Journal of Statistical Physics, Vol. 74, Nos. 1/2, 1994

Optimal Multigrid Algorithms for Calculating Thermodynamic Limits

A. Brandt,[1] M. Galun[1] and D. Ron[1]

Received February 22, 1991; final August 26, 1993

Beyond eliminating the critical slowing down, multigrid algorithms can also eliminate the need to produce many independent fine-grid configurations for averaging out their statistical deviations, by averaging over the many samples produced in coarse grids during the multigrid cycle. Thermodynamic limits can be calculated to accuracy ε in just $O(\varepsilon^{-2})$ computer operations. Examples described in detail and with results of numerical tests are the calculation of the susceptibility, the σ-susceptibility, and the average energy in Gaussian models, and also the determination of the susceptibility and the critical temperature in a two-dimensional Ising spin model. Extension to more advanced models is outlined.

KEY WORDS: Multigrid; Gaussian model; Ising spin model; XY model; Monte Carlo; thermodynamic limit; coarsening by approximation.

1. INTRODUCTION

The aim in statistical physics is to calculate various average properties of configurations governed by the Boltzmann distribution. This is usually done by measuring these averages over a sequence of Monte Carlo iterations. Unfortunately, such processes tend to suffer from several independent inefficiency factors that multiply each other and thus produce very expensive computations.

The best known of these inefficiency factors is the *critical slowing down* (CSD). This is the phenomenon, typical to simulations of critical systems, that with the increase in lattice size there also comes an increase in the number of Monte Carlo passes over the lattice needed to produce a new

[1] Department of Applied Mathematics and Computer Science, Weizmann Institute of Science, Rehovot 76100, Israel. E-mail: MAGALUN@weizmann.weizmann.ac.il

0022-4715/94/0100-0313$07.00/0

configuration which is statistically "useful," i.e., substantially independent of, or only weakly correlated to, a former configuration. Considerable efforts have been devoted to overcome this difficulty. For simple enough cases with real-state variables and at most mild nonlinearities, a general method to eliminate CSD is by classical multigrid methods, properly adapted. (Three different adaptations were introduced in ref. 14, in §7.1 of ref. 6, and in ref. 8. For an introduction to classical multigrid, see Sections 1 and 2 in ref. 5.) For models with severe nonlinearities or discrete variables, such as the ϕ^4 or Ising spin models, a number of publications report on simulation techniques, based on the Swendsen–Wang[18] stochastic clustering technique that partially[8,18] or completely[3,11,12,19] eliminate CSD (see survey[17]). This means that in a work just proportional to the number of gridpoints, a new, substantially independent configuration can be generated.

Optimal as this result is, other, no less important factors of inefficiency still remain intact. To calculate a thermodynamic quantity to a certain accuracy ε, one needs to produce $O(\sigma^2\varepsilon^{-2})$ essentially independent configurations to average out the deviation exhibited by each of them, where σ denotes the standard (i.e., the L_2 average) deviation. Also, the size of the grid must increase as some positive power of ε^{-1}. The main purpose of the present article is to show that multigrid techniques may overcome these additional inefficiency factors as well by introducing more statistical measurements at coarse levels and by other means, such as domain replication. (These techniques were first described in Appendix B of ref. 1.). More directly, what we intend to demonstrate below is that the multigrid structure can be used for measuring meaningful thermodynamic quantities in an *optimal* computational time.

Namely, we will show that thermodynamic limits (quantities obtained at the limit of infinie grids) can be calculated to accuracy ε using only $O(\varepsilon^{-2})$ *computer operations*. This is just the same order of complexity as needed to calculate, by statistical trials, any simple "pointwise" average, such as the frequency of "heads" in coin tossing. This means that in addition to eliminating the CSD factor, multigrid algorithms may also eliminate the "volume factor," which is equal to the total number of sites in the lattice. Both factors *multiply* the statistical factor (ε^{-2}) in the operation count of conventional algorithms.

Stated differently, what we will show is that the multigrid algorithm may effectively produce an independent sample in just $O(1)$ computer operations.

Our prime examples here will be the calculation of the susceptibility, the σ-susceptibility (an approximation to the XY model susceptibility), and the average energy in the Gaussian model, and the susceptibility and the

critical temperature in the two-dimensional Ising model. These cases are ideal for developing, testing, and demonstrating the new multilevel techniques, because of their simplicity and because analytical solutions are known and can be used for comparing the results and understanding the behavior of the numerical processes.

For the one-dimensional Gaussian model it is shown in Section 2 that the susceptibility and the σ-susceptibility (both at the limit of vanishing meshsize or infinite grid), as well as the average energy per degree of freedom (on arbitrarily large grids), can each be calculated to a relative accuracy ε in less than $10\varepsilon^{-2}$ random number generations (independently of the size of the grid). In two dimensions (cf. Section 4.1), the calculation of susceptibility required less than $40\varepsilon^{-2}$ random number generations, even for cases of strongly discontinuous coupling coefficients. It is also shown that the optimal multigrid algorithms for calculating susceptibility and energy cannot be the same; their "cycle index" must differ.

The generalization from the Gaussian to other models, with continuous state but not quadratic Hamiltonians, is not straightforward, but possible. The general approach is outlined in Section 4.3. An important feature is that it may be used for a direct and simple computational derivation of macroscopic dynamics for the model at hand (Section 4.4).

It is not clear whether for discrete models (or continuous models exhibiting a discrete-like behaviour, such as ϕ^4) *optimal* computations of thermodynamic limits [i.e., obtaining accuracy ε in $O(\varepsilon^{-2})$ operations] is always possible. As an example, we discuss in detail calculations with the two-dimensional Ising spin model. It is shown that the configurations produced within one multigrid cycle by stochastic freezes/deletions of the Swendsen–Wang type depend on each other in such a way that not much can be gained by introducing statistical measurements at coarse levels. Nevertheless, the results indicate that it may still be possible to calculate the thermodynamic quantities, such as the critical temperature, in optimal $[O(\varepsilon^{-2})]$ time. Moreover, it is shown that a more sophisticated (three-spin) coarsening, not of the Swendsen–Wang type, produces much less dependence within the multigrid cycle, making it possible to benefit much from making many measurements at its coarse levels.

2. ONE-DIMENSIONAL GAUSSIAN MODEL: FAST CALCULATIONS OF THERMODYNAMIC QUANTITIES

A multigrid algorithm for simple continuous-state models, such as the Gaussian model, has been described by us in ref. 6 and independently by Goodman and Sokal[8] and by Mack.[14] The three approaches are not the same: while Mack and we use linear interpolations, Goodman and

Sokal employ constant interpolation (blocking); and while Goodman and Sokal and we produce independent coarse level dynamics, Mack uses the "unigrid" approach (calculating moves of all scales on the same fine grid). In the present work we show that our multigrid Monte Carlo approach (unlike those of Mack and Goodman–Sokal) can be used not only for eliminating the CSD, but also for accelerated calculation of averages, especially those which depend on large-scale fluctuations. Thermodynamic limits can be calculated to accuracy ε in just $O(\varepsilon^{-2})$ computer operations.

The outline of this section is as follows. In Section 2.1 we use Fourier analysis to calculate in closed form some thermodynamic quantities. Analogous quantities on finite grids are introduced in Section 2.2. In Sections 2.3 and 2.4 we describe an extremely efficient multigrid algorithm for evaluating the discrete susceptibility and approaching its limit for zero meshsize. Some remarks on parallel processing are briefly given in Section 2.5 and numerical tests are reported in Section 2.6. Similar techniques and tests are reported in Section 2.7 for another type of susceptibility, the "σ-susceptibility," designed to approximate the susceptibility of sigma models in the limit of vanishing temperature. The optimal computation of the average energy per degree of freedom is described in Section 2.8. More detailed description of all the derivations and numerical tests can be found in ref. 7.

2.1. Continuous Case

To facilitate theoretical analysis of the algorithms, we treat the constant-coefficient case. But the same algorithms have been used for much more general situations, with a similar efficiency.

The Hamiltonian associated with the continuous case is

$$\mathcal{H}(u) = \int_0^L u_x^2 \, dx \tag{1}$$

where $u = u(x)$ is a function (configuration) defined for $0 \leqslant x \leqslant L$. Homogeneous Dirichlet boundary conditions, $u(0) = u(L) = 0$, are used for definiteness, though others could serve as well. Consequently, a general configuration $u(x)$ can be expanded by

$$u(x) = \sum_{j=1}^{\infty} c_j \sin(j\pi x/L) \tag{2}$$

where the Fourier coefficients c_j are real. By substituting (2) into (1) one gets

$$\mathcal{H}(u) = \frac{\pi^2}{2L} \sum_{j=1}^{\infty} j^2 c_j^2 \tag{3}$$

The magnetization is given by

$$M(u)=\frac{1}{L}\int_0^L u(x)\,dx=\frac{2}{\pi}\sum_{j=1}^{\infty}{}^{*}\frac{c_j}{j} \tag{4}$$

where $\sum^*$, here and below stands for a summation over *odd* integers. As the probability density of each configurations u is given by the density function of the Boltzmann distribution

$$P(u)=e^{-\mathscr{H}(u)/T}/Z(T) \tag{5}$$

straightforward calculations of the average magnetization $\langle M\rangle$, susceptibility $\langle M^2\rangle-\langle M\rangle^2$, and energy $\langle \mathscr{H}\rangle$ can easily be made using the above Fourier expansion, leading to the following results:

$$\langle M\rangle=0 \tag{6a}$$

$$\langle M^2\rangle-\langle M\rangle^2=\langle M^2\rangle=\frac{4LT}{\pi^4}\sum_{j=1}^{\infty}{}^{*}\frac{1}{j^4} \tag{6b}$$

$$\langle \mathscr{H}\rangle=\frac{T}{2}\sum_{j=1}^{\infty}1=\infty \tag{6c}$$

Although the Hamiltonian is not bounded, its differences associated with changing any c_j (or any discrete degree of freedom, such as those defined in Section 2.2 below) are well defined, hence it yields these well-defined statistics.

2.2. Discrete Case

In order to measure such statistical averages numerically, it is necessary to discretize the system. On a grid with meshsize $h=L/N$, the discretized Hamiltonian $\mathscr{H}_h(u)$, approximating (1), can be written as

$$\mathscr{H}_h(u)=\frac{1}{h}\sum_{i=1}^{N}[u(x_i)-u(x_{i-1})]^2 \tag{7}$$

where $x_i=ih$ $(0\leqslant i\leqslant N)$ are the gridpoints. For the simplicity of the multigrid algorithm (see Section 2.3) we have assumed $N=2^k$; the general case could, however, be calculated as well by handling the near-boundary points differently. Assuming again $u(x_0)=u(x_N)=0$, a general grid configuration can be represented by

$$u(x_i)=\sum_{j=1}^{N-1}c_j\sin(j\pi x_i/L) \tag{8}$$

leading together with (7) to

$$\mathscr{H}_h(u)=\frac{2L}{h^2}\sum_{j=1}^{N-1} c_j^2 \sin^2\left(\frac{j\pi h}{2L}\right) \tag{9}$$

The discrete magnetization is given by

$$M_h(u)=\frac{h}{L}\sum_{i=0}^{N} u(x_i)=\frac{h}{L}\sum_{j=1}^{N-1}{}^{*} c_j \frac{\cos[j\pi h/(2L)]}{\sin[j\pi h/(2L)]} \tag{10}$$

By using the probability distribution (5), where $\mathscr{H}_h(u)$ replaces $\mathscr{H}(u)$, one can derive

$$\langle M_h\rangle = 0 \tag{11a}$$

$$\langle M_h^2\rangle = \frac{Th^4}{4L^3}\sum_{j=1}^{N-1}{}^{*} \frac{\cos^2[j\pi h/(2L)]}{\sin^4[j\pi h/(2L)]} \tag{11b}$$

$$\langle \mathscr{H}_h\rangle = T(N-1)/2 \tag{11c}$$

As $N\to\infty$ (with fixed L, hence $h\to 0$), which is called here the thermodynamic limit, the results of the discrete case tend to those of the continuum. $\langle M_h^2\rangle$ exhibits a discretization error of $O(h^2)$. More precisely, Taylor expansion of each term in (11b) and comparison to (6b) yields

$$\langle M_h^2\rangle = \langle M^2\rangle - \frac{Th^2}{3\pi^2 L}\sum_{j=1}^{N/2}{}^{*}\frac{1}{j^2} + TLO(N^{-3}) \tag{11d}$$

where terms with $j>N/2$ are omitted from both $\langle M_h^2\rangle$ and $\langle M^2\rangle$ since they clearly have only $TLO(N^{-3})$ total contribution. From (11d) it follows that the *relative* discretization error $|\langle M_h^2\rangle - \langle M^2\rangle|/\langle M^2\rangle$ is $O(h^2/L^2)$. Clearly, by using a pth-order discretization, this relative error can be further reduced to $O(h^p/L^p)$. As will be explained later, the algorithm will not necessarily actually provide this $O(h^p/L^p)$ accuracy for any given gridsize, as this may turn out to be wasteful when statistical errors are taken into account as well. Instead, it will be constructed so as to keep the number of operations optimal with respect to the overall produced accuracy. The details of the algorithm are given next.

2.3. Description of the Multigrid Cycle

Consider the following generalized Hamiltonian, which includes an additional external magnetic field of density ϕ_i at gridpoint x_i:

$$\mathscr{H}_h(u)=\frac{1}{h}\sum_{i=1}^{N}(u_i-u_{i-1})^2 + h\sum_{i=1}^{N-1}\phi_i u_i \tag{12}$$

where $u_i = u(x_i)$. On the given (the finest) grid, $\phi_i = 0$ is actually prescribed, but the more general form is needed for the algorithm recursion.

The coarse grid with meshsize $H = 2h$ is constructed by taking every other fine-grid point: see Fig. 1. The coarse-grid function $u^H = (u_0^H, ..., u_I^H, ..., u_{N/2}^H)$ describes a *displacement* of the fine-grid function $u^h = (u_0, ..., u_i, ..., u_N)$; i.e., it modifies the latter through *interpolation and addition:*

$$u^h = \tilde{u}^h + I_H^h u^H \tag{13}$$

where $\tilde{u}^h$ is the fine-grid configuration at the stage of switching to the coarse grid and I_H^h denotes interpolation from grid H to grid h (we will use the *linear* interpolation, which is optimal here; see conclusion C in Section 2.4).

The fine-grid Hamiltonian $\mathscr{H}_h(u^h)$ resulting from that interpolation can be written as follows:

$$\mathscr{H}_h(\tilde{u}^h + I_H^h u^H) = \mathscr{H}_h(\tilde{u}^h) + \mathscr{H}_H(u^H) \tag{14}$$

where $\mathscr{H}_h(\tilde{u}^h)$ is given by (12) and $\mathscr{H}_H(u^H)$ is

$$\mathscr{H}_H(u^H) = \frac{1}{H} \sum_{I=1}^{N/2} (u_I^H - u_{I-1}^H)^2 + H \sum_{I=1}^{N/2-1} \phi_I^H u_I^H \tag{15}$$

with

$$\phi_I^H = \frac{-\tilde{u}_{i-2}^h + 2\tilde{u}_i^h - \tilde{u}_{i+2}^h}{2h^2} + \frac{\phi_{i-1}^h + 2\phi_i^h + \phi_{i+1}^h}{4} \qquad \left(I = \frac{i}{2} = 1, ..., \frac{N}{2} - 1\right) \tag{16}$$

representing fine-to-coarse induced field-like terms. These coarse terms are calculated from the details of the fine-grid configuration at coarsening and are fixed throughout the processing on the coarser level. The variables of the coarse grid u_I^H are initially set to zero, corresponding to zero initial displacements.

Notice that, having calculated the field ϕ^H once for all, one calculates $\mathscr{H}_H$ directly in terms of the coarse-grid configuration u^H; there is no need

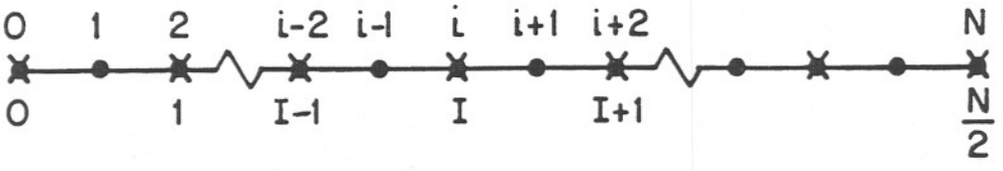

Fig. 1. Fine-grid points are denoted by ● and coarse-grid points by ×.

to perform explicitly (14) in order to compute energy differences associated with changes in u^H. One can therefore run a long Monte Carlo process with $\mathscr{H}_H$ before explicitly updating u^h by (13).

The entire algorithm can be described by a sequence of multigrid cycles for the finest level. A cycle for any given ("current") level is recursively defined by the following five steps.

1. ν_1 Monte Carlo sweeps are first made on the current level. Then, if this level is the coarsest, go to 5.
2. The next coarser level is created from the current one by determining the coarse field-like terms (16).
3. γ multigrid cycles for the coarse level are performed. (γ may change from one cycle of the current level to another in some periodic manner. The *cycle index* is the average value of γ, and for convenience it will also be denoted by γ. Thus γ need not be an integer, and may be smaller than 1.)
4. Update the current level by performing (13).
5. Additional ν_2 Monte Carlo sweeps are finally made on the current level.

The Monte Carlo sweeps are performed by changing each variable in its turn randomly according to its associated distribution, regarding its neighbors as fixed.

The values of ν_1, ν_2, and γ are discussed below.

The described cycle, even with $\gamma = 1$ (which is called a V cycle), would generate a new configuration substantially independent of the precycle one in a work just proportional to the number of gridpoints; it would thus eliminate the critical slowing down. By the term *substantially independent configuration* we mean that the correlation between any quantity of interest in the initial arbitrary configuration and in the one produced after k cycles decays like $e^{-k/\tau}$, where τ, the cycle autocorrelation time, is independent of the lattice size N. In fact, τ is very small, so there is almost no correlation between any quantities before and after even one cycle.

The crucial issue, however, turns out to be different; it is addressed next.

2.4. Fast Sampling of Susceptibility

Any observable which is bounded as $h \to 0$ [e.g., magnetization and susceptibility, but *not* energy; see (6c)] must be dominated by contributions from large-scale fluctuations [low-frequency Fourier components; cf.,

e.g., (6b)]. The main issue of the Monte Carlo simulation is therefore to sample quickly as many such fluctuations as possible. For this purpose a cycle index γ substantially larger than 1 may be used, and averages will be calculated over as many measurements as one could make *within* each cycle, especially at its coarsest level stages.

Consider, for instance, the calculation of $\langle M_h^2 \rangle$. Observe first that M_h can be evaluated on any level. Indeed, denoting by $\tilde{u}^h$ the fine-grid configuration at coarsening, (10) and (13) imply that

$$M_h = \frac{h}{L} \sum_{i=0}^{N} (\tilde{u}^h + I_{2h}^h u^{2h})_i = \frac{h}{L} \sum_{i=0}^{N} \tilde{u}_i^h + \frac{2h}{L} \sum_{I=0}^{N/2} u_I^{2h} \tag{17}$$

Employing this recursively, one gets that when working on level l, if its current configuration is u^{h_l}, then the (fine-grid) magnetization corresponding to it is

$$M_h = \sum_{j=0}^{l-1} \tilde{M}_{h_j} + M_{h_l} \tag{18}$$

where generally

$$h_j = 2^j h, \qquad M_{h_j} = \frac{h_j}{L} \sum_I u_I^{h_j}, \qquad \tilde{M}_{h_j} = \frac{h_j}{L} \sum_I \tilde{u}_I^{h_j}$$

and $\tilde{u}^{h_j}$ is the jth level configuration at the stage of switching to the next coarser level. Thus, many measurements of M_h^2 can be made within a cycle, and their average $\overline{M_h^2}$ can be used as an estimate for the discrete susceptibility $\langle M_h^2 \rangle$. In practice, measurements need be taken only on the coarsest level, and in fact after each relaxation sweep there, because only that is when substantial changes in M_h are introduced.

Let us now estimate the number m_i of relaxation sweeps the algorithm needs to perform on level i, i.e., on the grid with meshsize $h_i = 2^i h$ $[i = 0, 1, \ldots, l = \log_2(N/2)]$, in order to achieve relative accuracy ε in the estimation of the susceptibility. The relaxation sweep on level i strongly affects, hence effectively samples, only those Fourier coefficients c_j [cf. (2)] for which $j = O(L/h_i)$. Hence m_i depends on the contribution of these components to the deviations in measuring $\langle M^2 \rangle$. By (4)

$$M^2 = \frac{4}{\pi^2} \sum_{j,k}^{*} \frac{c_j c_k}{jk} \tag{19}$$

Consider first a term (j, k) in (19) for which both j and k are $O(L/h_i)$, hence the term is effectively sampled $O(m_i)$ times in a cycle. Since the standard deviation of the term is

$$\frac{4}{jk\pi^2} [\langle (c_j c_k)^2 \rangle - \langle c_j c_k \rangle^2]^{1/2} = O(j^{-2} k^{-2} LT) = O(h_i^4 L^{-3} T)$$

the standard deviation of its average over the $O(m_i)$ samples is $O(m_i^{-1/2}h_i^4L^{-3}T)$. There are $O(h_i^{-2}L^2)$ such terms, and their deviations can be considered independent of each other, hence their total contribution is $O(m_i^{-1/2}h_i^3L^{-2}T)$. In case $j=O(L/h_{i-r})$ and $k=O(L/h_i)$, where $r\geqslant 1$ (i.e., $h_i>h_{i-r}$), the term (j,k) in (19) is effectively sampled as follows: in an inner loop, for a (nearly) fixed value of c_j, the value of c_k is averaged $O(m_i/m_{i-r})$ times, yielding an average whose deviation is of the order

$$O\left(\frac{c_j}{jk}\left(\frac{m_i}{m_{i-r}}\right)^{-1/2}\langle c_k^2\rangle^{1/2}\right)=O\left(\frac{c_j}{jk^2}\left(\frac{m_{i-r}}{m_i}LT\right)^{1/2}\right)$$

Then, in an outer loop, the c_j in this average is averaged over $O(m_{i-r})$ samples, giving results with deviations of order

$$O\left(\frac{m_{i-r}^{-1/2}}{jk^2}\left(\frac{m_{i-r}}{m_i}LT\right)^{1/2}\langle c_j^2\rangle^{1/2}\right)=O\left(\frac{m_i^{-1/2}}{j^2k^2}LT\right)$$
$$=O(m_i^{-1/2}h_i^2h_{i-r}^2L^{-3}T)$$

There are $O(h_i^{-1}h_{i-r}^{-1}L^2)$ such terms, effectively independent, hence their total deviation is $O(m_i^{-1/2}h_i^{3/2}h_{i-r}^{3/2}L^{-2}T)$, which, when summed over integers $r\geqslant 0$, gives again $O(m_i^{-1/2}h_i^3L^{-2}T)=\langle M^2\rangle\, O(m_i^{-1/2}h_i^3L^{-3})$. Hence the *total relative* expected error in measuring $\langle M^2\rangle$ is

$$\varepsilon=O\left(\sum_{i=0}^{l}m_i^{-1/2}h_i^3L^{-3}\right)+O(h^pL^{-p}) \tag{20}$$

where the last term added here is the discretization error (cf. Section 2.2). The total work (operations) on all the levels is clearly

$$W=\sum_{i=0}^{l}m_iO(L/h_i) \tag{21}$$

The optimal choice for m_i (yielding either minimal ε for a given W or minimal W for a given ε) is obtained when $\partial\varepsilon/\partial m_i+\lambda_1\,\partial W/\partial m_i=0$, which, by (20) and (21), yields

$$m_i=\lambda_2h_i^{8/3}=\lambda_3 2^{8i/3} \tag{22}$$

where λ_1, λ_2, and λ_3 are independent of i. Relation (22) is realized by the cycle index $\gamma_{\text{opt}}=2^{8/3}\simeq 6.35$.

For *any* fixed cycle index γ we have $m_i=m\gamma^i$, where m is the total number of cycles performed. Since $h_iL^{-1}=O(2^{i-l})$, we can perform the summations in (20) and (21) and obtain

$$\varepsilon=O\left(m^{-1/2}\frac{\gamma^{-l/2}-2^{-3l}}{1-2^{-3}\gamma^{1/2}}\right)+O(2^{-lp}) \tag{23}$$

and

$$W = O\left(m \frac{\gamma^l - 2^l}{1 - 2\gamma^{-1}}\right) \tag{24}$$

for any $2 < \gamma < 2^6$. Actually, by choosing γ and the approximation order p so that γ is significantly smaller than 2^{2p}, the second term in (23) can be ignored, yielding $W = O(\varepsilon^{-2})$ and $m = O(\varepsilon^{-2}\gamma^{-l})$. While $\gamma = \gamma_{\text{opt}}$ indeed minimizes $W\varepsilon^2$, the value of $W\varepsilon^2$ [when the second term in (23) is negligible] is only 15% different from that minimum for any $4 \leqslant \gamma \leqslant 11$. The four main conclusions from this analysis are therefore as follows.

A. *Cycle index* γ. Generally, any $2 < \gamma < \min(64, 2^{2p})$ yields $W = O(\varepsilon^{-2})$. *Asymptotically* (for $\varepsilon \to 0$), the minimal value of $W\varepsilon^2$ is attained for $\gamma = 2^{8/3} \simeq 6.35$ and values very close to the minimum are obtained for $4 \leqslant \gamma \leqslant 11$. *In practice* (for realistic values of ε), the smaller values of γ in this range are better, since for them the influence of the second term in (23) is smaller.

B. *Discretization order p.* There is little advantage in raising the order beyond $p = 2$. It would allow the use of cycles with larger γ, but the dominating coarsest-grid work will remain essentially the same. The only slight advantage may be the smaller storage requirement, which is $O(N) \geqslant O(\varepsilon^{-1/p})$ (cf. Section 2.5).

C. *Interpolation order.* Linear interpolation (second-order interpolation) I_{2h}^h is good enough, as any smooth fluctuation v^h has an approximate configuration $I_{2h}^h u^{2h}$ such that $\tilde{u}^h + v^h$ and $\tilde{u}^h + I_{2h}^h u^{2h}$ have almost the same energy. This means that the probability density function for smooth movements on the coarse grid is nearly the physical one: the coarsening has introduced nearly no statistical bias into such movements. Thus, not much could be gained by using interpolation orders higher than 2 (meaning I_H^h higher than linear). Such higher interpolation order would also make the coarse-grid Hamiltonian substantially more complicated. A relatively little complication, together with almost all the possible gain, can be obtained by using third-order interpolation [based either on the function $x(L-x)$ or on $\sin(\pi x/L)$] only at the transition to the *coarsest* grid (which has only one degree of freedom).

On the other hand, the order of the interpolation operator I_{2h}^h should be *at least* 2, i.e., linear interpolation. This is because the coarser levels should accurately sample all the components slow to change under the current-level Monte Carlo process. It means that every slowly changing configuration u^h must have an approximate configuration of the form $I_{2h}^h u^{2h}$, and the two configurations should have approximately the same energy. The linear interpolation satisfies this requirement. A border case is

the *first-order* (*constant*) *interpolation* I_{2h}^{h}. For any smooth configuration u^h approximated by a function u^{2h} on a coarse grid with meshsize $2h$, the energy of $I_{2h}^{h}u^{2h}$ is about twice that of u^h. Hence, a very smooth component on the finest grid, u^h, that is effectively changed only on a coarse grid with meshsize $2^q h$, will be represented on that coarse grid by an approximation which has an energy 2^q times its own energy. Thus, the changes introduced to the amplitude of such an approximation on that coarse grid will be only $O(2^{-q/2})$ times the typical fluctuations of that amplitude. It follows that roughly 2^q visits to that coarse grid will be needed to accumulate a typical fluctuation. This means that *a cycle index* $\gamma \geqslant 2$ *must be used* to eliminate the CSD. Thus, for two reasons *constant interpolation* is not used. First, the work per W cycle ($\gamma = 2$) is $O(N \log N)$, so one cannot eliminate the CSD. Second, and much more importantly, a whole W cycle is needed to produce a single useful measurement, whereas in the algorithm above each additional movement on the coarsest grid generates another useful measurement.

D. *Number of cycles m. Asymptotically*, to avoid an error ε dominated by the discretization error [the second term in (23)], one should choose $m \leqslant O((16/\gamma)^l)$. Any larger number of cycles would do useless work of reducing the statistical error, because this error is already smaller than the discretization error. *In practice*, for realistic ε, the smallest m possible, i.e., $m = 1$, is preferable for minimizing the influence of the discretization error. This means that whenever the desired accuracy ε is reduced (or the available amount of processing W is increased) the work is increased not by increasing m, but by raising l, i.e., introducing new finer levels (processed very rarely, of course).

In summary, in computing susceptibility one can use second-order discretization and second-order interpolation, any cycle index in the range $4 \leqslant \gamma \leqslant 11$ (with some preference for the lower values), and any number of cycles $1 \leqslant m \leqslant O((16/\gamma)^l)$ (with some preference to $m = 1$): the overall computational work will always be dominated by the $m\gamma^l = O(\varepsilon^{-2})$ work on the coarsest grid, with ε being the relative accuracy that will be obtained in calculating the thermodynamic limit $\langle M^2 \rangle$.

2.5. Parallel Processing

The algorithm described above can use a very high degree of parallel processing. On each grid, the Monte Carlo sweeps can proceed simultaneously at half the gridpoints: first the odd, then the even. The calculation of the coarse-grid functions (16) can be done at all points in parallel. More important, the employment of the γ different coarse-grid cycles can proceed

in parallel to each other, and so can the γ^2 cycles of the still coarser level, etc. So, if an unlimited number [or actually $O(\varepsilon^{-2})$] processors are available, the algorithm can in principle be performed in $O(\log(1/\varepsilon))$ parallel steps. This of course ignores processor communication considerations, but these place only very mild restrictions here; e.g., the γ parallel multigrid cycles created on each coarsening need no communication to each other while running, and only minimal synchronization in collecting their results.

For keeping the total communication, as well as the total storage, at minimum, it may be desired to have as small N (number of points on the finest grid) as possible. For this, a discretization order $p=4$ may be preferred, since $N=O(\varepsilon^{-1/p})$. Still higher-order discretization would not help, since components with $O(h)$ wavelength contribute $O(h^4)$ to $\langle M^2\rangle$, so for an $O(\varepsilon)$ accuracy a grid must be used whose meshsize is no larger than $O(\varepsilon^{1/4})$.

2.6. Numerical Results

We have tested the multigrid algorithm with $\gamma=1$, 2, 3, 4, 6, 12, and 24 on grids of size up to 128. Our main aim was to show that for appropriate values of γ, optimal behavior is achieved; i.e., the average error in the approximation for $\langle M^2\rangle$, (6b), produced by a multigrid cycle is reduced by a factor $\sqrt{\gamma}$ upon using a (finest) grid twice finer (N twice larger), which increases the work by a factor γ. The susceptibility has been measured over just *one* cycle ($m=1$). In that cycle, M_h^2 has been measured, using (18), after each relaxation step on the coarsest level, hence $(\nu_1+\nu_2)\gamma^l$ times altogether. The average of these measurements, $\overline{M_h^2}$, is the approximation for $\langle M_h^2\rangle$, (11b), which in turn is also an approximation for the desired thermodynamic limit $\langle M^2\rangle$. The measured relative error is defined as $\varepsilon=|\overline{M_h^2}-\langle M^2\rangle|/\langle M^2\rangle$. We also define α to be the expected value of $\#\mathrm{RAN}\cdot\varepsilon^2$, where $\#\mathrm{RAN}$ is the amount of work spent per cycle, measured by the number of times a random number is generated, which (for $\gamma>2$) is dominated by $(\nu_1+\nu_2)\gamma^l$, the number of relaxation steps on the coarsest level. Thus, α should turn out constant if indeed the algorithm solves to accuracy ε in $O(\varepsilon^{-2})$ operations. For each value of N, α was estimated by averaging ε over an ensemble of 2000–5000 runs.

In Fig. 2 we present α for $\gamma=1$, 2, 3, 4, 6, 12, and 24 vs. (log) system size N. As expected, for $2<\gamma\leqslant 16$ the value of α tends to a constant as N increases. Also, the cycle with smaller values of γ is slightly more efficient (has smaller α) as expected (see conclusion A in Section 2.4). The graphs for $\gamma=4$, 6 are indistinguishable.

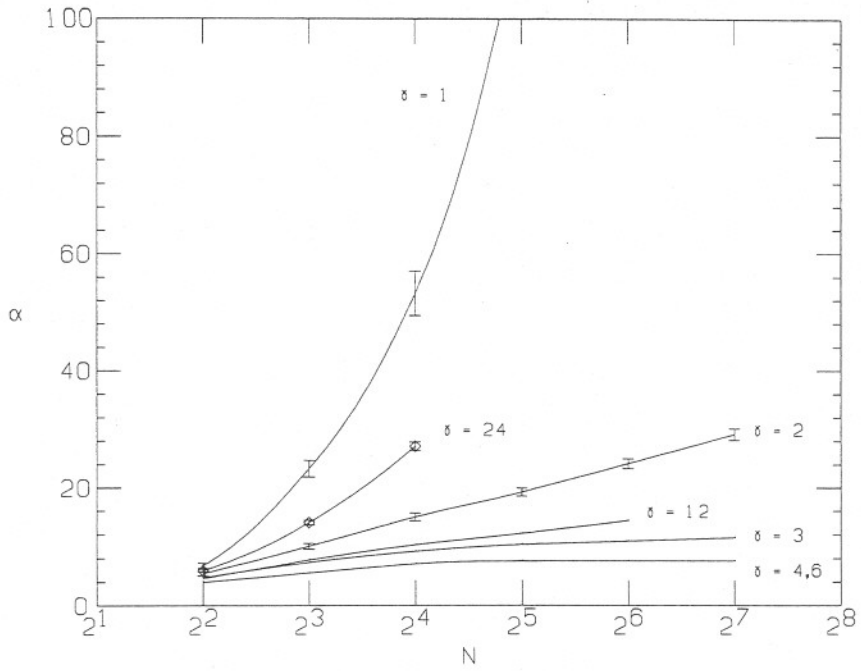

Fig. 2. Performance in measuring susceptibility. Each curve shows α (measuring computational work times the square of the obtained accuracy) as a function of the system size N for the indicated value of the cycle index γ.

2.7. σ-Susceptibility and Its Computation

We introduced a new type of observable, χ_σ, defined in the continuum limit by

$$\chi_\sigma = \left\langle \frac{1}{L}\int_0^L (u(x))^2\,dx - \left(\frac{1}{L}\int_0^L u(x)\,dx\right)^2 \right\rangle$$

and approximated in the discrete model by

$$\chi_\sigma^h = \left\langle \frac{1}{N}\sum_{i=0}^{N} u_i^2 - \left(\frac{1}{N}\sum_{i=0}^{N} u_i\right)^2 \right\rangle$$
$$= \left\langle \frac{1}{N}\sum_{i=0}^{N} u_i^2 \right\rangle - \langle M^2 \rangle$$

The significance of this observable is that it corresponds to the susceptibility of sigma models at the limit of small fluctuations. Indeed, assuming $u_i - \bar{u}$ to be small, where $\bar{u} = (\sum_{j=1}^{N} u_j)/N$, it is easy to see that

$$\frac{1}{N}\left\langle \left(\sum_{i=1}^{N} \cos u_i\right)^2 + \left(\sum_{i=1}^{N} \sin u_i\right)^2 \right\rangle \simeq N - N\chi_\sigma^h$$

the expression on the left-hand side being the susceptibility of the XY model with u_i as the angle parameters. We therefore call χ_σ the σ-susceptibility.

The observables χ_σ and χ_σ^h can be computed analytically as follows. By the Parseval identity and by (4)

$$\chi_\sigma = \left\langle \frac{1}{2} \sum_{j=1}^{\infty} c_j^2 - \left(\frac{2}{\pi} \sum_{j=1}^{\infty *} \frac{c_j}{j} \right)^2 \right\rangle$$

Hence [using the density function of the Boltzmann distribution (5)] the continuous σ-susceptibility is

$$\chi_\sigma = \frac{LT}{2\pi^2} \sum_{j=1}^{\infty} \frac{1}{j^2} - \frac{4LT}{\pi^4} \sum_{j=1}^{\infty *} \frac{1}{j^4} = \frac{LT}{24}$$

Similar calculations in the discrete case lead to

$$\chi_\sigma^h = \left\langle \frac{1}{2} \sum_{j=1}^{N-1} c_j^2 - \left(\frac{h}{L} \sum_{j=1}^{N-1 *} c_j \frac{\cos[j\pi h/(2L)]}{\sin[j\pi h/(2L)]} \right)^2 \right\rangle$$

$$= \frac{Th^2}{8L} \sum_{j=1}^{N-1} \frac{1}{\sin^2[j\pi h/(2L)]} - \frac{Th^4}{4L^3} \sum_{j=1}^{N-1 *} \frac{\cos^2[j\pi h/(2L)]}{\sin^4[j\pi h/(2L)]}$$

We claim that the same multigrid algorithm which is described in Section 2.3 can achieve optimal results in measuring the σ-susceptibility using appropriate cycle index γ. Fourier analysis analogous to that presented

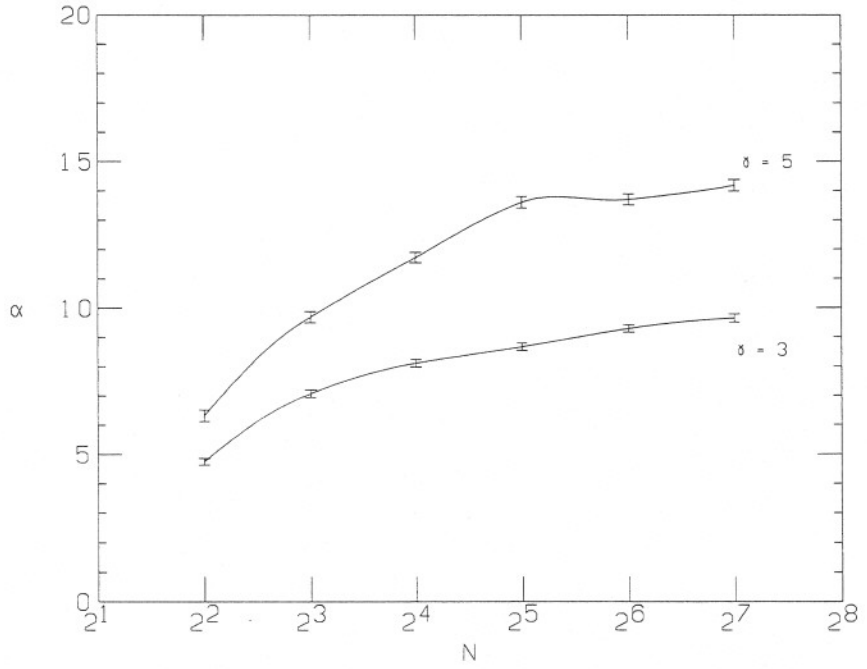

Fig. 3. Performance in measuring σ-susceptibility. Notation as in Fig. 2.

for χ (Section 2.4) shows that any $2<\gamma<8$ yields $W=O(\varepsilon^{-2})$. Numerical experiments for $\gamma=3, 5$ demonstrates that the algorithm solves to accuracy ε in $O(\varepsilon^{-2})$ operations, i.e., α tends to a constant as N increases. See Fig. 3.

2.8. Computing Average Energy per Degree of Freedom

From (6c) and (11c) it is evident that the average energy per degree of freedom is exactly $T/2$ both in the discretization, with any meshsize, and in the continuum limit (where that is exactly the average energy of each Fourier component). We study now the fast Monte Carlo calculation of this quantity.

Whereas the calculation of susceptibility has been shown to be heavily dominated by the coarsest level, the sampling of $\mathscr{H}$ presents the other extreme. Since most Fourier components are substantially affected by relaxation on the finest level, a measurement of $\mathscr{H}$ should be done after each relaxation sweep on that level, and the work should be dominated by the finest-grid sweeps.

Consider the calculation of the average energy on a given *fixed* finest grid, with meshsize h_0 and $N=L/h_0$ intervals. Since $\mathscr{H}(u)=(\pi^2/2L)\sum_{j=1}^{\infty} j^2c_j^2$ [see (3)], in each measurement of $\mathscr{H}$ each Fourier component contributes the following deviation:

$$\frac{j^2\pi^2}{2L}\left[\langle (c_j^2)^2\rangle-\langle c_j^2\rangle^2\right]^{1/2}$$

which is $O(T)$. For any level with meshsize h_i, the number of components with wavelength $O(h_i)$ is $O(h_i^{-1}L)$, and their total deviation in each measurement is therefore $O(h_i^{-1/2}L^{1/2}T)$. If grid h_i is averaged $m_i\geqslant 1$ times, this deviation drops to $O(m_i^{-1/2}h_i^{-1/2}L^{1/2}T)$. To obtain relative accuracy ε, this deviation should be less than $\varepsilon\langle \mathscr{H}_h\rangle=O(\varepsilon h_0^{-1}LT)$, hence it is necessary that $m_i\geqslant O(\varepsilon^{-2}h_0^2h_i^{-1}L^{-1})$. To guarantee that the deviations contributed from all levels do not accumulate unboundedly, the slightly stronger condition

$$m_i\geqslant O(\varepsilon^{-2}h_0^{2-\delta}h_i^{\delta-1}L^{-1}) \tag{25}$$

where δ is any (small) positive number, may be required.

In particular, $m_0\geqslant O(\varepsilon^{-2}N^{-1})$. On the other hand, for the total work to be *at most* $O(\varepsilon^{-2})$ it is necessary that $Nm_0\leqslant O(\varepsilon^{-2})$, hence

$$m_0=O(\varepsilon^{-2}N^{-1}) \tag{26}$$

Both (25) and (26) can be satisfied by any multigrid cycle with index $\gamma>0.5$. The total work will still be $O(\varepsilon^{-2})$ and independent of N iff $\gamma<2$.

Thus, the energy can be calculated to $O(\varepsilon)$ relative accuracy in $O(\varepsilon^{-2})$ operations by a multigrid cycle with index $0.5 < \gamma < 2$. Effectiveness diminishes as γ approaches either endpoint (0.5 or 2). The V cycle, i.e., $\gamma = 1$, is of course the most convenient.

Results are presented in Fig. 4. $\mathscr{H}_h$ has been measured, using (14), after each relaxation sweep on *each* level. The average of these measurements, $\overline{\mathscr{H}_h}$, is an approximation for $\langle \mathscr{H}_h \rangle$. The measured relative error is defined as

$$\varepsilon = \left| \frac{\overline{\mathscr{H}_h}}{N-1} - \frac{T}{2} \right| \Big/ \frac{T}{2}$$

The values of α, defined as in Section 2.6, are shown for $\gamma = 0$ (simple Monte Carlo), 1, 2, and 3 as a function of the number m_0 of sweeps over a grid with $N = 64$ points. In all cases $\nu_1 = \nu_2 = 1$ is employed. For clarity of results, the work in first equilibrating the system by ten V cycles (actually a much smaller number would suffice) is not taken into account in calculating α. Indeed, for small ε this work should be negligible.

The results show the expected slowing down for $\gamma = 0$ and the optimality of $\gamma = 1$. The work on grid $h_i = 2^i h_0$ is $Nm_0(\gamma/2)^i$, so the total work (or #RAN) is $2Nm_0$ for $\gamma = 1$, $Nm_0 \log_2 N$ for $\gamma = 2$ and $Nm_0[(\gamma/2)^{\log_2 N} - 1]/(\gamma/2 - 1)$ for $\gamma > 2$. Hence, for $N = 64$ the work is

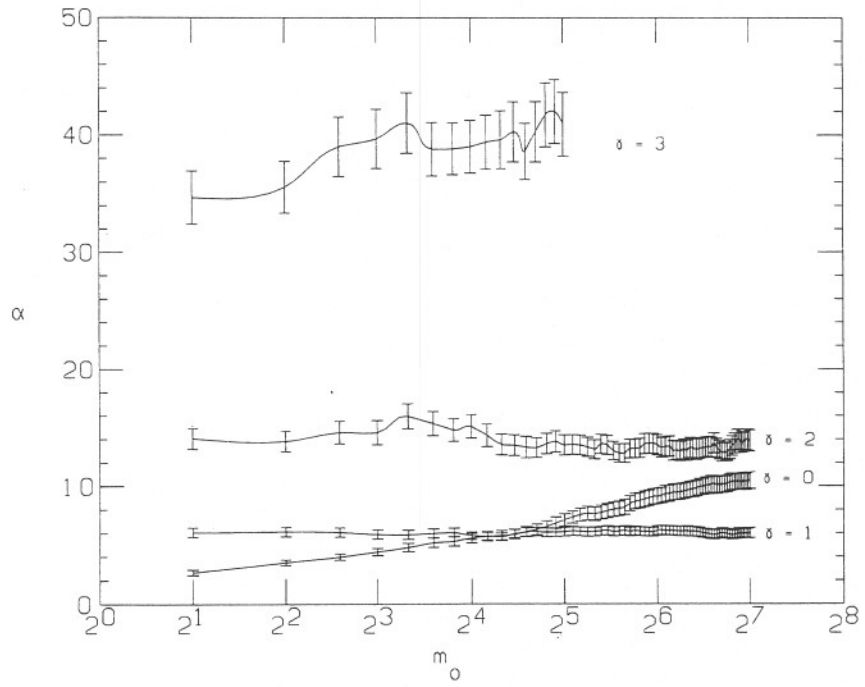

Fig. 4. Performance in measuring average energy. Each curve shows α (measuring computational work times the square of the obtained accuracy) as a function of the number m_0 of sweeps over a system with $N = 64$ points for the indicated value of the cycle index γ.

$2Nm_0$, $6Nm_0$, and $20.8Nm_0$ for $\gamma = 1, 2$, and 3, respectively. The relative values of α seen in Fig. 4 almost exactly correspond to this increase of work with γ, showing in comparison to it only a slight initial decrease (upon an initial increase of γ for fixed m_0). The slight decrease is due to the increased accuracy of smoother components, which are the only ones to benefit from

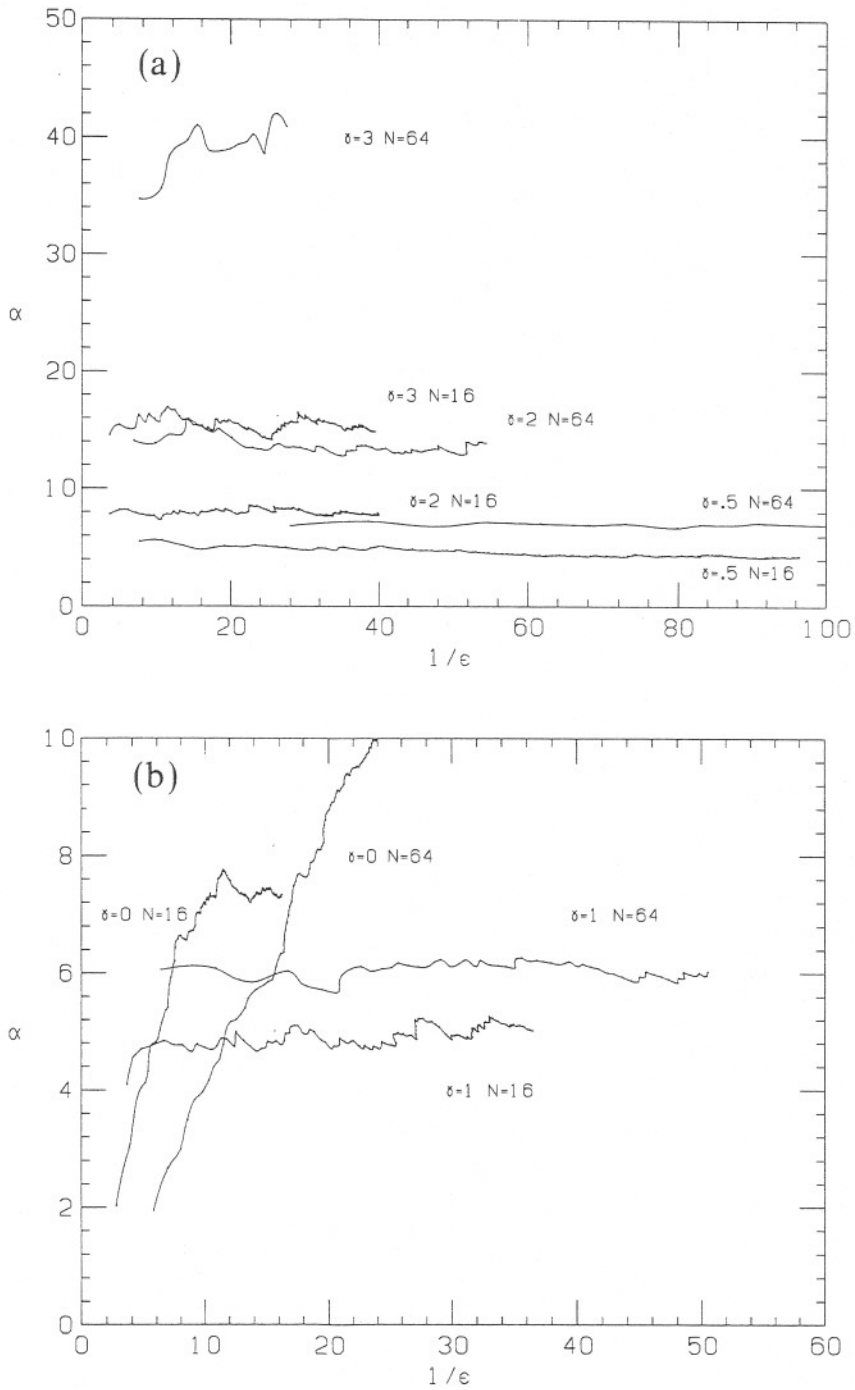

Fig. 5. (a) Performance in measuring average energy. Each curve shows α (measuring computational work times the square of the obtained accuracy) as a function of $1/\varepsilon$, for the indicated values of the system size N and the cycle index γ. (b) Same as in Fig. 4a, but differently scaled for different parameters.

higher γ (initially; when γ increases further, the additional benefit is negligible).

Generally, for any fixed N, α is approximately a constant if $\gamma > 0.5$, but for $\gamma \geqslant 2$ this constant increases with N, being $O(N^{\log \gamma/\log 2 - 1})$ for $\gamma > 2$ and $O(\log N)$ for $\gamma = 2$. Figures 5a and 5b demonstrate this.

3. TOWARD OPTIMAL ALGORITHMS FOR ISING MODELS

3.1. The Multigrid Cycle

A multigrid Monte Carlo method, based on a stochastic coarsening procedure, has been applied to the two-dimensional Ising model, successfully producing a CSD-free sequence of statistically independent configurations. See refs. 11, 12, and 15 for details; the following is a brief description.

Consider the ferromagnetic Ising model Hamiltonian associated with a spin configuration s:

$$\mathscr{H}(s) = -\sum_{\langle i,j \rangle} J_{ij} s_i s_j \qquad (J_{ij} > 0) \tag{27}$$

where s_i is the (1 or -1) value of the spin at site i and where $\langle i, j \rangle$ runs over all pairs of nearest-neighbor sites on a square, doubly periodic lattice. The basic strategy of the algorithm is the stochastic generation of a new, coarser Hamiltonian:

$$\mathscr{H}^1(s^1) = -\sum J_{ij}^1 s_i^1 s_j^1 \qquad (J_{ij}^1 \geqslant 0) \tag{28}$$

with a decreased number of degrees of freedom. Each new "spin" s_i^1 is actually a *block* of one or more spins s_α, so that each spin s_α belongs to one and only one particular block s_i^1. The process of creating the Hamiltonian $\mathscr{H}^1$ from $\mathscr{H}$ is referred to as going from fine to coarse level, or *coarsening*. Given $\mathscr{H}^1$, usual Monte Carlo sweeps can be performed to generate transitions in the phase space of the new level. The process of restoring finer-scale degrees of freedom, i.e., interpreting each flipped s_i^1 as a simultaneous flip of all the spins s_α in that block, and then returning to work with $\mathscr{H}$, is called *uncoarsening*. The description of the coarsening procedure and the organization of the coarsening/uncoarsening steps are given next.

The stochastic blocking is performed by scanning the fine-grid interactions ($J_{ij} s_i s_j$) one by one, in any convenient order, each interaction in its turn being either *kept alive* or *terminated*, according to a certain criterion (given below). If it is "kept alive," then it actually just stays unchanged. If it is "terminated," then with a probability $P_{ij} = \exp[-J_{ij}(s_i s_j + 1)/T]$ it is

deleted, i.e., the interaction between the spins is omitted from the coarse Hamiltonian; and with probability $1-P_{ij}$ it is *frozen*: its two interacting spins are blocked together so that both are flipped simultaneously in all subsequent coarser level moves (including still-coarser-level moves subsequently made), until uncoarsening takes place. This particular way of terminating bonds has been introduced by Swendsen and Wang,[18] and we will denote it SW. Its particular prescription for P_{ij} guarantees detailed balance and is also such that a bond connecting antiparallel spins (for which $s_i s_j = -1$) is deleted with probability 1, and thus only parallel spins can be blocked together.

Consequently, by freezing more bonds, blocks of increased sizes are created. The potential size of the produced block, i.e., the number of spins s_α it would include upon freezing a particular candidate bond, has been used as a criterion for deciding whether to terminate that bond or keep it alive; e.g., by adopting the rule of keeping the bond alive iff its freezing would produce a block of more than, say, four spins.

When the process of stochastic coarsening is completed, all spins are grouped into two different kinds of blocks: the *disconnected blocks*, each of which is separated from all others by boundaries of deletions and thus have no remaining alive interactions, and the *interacting blocks*, which still have alive bonds between them. The coupling J_{ij}^1 between two interacting blocks s_i^1 and s_j^1 is calculated by summing up all alive bonds connecting them. Starting with $J_{ij}=1$ on the fine level, stronger couplings ($J_{ij}^1 > 1$) may appear between the interacting blocks. Thus, a new Hamiltonian $\mathscr{H}^1$ is constructed.

The entire process can be repeated recursively: to be effective, the Monte Carlo simulation of $\mathscr{H}^1$ itself includes both conventional Monte Carlo sweeps and stochastic coarsening/uncoarsening steps. The next stochastic blocking is employed by regarding the blocks s_i^1 of the coarse level as now being the spins from which the blocks of the next coarser level are constructed. Each new, coarser block represents a block of blocks s_i^1 and in turn can be referred to as a block of spins s_α of the finest level. Repeating this recursively, a *sequence* of increasingly coarser *levels* is created. Each level k consists of a list of blocks, into which the original set of spins s_α (which in this notation are also s_α^0) is uniquely decomposed. This list is actually a union of two sublists: the disconnected blocks created at all finer levels up to and including the current one, and the interacting blocks, denoted s_i^k, which are coupled by the k-level Hamiltonian

$$\mathscr{H}^k(s^k) = -\sum J_{ij}^k s_i^k s_j^k \qquad (k=0, 1, 2,\ldots) \tag{29}$$

Progressing to increasingly coarser levels, the number of alive bonds keeps decreasing until at the coarsest level none exists. At that stage all

bonds are either frozen or deleted, so the original set of spins s_α is completely decomposed into disconnected blocks.

The entire algorithm can be described as a sequence of multigrid cycles for the *finest* level 0, where a multigrid cycle for *any* given level k (the current "fine" level, assumed not to be the coarsest level) is recursively defined as consisting of the following five steps.

1. ν_1 Monte Carlo sweeps are first made on the fine level, using the Hamiltonian (29).
2. The next coarser level $(s^{k+1}, \mathscr{H}^{k+1})$ is created from the fine level by the above stochastic coarsening process.
3. γ multigrid cycles for the coarse level are performed. If, however, this coarse level is the *coarsest*, do nothing.
4. Each interacting block s_i^{k+1} whose final value is different from its initial value is translated into flipping all the fine-grid spins s_α^k which belong to it. The spins s_α^k in each disconnected block are flipped simultaneously with probability 1/2.
5. ν_2 Monte Carlo sweeps are finally made on the fine level.

We have used cycles with $\nu_1 = \nu_2 = 1$ and $\gamma = 2$. The parameter γ is called the cycle index, and cycles with $\gamma = 2$ are called W-cycles.

The obtained results were in a certain respect very satisfactory: the CSD seem to have been completely eliminated, meaning that in a work proportional to the gridsize a new, substantially independent configuration was created. It was later proved[13] that some similar algorithms (in which, however, the ordering of bond termination is the same in all cycles) must suffer a (very slight) slowing down. This raises the suspicion that the same may be true here, although the proof does not strictly apply to our described algorithm (which generates different bond termination ordering in each cycle: see the above rule for keeping bonds alive). At any rate, the practical behavior for all tested gridsizes (up to 128×128) exhibited no trace of slowness.

3.2. Dependence of Configurations within a Cycle

Measuring the desired averages (observables) on the sequence of configurations produced by multigrid cycles on the finest level converges, to be sure, faster than measurements on the simple (only single-spin) Monte Carlo simulations. But still, as expected, the obtained statistics were slow in averaging out the deviation (from the observable average) exhibited by each configuration. If a standard deviation σ is contributed by the features of some scale, these features have to completely change $O((\sigma/\varepsilon)^2)$ times in

order to obtain accuracy ε. The trouble is presumably mostly related to large-scale deviations, because only few samples of them are contained in each produced configuration, while small-scale fluctuations, one could hope, are effectively averaged out in each given configuration.

The main idea for overcoming this difficulty was to average over as many different configurations *within* each cycle as possible. Take, for example, the calculation of the mean of the squared magnetization per site $\langle M^2 \rangle$, where $M^2 = (1/N)(\sum_{i=1}^{N} s_i)^2$ and N is the number of spins in the finest level. Instead of measuring M^2 once per cycle, compute it each time the algorithm visits the coarsest level, that is, 2^{l-1} times per W-cycle, where l is the number of levels. Moreover (as pointed out in the Appendix of ref. 11), on each visit to the coarsest level it is possible to average immediately over all the different spin configurations referring to their decomposition into disconnected blocks. More precisely, it is easy to show that if there are μ disconnected blocks consisting of $n_1, n_2, \ldots, n_\mu$ spins, respectively (where $\sum_{i=1}^{\mu} n_i = N$), then the average of M^2 taken over the 2^μ equally probable different spin configurations allowed by these blocks is given by

$$\overline{M^2} = \frac{1}{N} \sum_{i=1}^{\mu} n_i^2 \tag{30}$$

Experiments showed that, disappointingly, near the critical temperature the convergence with such many-per-cycle measurements was *not* substantially faster than with once-per-cycle measurements. We have compared the standard deviation exhibited by the once-per-cycle measurements with that of the average of the 2^{l-1} measurements (30) in one cycle, and found that on grids up to 128×128 the latter was not even half as small as the former. Even though this may somewhat improve on extremely large grids, not much can be gained in the practical range.

To understand this behavior, one should first observe that the average (30) of M^2 taken over the 2^μ equally probable configurations is heavily dominated by the largest block, hence they are very strongly correlated. This is particularly true at the critical (or lower) temperatures, where the size of the largest block far exceeds all others.

Consider next the relations between the 2^{l-1} coarsest configurations obtained within a single cycle. What the experiments show is that they, too, are correlated to each other. Each of them, in other words, strongly depends on both the finest configuration at which the cycle begins and on the first level of stochastic coarsening performed on it. A detailed study showed that the reason is that the stochastic coarsening tends to produce many more deletions along boundaries (between regions of opposite signs

in the current configuration) than elsewhere. As a result, in all subsequent coarse-grid calculations, there appears a statistical bias to retain many of these boundaries. In particular, the largest block is likely to remain roughly the same.

Thus, in reducing the number of degrees of freedom one loses not just fine-scale fluctuations, but large-scale ones as well.

3.3. Reducing the Dependence

At first, this strong correlation between scales appears to be a necessary property of discrete-state models. But then, a similar situation is encountered when *constant* interpolation is used even for the Gaussian model (cf. Section 5.3 in ref. 5). Since the SW termination resembles constant interpolation, the question now is whether a better coarsening technique, capturing some features from *linear* interpolation, can be devised for Ising spins, so as to reduce the dependence between the configurations produced within the same cycle, i.e., by the same coarsening. Let us denote by $\chi_0 = \langle M^2 \rangle$ the true susceptibility, by $\sigma_0 = \langle (M^2 - \chi_0)^2 \rangle^{1/2}$ the standard deviation from χ_0 of M^2 of any single configuration, by χ_1 the average of M^2 for the Hamiltonian $\mathscr{H}^1$, and by $\sigma_1 = \langle (\chi_1 - \chi_0)^2 \rangle^{1/2}$ the standard deviation of χ_1 from χ_0. The above coarsening, based on the SW termination, produced $0.5\sigma_0 < \sigma_1 < \sigma_0$. The question then is whether a better coarsening technique can produce σ_1 much smaller than σ_0.

One obvious difference between constant and linear interpolation is that the latter relates a given variable to *two* neighbors, not one. Thus, our first attempt at a linear-like interpolation is to replace the two-spin SW coarsening with the following *three-spin coarsening* (3SC).

For simplicity we describe (and have developed and tested) only the case of uniform bonds (constant J_{ij}); this is not essential, but introduces simplifying symmetries. Denote by $\beta = J_{ij}/T$ the uniform thermal binding between neighbors. Consider a spin s_0 with two neighbors, s_- and s_+, say. The current Hamiltonian has the form

$$\frac{1}{T}\mathscr{H} = -\beta s_0 s_- - \beta s_0 s_+ - \cdots$$

where the dots stand for all the other terms of $\mathscr{H}$. Three other Hamiltonians are offered as alternatives:

$$\frac{1}{T}\mathscr{H}_1 = -\infty s_0 s_- - a s_0 s_+ - \cdots$$

$$\frac{1}{T}\mathcal{H}_2 = -as_0 s_- - \infty s_0 s_+ - \cdots$$

$$\frac{1}{T}\mathcal{H}_3 = -bs_- s_+ - \cdots$$

The ∞ value in $\mathcal{H}_1$ ($\mathcal{H}_2$) means that s_0 and s_- (s_+) are blocked together. Note that in $\mathcal{H}_3$ the two bonds between s_0 and its two neighbors are deleted, but a new direct bond is introduced between the neighbors themselves. One selects $\mathcal{H}_i$ with probability P_i ($i=1, 2, 3$), where $P_1+P_2+P_3=1$. To obtain detailed balance, these probabilities are taken to depend on the current value of s_-, s_0, and s_+ according to Table I—plus the obvious rule that $P_i(-s_-, -s_0, -s_+) = P_i(s_-, s_0, s_+)$—and the values of a and b are taken so that

$$e^{2a} = (e^{2\beta} - e^{-2\beta})/(2-2p_*)$$

$$e^{2b} = e^{-2\beta}/p_*$$

p_* is a small positive parameter. We chose $p_* = 0.15$, but other values in the range $0.05 \leqslant p_* \leqslant 0.2$ are perhaps as good.

The detailed balance of this, and also that of SW and other coarsening schemes, is a special case of the following theorem, which generalizes the Kandel–Domany[10] Theorem.

Detailed Balance Theorem. Let u denote a configuration of a model, $\mathcal{H}(u)$ its Hamiltonian, and $\mathcal{H}_1(u)$, $\mathcal{H}_2(u)$,... some alternative Hamiltonians, where the use of the Hamiltonian $\mathcal{H}_i(u)$ also means restriction of the configurations u to a subset where some functionals $F_{i1}(u)$, $F_{i2}(u)$, etc., are constant. (Thus, upon selecting $\mathcal{H}_i$ we also *freeze* F_{ij}.) Then, in a Monte Carlo process with current configuration $\tilde{u}$, replacing $\mathcal{H}(u)$ by $\mathcal{H}_i(u)$ in probability $P_i(\tilde{u}) \geqslant 0$ maintains detailed balance provided

$$P_i(\tilde{u}) = f_i(F_{i1}(\tilde{u}), F_{i2}(\tilde{u}),\ldots)\, e^{\mathcal{H}(\tilde{u}) - \mathcal{H}_i(\tilde{u})} \tag{31}$$

Table I

s_-	s_0	s_+	P_1	P_2	P_3
+	+	+	$\frac{1}{2}(1-e^{-4\beta})$	$\frac{1}{2}(1-e^{-4\beta})$	$e^{-4\beta}$
+	−	+	0	0	1
+	+	−	$1-p_*$	0	p_*
+	−	−	0	$1-p_*$	p_*

where f_i are arbitrary functions and where

$$\sum_i P_i(\tilde{u}) = 1 \qquad \text{for any } \tilde{u} \tag{32}$$

Note that if no real freezing is done with a certain $\mathscr{H}_i$, then the coefficient f_i can only depend on i, and not in any way on $\tilde{u}$ (which is the special case proved in ref. 10).

Proof. Denote by $P(u^1 \to u^2)$ the probability of obtaining the configuration u^2 at any stage after starting with $\tilde{u} = u^1$. Then clearly

$$P(u^1 \to u^2) = \sum_k^{\{1,2\}} P_k(u^1)\, P_k(u^1 \to u^2) \tag{33}$$

where $P_k(u^1 \to u^2)$ is the probability of reaching u^2 from u^1 under the Hamiltonian $\mathscr{H}_k$, and where $\sum_k^{\{1,2\}}$ sums only over such k for which $F_{kj}(u^1) = F_{kj}(u^2)$ $(j = 1, 2, \ldots)$. But for each such k, by (31),

$$\frac{P_k(u^1)}{P_k(u^2)} = e^{\mathscr{H}(u^1) - \mathscr{H}_k(u^1) - \mathscr{H}(u^2) + \mathscr{H}_k(u^2)} = e^{\mathscr{H}(u_1) - \mathscr{H}(u^2)} \frac{P_k(u^2 \to u^1)}{P_k(u^1 \to u^2)}$$

Hence, by (33),

$$P(u^1 \to u^2) = P(u^2 \to u^1)\, e^{\mathscr{H}(u^1) - \mathscr{H}(u^2)}$$

which is the desired detailed balance. ∎

We have tested 3SC on an $L \times L$ periodic grid by applying the coarsening step for all triplets s_-, s_0, and s_+ at grid positions $(j, 2k-1)$, $(j, 2k)$, and $(j, 2k+1)$, respectively, such that $j+k$ is even. We compared it with an SW coarsening that terminated all the corresponding (s_0, s_-) and (s_0, s_+) bonds. Results at the critical temperature are summarized in Table II. They show that for 3SC, unlike SW, the ratio σ_1/χ_0 decreases with L. This means that if the susceptibility is measured on the first coarse grid,

Table II

L	χ_0	σ_0	σ_1 SW	σ_1 3SC
4	12.2		1.8	0.7
8	41.4		7.2	1.5
16	139.5	56.8	25.6	4.0
32	470.2	192.5	81.6	10.6

without ever returning to the fine, the average error is small: it tends to 0 as L increases.

The observation that has led to the construction of 3SC is that the basic flaw in the SW coarsening is the introduction of many deletions, usually clustered along well-defined lines: the lines of current boundaries of spin alignment. These lines therefore exhibit in $\mathscr{H}^1$ weakened couplings, and are thus likely to persist as boundaries of spin alignment also on coarse grids. This means strong correlation between different coarse-grid configurations. In 3SC the introduction of such weakened-coupling lines is minimized.

This is just a first attempt; it all may well be done better. Observe that the blocks created by 3SC are not necessarily *contiguous*: the Hamiltonian $\mathscr{H}_3$ creates a bond between s_- and s_+, so that later they may be blocked together without having the points in between, such as s_0, included in the block. More general schemes may create blocks that are not necessarily *disjoint*. And so forth: the possibilities are many.

It is not clear at this point whether the ideal statistical efficiency is always attainable. What *has* been established, we believe, is that it is possible to benefit greatly from making many measurements at the coarse levels of a multilevel Monte Carlo algorithm, even in discrete-state models, if a suitable coarsening scheme is used.

3.4. Optimal Calculation of T_c

Even though the SW coarsening is not optimal, as explained above, it can still be used in an optimal calculation of certain thermodynamic quantities. As the simplest example of such a quantity we chose the critical temperature T_c itself. The tests reported below indicate that the above multigrid cycle (Section 3.1) can directly be used for a very inexpensive, in fact optimal determination of T_c. More precisely, a sequence of increasingly better approximations to T_c is obtained on increasingly larger grids by performing only a few cycles on each. To achieve an accuracy ε in T_c, the amount of computational work turns out to be $\alpha\varepsilon^{-2}$, where the average value of α is about 100.

The algorithm is based on the following measurements. In every W-cycle, performed with some temperature T, the algorithm visits the coarsest level many times. At each such visit the domain is completely decomposed into disconnected blocks, and the ratio r between the number of spins in the largest of these blocks and the total number N of spins in the lattice is measured. Denote by $\bar{r}$ the average of these r's within one W-cycle; clearly $0 < \bar{r} \leqslant 1$. Generally, if $\bar{r} < \eta$ (η chosen as described below),

then we say that the cycle "indicates" a supercritical temperature: $T > T_c$ (or $\beta = 1/T < \beta_c$). Similarly, if $\bar{r} > \eta$, then the cycle indicates a subcritical temperature: $T < T_c$ (or $\beta > \beta_c$).

This definition of T_c is suitable for fairly large grids if η is chosen reasonably small. In fact, in the range we have calculated (grids up to 128×128), $\eta = 0.5$ has already proved to be small enough. In principle, for much larger grids the definition of T_c should be modified to allow for the fact that $\bar{r}$ becomes small near T_c, even in the subcritical range. One possible modification is to replace r by r', defined as the ratio between the *length* of the largest block and the length L of the domain. This quantity r', or its average within a cycle, does not become small near T_c. [The length of a block B can, for example, be defined as $1 + \max_B |i_1 - j_1|$, where the max is taken over all pairs of points $i = (i_1, i_2)$ and $j = (j_1, j_2)$ such that both i and j are in B. For locating T_c on very large grids it is enough to calculate the *approximate* length of the largest block. The coarsening process can very inexpensively incorporate a procedure that supplies each block B on some coarse level with its approximate $\min_B i_1$ and $\max_B i_1$, from which similar quantities can be calculated for all blocks at all coarser levels.] In the practical range of our calculations, however, this more elaborate quantity r' proved unnecessary. What our simpler procedure, based on r, calculates is in fact another thermodynamic quantity, $T(\eta)$, which is the temperature for which the average magnetization per site is η, i.e., the temperature for which $\lim_{N \to \infty} \langle |\sum_{i=1}^{N} s_i|/N \rangle = \eta$. But the difference $|T_c - T(0.5)|$ is below the accuracy one can obtain with grids up to 128×128, so we had no motivation to run tests with r' instead of r.

The experiments show that using an $L \times L$ lattice, an interval of roughly $1/L$ around β_c is the best approximation one could get for the critical value; i.e., within that interval, the criterion does not correctly distinguish between sub- and supercritical temperatures. To get an approximation twice more accurate, it is therefore necessary to switch to a four times larger grid: $2L \times 2L$. [This observation is of course in agreement with the known critical exponent $\nu = 1$, i.e., with the correlation length being proportional to $(T - T_c)^{-1}$.] In order to save work, the algorithm is constructed so that much of the search for T_c on any given grid is carried over from smaller grids. Since each interval is being further corrected by a larger grid's interval, it is not important to check the criterion precisely, and for practical purposes it is sufficient to perform only *one* W-cycle at each temperature. In this way, on each of the grids a computational work equivalent to just a few Monte Carlo passes is enough for determining T_c to within an interval roughly as narrow as can ever be obtained on that grid. The interval will get narrower and narrower as the grid becomes larger and larger, until a desired accuracy is obtained. The details of the algorithm are as follows.

Initialization. Start from a very small random grid—level 0 (say 4×4). Set an initial temperature $T_0 = T_0^0 = 1/\beta_0^0$ (the subscript stands for the level, the superscripts for the sequence of temperatures within each level). For instance, $\beta_0^0 = 0$, which is an infinite temperature and hence, for sure, supercritical. Finally, choose some $\Delta\beta_0 > 0$, the step in which the temperature is lowered. (The values used by us are shown in Fig. 6 below.)

For each level $i = 0, 1, \ldots$ do the following three steps.

1. Perform one W-cycle (with $\beta = \beta_i^0$) to reach near equilibrium [erasing in particular lower-level periodicity (see step 3 below)].
2. Make one W-cycle for each $\beta_i^k = \beta_i^0 + k\,\Delta\beta_i$, $k = 0, 1, \ldots$, until either $\bar{r} > \eta$ for $\Delta\beta_i > 0$ or $\bar{r} < \eta$ for $\Delta\beta_i < 0$ is obtained for some k. If this condition is already satisfied at the first step (W-cycle), then $\Delta\beta_i \leftarrow -\Delta\beta_i$, i.e., switch the direction of the search.
3. The β_i^k for which $\bar{r}$ has first passed η will be denoted β_i and will serve as our final approximation to β_c on level i. Switch to level $i+1$: its grid is four times larger (factor two in each direction), and its initial configuration is the current configuration on level i extended periodically in each direction (exploiting its doubly periodic boundary conditions). Set $\beta_{i+1}^0 = \beta_i$ and $\Delta\beta_{i+1} = -\Delta\beta_i/2$. Go to step 1 with $i+1$ replacing i.

The step ratio $\Delta\beta_{i+1}/\Delta\beta_i = -1/2$ is reasonable due to the known ratio, mentioned above, between the accuracy obtainable on the corresponding grids. If this ratio were not known (i.e., if we did not know that in this

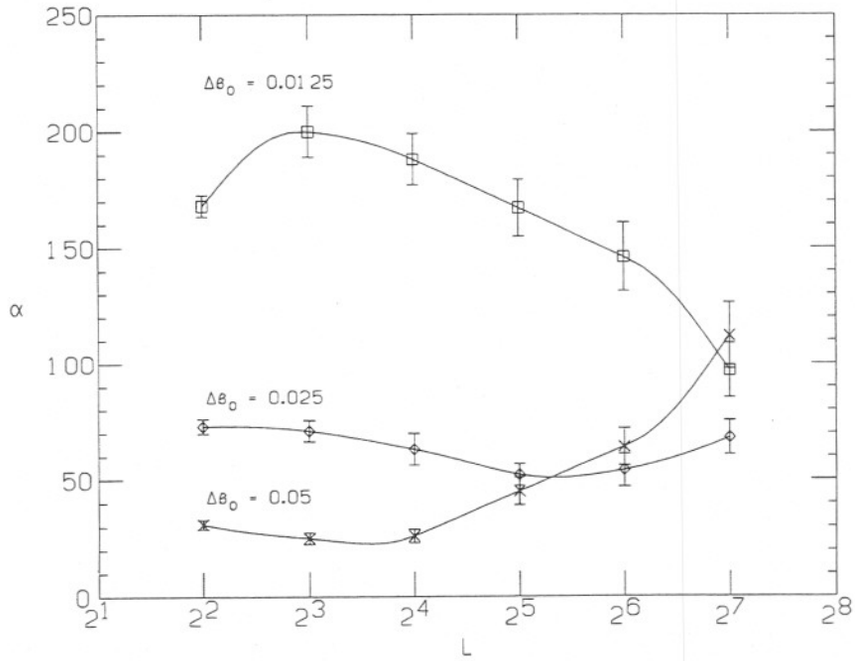

Fig. 6. Each curve shows α as a function of the system size L, for the indicated value of $\Delta\beta_0$.

particular model $\nu = 1$), the value of $\Delta\beta_{i+1}$ would have to be determined in an adaptive way, decreasing it faster whenever too many steps were required in the former lattice size.

This algorithm produces an open-ended sequence of increasingly better approximations to β: $\beta_0, \beta_1, \ldots$, as long as required or allowed by computing resources.

It should be noticed that the algorithm saves a lot of work by not insisting on exact measurements of $\langle \bar{r} \rangle$, or even on exact equilibration, at each temperature. It just senses the approximate point at which β_c, the critical temperature, is passed. More accuracy will not help, since on each grid that point is only fuzzily defined.

Numerical results. The accuracy ε of β_i, the ith approximation for β_c, is defined by the difference $|\beta_i - \beta_c|$, where $\beta_c = 0.4406868$ is the known thermodynamic limit. The amount of work, denoted by #RAN, has been measured by us by the number of times a random number is generated. This effectively counts the work in the Monte Carlo sweeps and in the stochastic coarsenings, which are indeed the most time-consuming processes. For an algorithm to be optimal, the quantity $\alpha = \varepsilon^2(\#\text{RAN})$ should be roughly constant, or at least bounded. In Fig. 6 the values of α obtained in our experiments are shown for grid sizes up to 128×128. Each shown value of α is averaged over an ensemble of about 100 systems.

The results show that the algorithm is not sensitive, within limits, to changes in $\Delta\beta_0$: asymptotically, T tends to the correct value of T_c, and at approximately the same rate α. The behavior of α as a function of the system size can, however, serve as a good indicator for fixing the stepsize $\Delta\beta_0$. Large values of $\Delta\beta_0$ yield fast localization of the first β_i's, hence small values of α for small L. But these β_i's are less accurate, hence more localization work is needed later, thus yielding a typical rise in α for large L. Too small values of $\Delta\beta_0$, on the other hand, show much increased values of α for small L. This is explained as follows. A small $\Delta\beta_0$ necessarily implies many steps at each grid size. Not only is the amount of work thus increased, but so is, on small enough grids, the probability of *accidentally* crossing the threshold η too early, implying a bad approximation for T_c, i.e., a large α. For larger grids the test becomes more reliable (accidental crossing is much less likely), hence a smaller step provides a better approximation for T_c and in turn a smaller α.

4. EXTENSIONS

The multilevel computational methods for eliminating CSD and for fast calculation of thermodynamic limits described in the previous sections

are being extended to increasingly more complicated models. Some initial steps and results are briefly described below. Fuller discussion will appear elsewhere.

4.1. Generalized Gaussian Models

The algorithm of Section 2 has been extended to the generalized Gaussian Hamiltonian in d dimensions,

$$\mathscr{H}(u) = \int_{\Omega} a(x) \sum_{j=1}^{d} \left(\frac{\partial u}{\partial x_j} \right)^2 dx_1, ..., dx_d \tag{34}$$

where $u = u(x)$ and $a(x) > 0$ are real functions defined for $x = (x_1, ..., x_d) \in \Omega$, and Ω is any domain in R^d on the boundary of which values of u are being prescribed.

Of special interest are cases where $a(x)$ is strongly discontinuous, changing by orders of magnitude from one subdomain to another. Such models can no longer be analyzed, nor accelerated, by Fourier methods. The usual (point-by-point) Monte Carlo process can suffer in some such cases far greater slowness than in the constant-coefficient $[a(x) \equiv 1]$ case: the number of sweeps for producing an effectively independent configuration may grow proportionately to $h_0^{-2} r$, where h_0 is the meshsize and $r = \max_{x,y} a(x)/a(y)$. Hence, the total number of sweeps to obtain accuracy ε for any thermodynamic limit may grow as $O(\varepsilon^{-2} h_0^{-2-d} r)$, where h_0 decreases as some positive power of ε.

We have found that in order to reach optimality, the multigrid algorithm for such cases must differ from the one described in Section 2, mainly in the following two points.

1. *Weighted interpolation.* Instead of the simple linear interpolation used in (13), *weighted* interpolation must be used, with the weights in each direction being proportional to the size of $a(x)$ extending in that direction. [See ref. 1, where this weighted interpolation is described for the energy minimization ($T = 0$) problem.]

2. *Variable sampling.* The Monte Carlo process should sample more frequently regions with smaller values of the coupling $a(x)$. A general rule which has been derived for the optimal calculation of susceptibility, for example, is that the number of Monte Carlo steps at each gridpoint x on each grid with meshsize h should be proportional to $a(x)^{-2/3} h^{(2d+4)/3}$. This rule also dictates the relative amount of sampling on different grids, hence the cycle index.

Such algorithms were implemented for various strongly discontinuous cases in one and two dimensions ($d = 1, 2$). The results were as good as

those in Section 2, obtaining accuracy ε in approximating the infinite-grid susceptibility in less than $40\varepsilon^{-2}$ random number generations. This bound on the amount of work is independent of the above ratio r.

4.2. *XY* Models

Tests have been conducted on the one- and two-dimensional XY models, with various approaches to the coarsening process. The work was done partly in collaboration with S. Shmulyian, as reported in ref. 16.

The results show that optimal algorithms (eliminating both the critical slowing down and the volume factor) can be constructed, at least in the low-temperature range. To obtain such optimality, however, several additional algorithmic ideas must be introduced. The main ones, not restricted to the XY model, are surveyed below.

4.3. Coarsening by Approximation

If the given discrete Hamiltonian $\mathscr{H}_h(u^h)$ depends polynomially on u^h [e.g., quadratic dependence, as in (7)], and if each u_i^h is an unrestricted real or complex number or vector, then the coarse-grid Hamiltonian $\mathscr{H}_H(u^H)$ can exactly be derived, yielding again a polynomial of the same order [e.g., cf. (15)]. Simple expressions for $\mathscr{H}_H$, generally similar in form to $\mathscr{H}_h$, can also be derived in many other cases by using a low-order interpolation I_H^h (e.g., first-order, or "constant," interpolation,[8] i.e., blocking several fine-grid variables so that their changes by the coarse grid, $I_H^h u^H$, are identical). However, such low-order interpolations are not optimal, even when they yield substantial multigrid acceleration. In the Gaussian case, for example, and in many other (e.g., "asymptotically free") cases as well, the interpolation must be at least of second order (e.g., linear interpolation) to produce the full required mobility on the coarse levels (see item C in Section 2.4).

Second-order interpolation in non-Gaussian models is not a straightforward linear interpolation. The latter would usually produce a result outside the given states of the model [e.g., an (X, Y) state such that $X^2 + Y^2 \neq 1$] and should therefore be corrected [e.g., normalized by being multiplied by $(X^2 + Y^2)^{-1/2}$]. In other models or other representations (e.g., the XY model represented in terms of angles), straightforward linear interpolation is not well defined (e.g., because adding 2π to one of the angles would change the result). The definition of the interpolation should then include a deviation from linear interpolation, and contain bifurcations (separate formulas for different cases). To maintain detailed balance in

such cases is usually facilitated by writing $\mathscr{H}_H$ not in terms of the displacement function u^H, but in terms of the "full approximation scheme" (FAS) function $\hat{u}^H = I_h^H \tilde{u}^h + u^H$, where I_h^H is some fine-grid-to-coarse-grid restriction operator; e.g., I_h^H can be the simple "injection," defined by $(I_h^H \tilde{u}^h)(x) = \tilde{u}^h(x)$ (assuming that each coarse-grid point x also belongs to the fine grid). Thus, instead of $\mathscr{H}_H(u^H)$, it would usually be simpler to write the coarse-grid Hamiltonian as $\hat{\mathscr{H}}_H(\hat{u}^H) = \mathscr{H}_H(\hat{u}^H - I_h^H \tilde{u}^h)$. Even then, the resulting expression of $\hat{\mathscr{H}}_H$ as function of $\hat{u}^H$ is substantially more complicated than $\mathscr{H}_h(u^h)$. The complexity will be similarly further increased upon each additional coarsening. To avoid such compounded complexity, approximation methods are used, as follows.

For definiteness of the description, assume that the model is two dimensional and that the Hamiltonian $\mathscr{H}_h(u^h)$ can be written in terms of *stresses*, i.e., differences $u_i^h - u_j^h$ with i and j being nearest neighbors. Then, with a proper second-order interpolation I_H^h, the coarse-grid Hamiltonian $\hat{\mathscr{H}}_H(\hat{u}^H) = \mathscr{H}(\tilde{u}^h + I_H^h(\hat{u}^H - I_h^H \tilde{u}^h))$ can be written as a sum $\hat{\mathscr{H}}_H(\hat{u}^H) = \sum_q V^q(w_q)$, where each V^q is a (complicated and possibly bifurcated) expression in terms of w_q, which is the vector of the three stresses (of $\hat{u}^H$) belonging to the same coarse-grid plaquette. To curb the Hamiltonian complexity growth, we want to replace each $V^q(w_q)$ by a simpler *approximate* expression $\tilde{V}^q(w_q)$. Various such simplifying approximations can be constructed, the only important rules being the following.

(i) $\tilde{V}^q(w_q) \geq V^q(w_q)$ for any w_q, with equality obtained at least at one w_q.

(ii) For a small temperature (large β), the function $V^q(w_q)$ almost surely has a minimum at a point w_q^* where it is at least twice differentiable. Then, for such a temperature, the values of $\tilde{V}^q$ and all its first- and second-order derivatives are required to coincide with those of V^q at w_q^*.

(iii) $\tilde{V}^q$ should retain the topological properties of the model, such as 2π-periodicity in angle variables, to allow large-scale topological changes.

Just replacing V^q by $\tilde{V}^q$ would of course introduce statistical errors. Instead, to maintain exact detailed balance, the transition is done stochastically; namely, the transition is done in probability

$$P_t(\tilde{u}^h) = e^{\beta(V^q(w_q^0) - \tilde{V}^q(w_q^0))}$$

where w_q^0 is the value of w_q at the time of coarsening, i.e., for $\hat{u}^H = I_h^H \tilde{u}^h$. When transition does not occur, i.e., in probability $P_f(\tilde{u}^h) = 1 - P_t(\tilde{u}^h)$, the stresses w_q are *frozen*. By the Detailed Balance Theorem (see Section 3.3), this probabilistic choice between the simplifying transition and freezing maintains detailed balance. The meaning of freezing any stress $w = \hat{u}_i^H - \hat{u}_j^H$

is, of course, that the neighboring values $\hat{u}_i^H$ and $\hat{u}_j^H$ change henceforth (until the uncoarsening stage) *simultaneously*, keeping their differences fixed.

Rule (i) above guarantees that $P_t \leqslant 1$. Rule (ii) sees to it that for large β freezing is rare, and the system behaves virtually as in the Gaussian case. By increasing the order of approximation [requiring in (ii) more derivatives to coincide], freezing can be made even rarer, or remain rare even for smaller β. Such higher-order approximations increase of course the complexity of $\tilde{V}^q$, but the complexity per gridpoint remains fixed at subsequent coarsening steps: the Hamiltonians retain a fixed general form at all coarser levels.

In Section 2.3 we have seen that the coarse-field ϕ^H is stochastic: it depends on $\tilde{u}^h$, the fine-grid configuration at the time of coarsening. In non-Gaussian models other coefficients in the Hamiltonian depend on $\tilde{u}^h$ as well. The variable coefficients thus created imply that the methods described in Section 4.1 should be used. In particular, the interpolation at coarser levels needs to be *weighted* proportionately to the coupling strength in order to maintain full mobility at the coarsest levels.

The stochastic simplification of V^q to $\tilde{V}^q$ should actually be made just before the *next* coarsening step (the transition from grid $H=2h$ to grid $2H=4h$). Then, wherever freezing occurs, it just corresponds to a special choice of the I_{2H}^H interpolation weights. Since this choice in fact means constant interpolation, which yields simpler coarser interactions, the simplification of V^q to $\tilde{V}^q$ need not be done (hence freezing will not occur) at certain plaquettes adjacent to the frozen one.

In case of *gauge fields in a higher dimension*, the above approach is applicable, too, except that "stresses" should be replaced by "topological charges around plaquettes," and a "plaquette" should be changed to a "cube" of the proper dimension.

4.4. Domain Replication and Macroscopic Dynamics

If the fine-grid Hamiltonian uses periodic boundary conditions, instead of using cycle index γ (cf. Section 2.3), the domain can be "replicated" γ times; i.e., the coarse Hamiltonian can be extended periodically to a domain γ times larger. [To extend equally in all d coordinates, $\gamma = 2^d$ can be chosen. If such γ is too large (cf. Sections 2.4A, 2.7, and 2.8), alternate coarsening levels can use alternate coarsening directions.] No return to finer levels will ever be needed if they have already provided enough statistics (i.e., if the required number of cycles is 1, as indeed recommended in Section 2.4D). This is possible to do exactly when needed, i.e., when the size of the domain required to yield accuracy ε in some calculation is such

that it would contain more than $O(\varepsilon^{-2})$ finest-grid sites: the computational cost can still be only $O(\varepsilon^{-2})$, since the finest level is not employed over the entire domain.

One can make a sequence of such domain replications. At each step the domain is first coarsened, then replicated, then a simple multigrid cycle is made on the extended domain to reach an equilibrium. (Compare to the process in Section 3.4.) In this way one reaches ever larger domains, covered by increasingly coarser grids. After sufficiently many such steps one may reach Hamiltonians that represent the macroscopic dynamics of the system. (The assumption here is that the finest level need not interact with grids *many* times coarser. Indeed, any movement on such very coarse grids is very nearly seen as just a constant shift of the field on the finest scale. Such a constant shift does not normally interact with local fluctuations.)

5. SUMMARY

The calculation of an average quantity Q for an infinite system (a "thermodynamic limit" of finite systems) to within some prescribed accuracy ε by a Monte Carlo process usually requires the following three factors of complexity.

1. First, one should employ a large enough computational lattice $N \times N \times \cdots = N^d$, whose linear dimension N should usually increase as ε decreases: $N = N(\varepsilon)$; presumably N grows like $\varepsilon^{-\rho}$, where ρ is positive.
2. On this lattice one carries out a Monte Carlo process which produces a sequence of configurations, each configuration (from a certain point on) appears in its physical probability. Many of these configurations add nothing to the statistical measurement of Q, because they strongly depend on each other. The process requires $O(N^z)$ Monte Carlo sweeps, hence $O(N^{d+z})$ computer operations, to create each new, effectively independent configuration. The critical exponent z is of course nonnegative. The critical slowing down is the case where z is positive.
3. It is not enough to create one independent configuration, because any such configuration has a deviation from Q. If the standard deviation is σ, one would need $O(\sigma^2/\varepsilon^2)$ independent configurations in order to measure Q to the desired accuracy ε.

Taking these three factors together, one would overall need

$$O(N^{d+z}\sigma^2/\varepsilon^2) = O(\sigma^2/\varepsilon^{2+\rho(d+z)}) \tag{35}$$

computer operations in order to obtain an error smaller than ε.

The purpose of previous multigrid and cluster algorithms has been to reduce z as much as possible. The purpose of the multigrid techniques presented in this work is to eliminate the entire exponent $\rho(d+z)$ from (35), i.e., to *obtain an error smaller than* ε *in only* $O(\sigma^2/\varepsilon^2)$ *overall computer operations.*

This potential efficiency is especially good news for higher- (e.g., four-) dimensional problems: the work increase with accuracy is essentially independent of the dimension.

It is shown above in detail, especially in Section 2, how to achieve such optimal results in some simple cases. The possible extension to more advanced asymptotically free models is discussed in general terms in Section 4.

The parameters of the multigrid algorithm, such as the cycle index γ and the coarse-to-fine interpolation order, depend not only on the involved model and its discretization, but also on the measured quantity Q. In d-dimensional models and 1:2 coarsening ratio, for calculating quantities dominated by large-scale fluctuations (e.g., susceptibility), $2^d < \gamma \leqslant 2^{2p}$ must be used, where p is the order of discretization. For quantities dominated by small-scale fluctuations (e.g., the energy per degree of freedom), $\gamma < 2^d$ is needed to obtain accuracy ε in $O(\varepsilon^{-2})$ computational work.

At least second-order (e.g., linear-polynomial) coarse-to-fine interpolation is necessary for optimal calculations of asymptotically free models.

ACKNOWLEDGMENTS

We are grateful to R. Benav for useful discussions, and to E. Domany and D. Kandel for a basic correction to our method in Section 3.4. The research was supported in part by grants No. I-131-095.07/89 from the German–Israeli foundation for Research and Development (GIF), No. 399/90 from the Israeli Academy of Science and Humanities, AFOSR-91-0156 from the U.S. Air Force, and NSF DMS-9015259 from the U.S. National Science Foundation.

REFERENCES

1. R. E. Alcouffe, A. Brandt, J. E. Dendy, Jr., and J. W. Painter, The multi-grid methods for the diffusion equation with strongly discontinuous coefficients, *SIAM J. Sci. Stat. Comp.* **2**:430–454 (1981).
2. A. Brandt, Multigrid Techniques: 1984 Guide, with Applications to Fluid Dynamics [available as GMD Studien Nr. 85, GMD-AIW, Postfach 1240, D-5205, St. Augustin 1, Germany].
3. A. Brandt, Multilevel computations: Reviews and recent developments, in *Preliminary Proceedings 3rd Copper Mountain Conference on Multigrid Methods* (April 1987); see also

in *Multigrid Methods: Theory Applications and Super-computing*, S. F. McCormick, ed. (Marcel Dekker, New York, 1988), pp. 35–62.
4. A. Brandt, The Weizmann Institute research in multilevel computation: 1988 report, in *Proceedings 4th Copper Mountain Conference on Multigrid Methods*, J. Mandel *et al.*, eds. (SIAM, 1989), pp. 13–53.
5. A. Brandt, Multigrid methods in lattice field computations, *Nucl. Phys. B (Proc. Suppl.)* **26**:137–180 (1992).
6. A. Brandt, D. Ron, and D. J. Amit, Multi-level approaches to discrete-state and stochastic problems, in *Multigrid Methods*, W. Hackbusch and U. Trottenberg, eds. (Springer-Verlag, Berlin, 1986), pp. 66–99.
7. M. Galun, Optimal multigrid algorithms for model problems in statistical mechanics, M.Sc. Thesis, Weizmann Institute of Science (1992).
8. J. Goodman and A. D. Sokal, Multigrid Monte Carlo methods for lattice field theories, *Phys. Rev. Lett.* **56**:1015–1018 (1986).
9. S. Gottlieb, W. Liu, D. Toussaint, and R. L. Sugar, Testing an exact algorithm for simulation of fermionic QCD, *Phys. Rev. D* **35**:2611 (1987).
10. D. Kandel and E. Domany, General cluster Monte Carlo dynamics, *Phys. Rev. B* **43**:8539 (1991).
11. D. Kandel, E. Domany, and A. Brandt, Simulations without critical slowing down—Ising and 3-state Potts models, *Phys. Rev. B* **40**:330 (1989).
12. D. Kandel, E. Domany, D. Ron, A. Brandt, and E. Loh, Jr., Simulations without critical slowing down, *Phys. Rev. Lett.* **60**:1591 (1988).
13. X.-J. Li and A. D. Sokal, Rigorous lower bound on the dynamic critical exponent of some multilevel Swendsen–Wang algorithms, *Phys. Rev. Lett.* **67**:1482 (1991).
14. G. Mack and A. Pordt, Convergent perturbation expansions for Euclidean quantum field theory, *Commun. Math. Phys.* **97**:267 (1985); G. Mack, in *Nonperturbative Quantum Field Theory*, G. t'Hooft *et al.*, eds. (Plenum Press, New York, 1988), p. 309.
15. D. Ron, Development of fast numerical solvers for problems in optimization and statistical mechanics, Ph.D. Thesis, Weizmann Institute of Science (1989).
16. S. Shmulyian, Multilevel Monte Carlo algorithms for spin models, M.Sc. Thesis, Weizmann Institute of Science (1993).
17. A. D. Sokal, How to beat critical slowing-down: 1990 update, *Nucl. Phys. B (Proc. Suppl.)* **20**:55–67 (1991).
18. R. H. Swendsen and J. S. Wang, Nonuniversal critical dynamics in Monte Carlo simulations, *Phys. Rev. Lett.* **58**:86–88 (1987).
19. U. Wolff, Collective Monte Carlo updating for spin systems, *Phys. Rev. Lett.* **62**:361–364 (1989).