. Sokal, J. Stat. Phys. **87**, 1 (1997).
[13] Our scheme contains all diagrams, including *disconnected* ones (see Fig. 1). In a finite system at any finite temperature the series is convergent. This is to be contrasted with standard treatments, where the thermodynamic limit is taken explicitly with simultaneous elimination of disconnected diagrams (see, e.g., [10]): The resulting series is convergent only above the critical point.
[14] N. V. Prokof'ev and B. V. Svistunov, cond-mat/0103146.
[15] H. W. J. Blöte, E. Luijten, and J. R. Heringa, J. Phys. A **28**, 6289 (1995).
[16] H. G. Ballesteros *et al.,* Phys. Lett. B **387**, 125 (1996).
[17] M. Hasenbusch and T. Török, J. Phys. A **32**, 6361 (1999).
[18] V. A. Kashurnikov, N. V. Prokof'ev, and B. V. Svistunov, Phys. Rev. Lett. **87**, 120402 (2001).