Multigrid Monte Carlo Method for Lattice Field Theories

Jonathan Goodman and Alan D. Sokal
Courant Institute of Mathematical Sciences, New York University, New York, New York 10012
(Received 17 January 1986)

We propose a stochastic generalization of the multigrid method, which reduces critical slowing down in Monte Carlo computations of lattice field theories. For free fields, critical slowing down is completely eliminated. For a ϕ^4 model, numerical experiments show a factor of ≈ 10 reduction, over a standard heat-bath algorithm, in the work needed to get a given accuracy (error-bar size).

PACS numbers: 11.15.Ha, 02.70.+d, 05.50.+q

Monte Carlo calculations of critical phenomena in statistical mechanics[1] and of the continuum limit in quantum field theory[2] have been greatly hampered by critical slowing down: The autocorrelation time τ of the traditional Monte Carlo algorithms grows rapidly as the critical point is approached.[3] We present here a new class[4] of Monte Carlo methods, called "multigrid Monte Carlo" (MGMC), that should have shorter autocorrelation times in the critical region. For Gaussian (free) fields, a rigorous analysis[5] shows that critical slowing down is completely eliminated; the gain in efficiency over traditional algorithms thus grows without bound as the critical point is approached. For ϕ^4 fields, numerical experiments (reported below) show a gain in efficiency, over a single-site heat-bath algorithm, by a factor of ≈ 10. The MGMC idea applies to many other models, including plane rotators and U(1) lattice gauge theories.[4,6]

The MGMC method is a stochastic generalization of the multigrid (MG) method for solving finite-difference equations.[7] It is philosophically similar to, but technically quite different from, the block-spin renormalization group.[8] Consider, for purposes of illustration, a scalar field theory on a periodic L^d lattice Λ with Hamiltonian

$$H(\phi)=\frac{\alpha}{2}\sum_{|x-x'|=1}(\phi_x-\phi_{x'})^2+\sum_{x\in\Lambda}P(\phi_x), \quad (1a)$$

where

$$P(\phi_x)=\lambda\phi_x^4+(A/2)\phi_x^2+h\phi_x. \quad (1b)$$

The problem is to generate random configurations ϕ with probability distribution

$$d\mu(\phi)=Z^{-1}e^{-H(\phi)}\,d\phi. \quad (2)$$

This is usually done by simulation of a Markov chain with some transition probability $P(\phi\to\phi')$ that preserves the measure (2). The goal is to find such a Markov chain that can be efficiently simulated on a computer and has a small autocorrelation time τ.

First note that if $P_1,\ldots,P_n$ are transition probabilities that preserve (2), then so is their product $P_1P_2\cdots P_n$. Our methods are products of this form where the individual transformations P_k are "partial resamplings." Suppose that we have a decomposition of the vector space of fields $\Phi=R^\Lambda$ as a direct sum $\Phi=\Psi\oplus\Pi$, so that each field is uniquely written as $\phi=(\psi,\pi)$, $\psi\in\Psi$, $\pi\in\Pi$. The conditional probability distribution of ψ given π is then

$$d\nu(\psi|\pi)=Z(\pi)^{-1}e^{-H(\psi,\pi)}\,d\psi, \quad (3)$$

where we have written $H(\psi,\pi)$ for $H(\phi)$. If $\phi=(\psi,\pi)$ is a random field with distribution $d\mu$ and $\phi'=(\psi',\pi)$ where ψ' has the distribution $d\nu(\cdot|\pi)$, then ϕ' also has the distribution $d\mu$. We emphasize that this is true whether or not ψ' is independent of ψ. An example is the single-site heat-bath method: The sites are ordered $x_1,x_2,\ldots,x_N$ ($N=|\Lambda|=L^d$) and P_k is an independent resampling of ϕ_{x_k}—that is, $\psi=\phi_{x_k}$, $\pi=(\phi_{x_1},\ldots,\phi_{x_{k-1}},\phi_{x_{k+1}},\ldots,\phi_{x_N})$, and $\psi'=\phi'_{x_k}$ is independent of $\psi=\phi_{x_k}$. The single-site Metropolis method is similar, but ψ' is not independent of ψ.

In MGMC, these single-site resamplings are supplemented by certain "collective mode" resamplings.[9] Assume that L is even, and divide the lattice Λ into $(L/2)^d$ cubical blocks with 2^d sites each. These blocks B_y correspond to the sites of a coarser lattice, Λ_c, with $L/2$ sites on a side. Define the coarse-lattice variables (fine-lattice block means)

$$\psi_y=2^{-d}\sum_{x\in B_y}\phi_x,$$

and let π denote some set of complementary variables. The MGMC method supplements the single-site resamplings with resamplings of the coarse-lattice field $\psi=\{\psi_y\}_{y\in\Lambda_c}$. These coarse-lattice resamplings are themselves carried out by the MGMC method, and so on recursively (see below for details).

The computational efficiency of MGMC rests on the fact that the "conditional Hamiltonian" $H_c(\psi|\pi)=H(\psi,\pi)$, which determines $d\nu$ in (4), has the form

$$H_c(\psi|\pi)=\frac{\alpha'}{2}\sum_{|y-y'|=1}(\psi_y-\psi_{y'})^2+\sum_y P'_y(\psi_y), \quad (4a)$$

where

$$P'_y(\psi_y)=\lambda'\psi_y^4+\tfrac{1}{2}A'_y\psi_y^2+h'_y\psi_y. \quad (4b)$$

The coefficients in (4) are determined by direct substi-

tution into H; the site-dependent coefficients A_y' and h_y' depend on π, but $\alpha' = 2^{d-1}\alpha$ and $\lambda' = 2^d\lambda$ do not. Since the coarse-lattice Hamiltonian is of the same form as the fine-lattice Hamiltonian (except for site-dependent coefficients), any algorithm capable of resampling the fine-lattice field ϕ—including MGMC itself—can also be used to resample the coarse-lattice field ψ. The only difference is that the coarse lattice has a factor of 2^d fewer points, and so the computer work on the coarse lattice will be a factor of 2^d less.

In the MGMC algorithm this idea is carried out recursively at all levels. Let L be, for simplicity, a power of 2. The MGMC resampling of the L^d lattice consists of the following operations: First perform m_1 heat-bath sweeps on the L^d lattice; then compute the conditional Hamiltonian for the $(L/2)^d$ lattice and perform γ MGMC resamplings of this field; then add the result back to the L^d field and perform m_2 heat-bath sweeps on the L^d lattice. Note that for each heat-bath sweep on the L^d lattice, the algorithm performs γ heat-bath sweeps on the $(L/2)^d$ lattice, γ^2 sweeps on the $(L/4)^d$ lattice, and so on. Provided that $\gamma < 2^d$, the total work for one MGMC cycle is of order L^d, *independent of the number of levels—i.e.,* it is only a constant factor $(1-\gamma 2^{-d})^{-1}$ more expensive than m_1+m_2 heat-bath sweeps on the finest lattice alone. The most common cases are $\gamma=1$ ("V cycle") and $\gamma=2$ ("W cycle"). The order of operations is the same as for ordinary MG[7]: Just substitute "heat-bath sweep" for "smoothing sweep".

There is an alternate way of looking at MGMC that gives additional insight.[10] Let A be a subset of the lattice sites, define the one-dimensional variable

$$\psi = |A|^{-1} \sum_{x \in A} \phi_x,$$

and let π be the complementary variables. To resample ψ is to propose a move $\phi \to \phi + t\chi_A$, compute the conditional Hamiltonian $H_{\text{cond}}(t|\phi) = H(\phi + t\chi_A)$, and choose t with the probability density $\sim \exp[-H_{\text{cond}}(t|\phi)]$. The reader can convince herself that MGMC is mathematically equivalent to taking A to run through single-element subsets, cubes of side 2, cubes of side 4, etc., in an order that depends on m_1, m_2, and γ. However, the work required to perform an MGMC cycle in this way is of the order of $L^d \log L$ for $\gamma=1$ and $L^{d+\log_2\gamma}$ for $\gamma \geq 2$.

For the Gaussian (free field) case $P(\phi_x) = m^2\phi_x^2/2$, a rigorous analysis[5] shows that critical slowing down is completely eliminated (for suitably chosen m_1 and m_2 and for $\gamma \geq 2$): that is, the autocorrelation time τ of the MGMC method is *bounded* as $m \to 0$ (criticality). Indeed, the behavior of MGMC for Gaussian fields is completely governed by the behavior of the corresponding deterministic multigrid method for solving linear finite-difference equations. More generally, consider any quadratic Hamiltonian $H(\phi) = \frac{1}{2}(\phi, K\phi) - (b, \phi)$ and any linear stochastic iteration of the form

$$\phi^{(n+1)} = M\phi^{(n)} + Nb + T\xi^{(n)}, \tag{5}$$

where M, N, and T are fixed matrices and the $\xi^{(n)}$ are collections of independent mean-zero Gaussian random variables. Methods based on heat-bath resamplings (including MGMC) are of this form. Averaging over the ξ variables we find[4,5] that the connected time-autocorrelation function is

$$\begin{aligned} C_{xx'}(n) &= \langle \phi_x^{(n)} \phi_{x'}^{(0)} \rangle - \langle \phi_x^{(n)} \rangle \langle \phi_{x'}^{(0)} \rangle \\ &= (M^n K^{-1})_{xx'}. \end{aligned} \tag{6}$$

That is, the matrix M determines the autocorrelation functions of the Monte Carlo algorithm. Another way to state this relationship is that the transition probability $P(\phi^{(n)} \to \phi^{(n+1)})$ induces on the Fock space $L^2(R^\Lambda, d\mu)$ an operator $\Gamma(M)$ that is the second quantization[11] of the operator M on the "energy Hilbert space" (R^Λ, K), from which it follows that

$$\| \Gamma^n(M)|_{1^\perp} \|_{L^2(d\mu)} = \| M^n \|_{(R^\Lambda, K)}.$$

Now, to each Monte Carlo algorithm of the form (5) for generating Gaussian random fields, there corresponds a deterministic algorithm of the form

$$\phi^{(n+1)} = M\phi^{(n)} + Nb \tag{7}$$

for solving the linear equation $K\phi = b$. For example, the deterministic analog of the single-site heat-bath process is the single-site Gauss-Seidel[12] process. (One way to see this is to imagine what would happen if all the random numbers $\xi^{(n)}$ were zero.) For the operator $K = -\Lambda + m^2$ on a periodic lattice, the spectral radius of the Gauss-Seidel matrix M_{GS} is approximately[12] $1 - \text{const} \times m^2$, and so the autocorrelation time of the heat-bath process is of order m^{-2}, a well-known result. Similarly, the autocorrelation time of the MGMC process is determined by the spectral radius of M_{MG}, the iteration matrix of the corresponding deterministic multigrid method ("Galerkin with piecewise-constant injection"). A multigrid convergence theorem[13] states that $\| M_{\text{MG}} \|_{(R^\Lambda, K)} \leq \text{const} < 1$ with a constant independent of m and L. Hence there is no critical slowing down. The successive over-relaxation Monte Carlo method of Adler and Whitmer[14] can also be analyzed in this way, showing that their optimal ω is the same as the optimal ω for ordinary successive over-relaxation.[12]

Our numerical experiments on the two-dimensional ϕ^4 theory show that MGMC does *not* eliminate critical slowing down for models with a double-well (nonconvex) Hamiltonian. Figures 1 and 2 show our results at $\alpha=1$, $\lambda=0.1$, $h=0$ on a 128×128 periodic lattice with $m_1=m_2=1$ and $\gamma=2$ (W cycle). The heat-bath

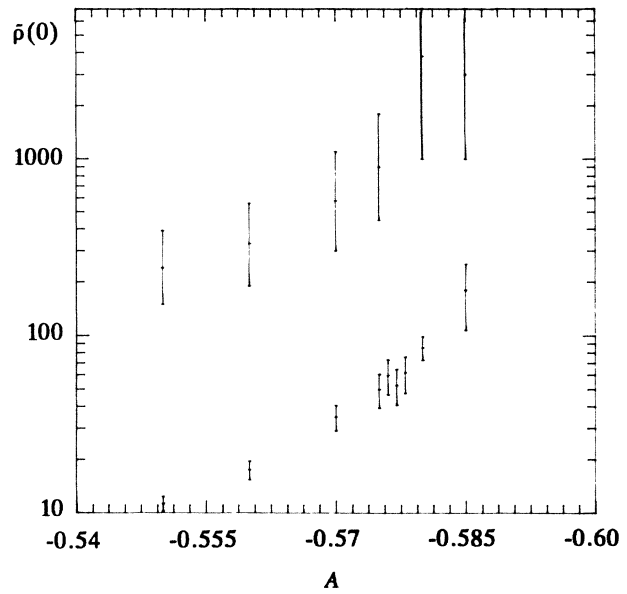

FIG. 1. Autocorrelation time $\tilde{\rho}(0)$ as a function of bare mass squared A. Lower trace is MGMC algorithm; upper trace is single-site heat-bath algorithm. Error bars are ± 1 standard deviation. Heat-bath error bars are highly subjective.

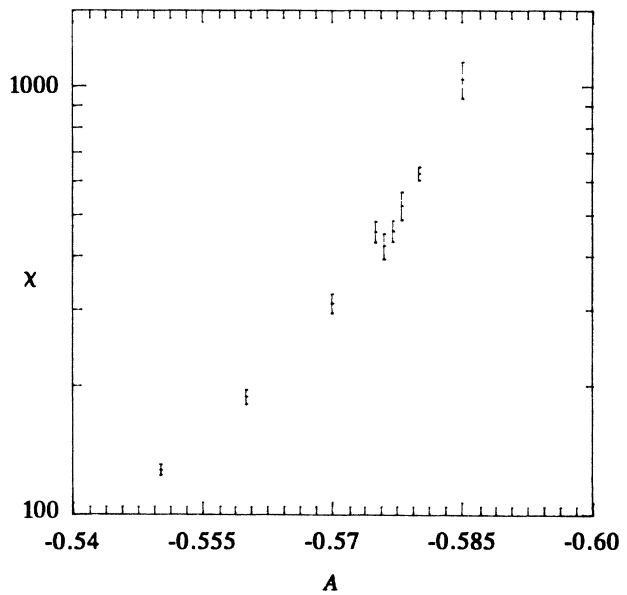

FIG. 2. Susceptibility χ as a function of bare mass squared A. Error bars are ± 1 standard deviation.

sweeps used the red-black ordering[7] of the sites. As a measure of the autocorrelation time we use[15]

$$\tilde{\rho}(0) = \sum_{t=-\infty}^{\infty} \rho(t),$$

where $\rho(t)$ is the normalized time-autocorrelation function of the total magnetization $\sum_x \phi_x$. (This is typically one of the slowest modes of the system.) Estimates and error bars for $\tilde{\rho}(0)$ and the susceptibility χ are computed by standard procedures of statistical time-series analysis.[16] Each run was 10000 MGMC iterations (except $A = -0.58$, which was 30337 iterations), and the first $\approx 20\tau$ iterations were discarded from the analysis. Results from a single-site heat-bath algorithm ($m_1 = m_2 = 1$, $\gamma = 0$) are shown in Fig. 1 for purposes of comparison, but the error bars are relatively large (and highly subjective) precisely because the critical slowing down in the heat-bath case is so severe.[17] The MGMC algorithm is seen to have an autocorrelation time roughly 20 times smaller than the heat-bath algorithm. Since each MGMC iteration takes twice the work of a heat-bath iteration, the gain in efficiency is a factor of ≈ 10. Preliminary runs at $\lambda \approx 1.0$ show a significant but probably lesser advantage for MGMC. Indeed, heuristic arguments[4] based on the double-well nature of the ϕ^4 potential indicate that the efficiency gain near criticality should approach a constant factor $F(\lambda)$; here $F(\lambda)$ is a decreasing function of λ which approaches $+\infty$ as $\lambda \to 0$, in accordance with the absence of critical slowing down for Gaussian MGMC.

There are many variants of the MGMC algorithm. We have used heat-bath updating, but Metropolis updating could also be used. (The hit size would have to be adjusted as a function of the level and perhaps also of the Hamiltonian.) We have used piecewise-constant injection for the coarse-grid correction, but higher-order interpolations (e.g., piecewise linear) could also be used, at the expense of some extra complexity (particularly in the non-Gaussian case). We note that the red-black ordering is particularly well adapted to vector or parallel processing.

There is a version of MGMC that applies to the plane-rotator model[4]; we add a constant angle to each spin in a 2^d block. Numerical experiments with MGMC on a $d = 2$ plane rotator are now being conducted by Edwards.[6] Zwanziger has found an MGMC method for the U(1) lattice gauge theory,[4] but we have not yet found one for non-Abelian gauge theories.

We emphasize that the conditional coarse-lattice Hamiltonian employed in the MGMC method is *not* the same as the renormalized Hamiltonian given by a block-spin renormalization-group (RG) transformation. The RG transformation computes the *marginal*, not the conditional, distribution of the block means—that is, it *integrates* over the complementary degrees of freedom (the π variables) rather than fixing them. Unlike our conditional Hamiltonian, the marginal Hamiltonian cannot be computed in closed form. Perhaps MGMC could be made more effective if it could be modified to resemble more closely the RG

transformation.

A related "collective mode" approach to reducing critical slowing down was proposed recently by Batrouni *et al.*[18]: It uses a Langevin equation and a fast Fourier transform. MGMC provides an alternative and possibly more efficient way of solving the Langevin equation. A collective-mode Monte Carlo algorithm for the Potts models (including Ising) has been proposed recently by Swendsen;[19] since MGMC, as currently formulated, works only for continuous spins, the two approaches are complementary. Ideas somewhat related to MGMC have been proposed by Schmidt[20] and by Chorin.[21]

Details of this work will appear subsequently.[4-6]

This work grew out of discussions with Dan Zwanziger, to whom we gratefully record our thanks. The computations reported here were performed on the Elxsi 6400 computer at the U.S. Department of Energy Courant Mathematics and Computing Laboratory; the computer programs are available from the authors. This work was supported in part by National Science Foundation Grants No. DMS-8400955 and No. DMS-8501953.

---

[1]*Monte Carlo Methods in Statistical Physics,* edited by K. Binder (Springer, Berlin, 1979); *Applications of the Monte Carlo Method in Statistical Physics,* edited by K. Binder (Springer, Berlin, 1984); D. Stauffer, J. Appl. Phys. **53**, 7980 (1982); K. Binder, J. Comput. Phys. **59**, 1 (1985).

[2]M. Creutz, L. Jacobs, and C. Rebbi, Phys. Rep. **95**, 201 (1983), and in *Lattice Gauge Theories and Monte Carlo Simulations,* edited by C. Rebbi (World Scientific, Singapore, 1983).

[3]Each block of data of length $\approx 2\tau$ can be considered, roughly speaking, to contribute one "statistically independent" datum point. Therefore, the "effective sample size" from a run of N sweeps is $\approx N/2\tau$. For a more detailed treatment, see K. Binder, in *Monte Carlo Methods in Statistical Physics,* edited by K. Binder (Springer, Berlin, 1979), Sects. 1.2.3 and 1.2.4, and A. Berretti and A. D. Sokal, J. Stat. Phys. **40**, 483 (1985), Sect. 4.1.

[4]J. Goodman, A. D. Sokal, and D. Swanziger, to be published.

[5]J. Goodman and A. D. Sokal, to be published.

[6]R. Edwards, J. Goodman, A. D. Sokal, and D. Zwanziger, to be published.

[7]A. Brandt, Math. Comp. **31**, 333 (1977), and in *Multigrid Methods,* edited by W. Hackbusch and U. Trottenberg, Lecture Notes in Mathematics Vol. 960 (Springer, Berlin, 1982); W. Hackbusch, *Multi-Grid Methods and Applications* (Springer, Berlin, 1985).

[8]The resemblance between the multigrid method and the renormalization group has often been noted; see e.g., O. A. McBryan, Physica (Amsterdam) **124A**, 481 (1984), p. 488.

[9]The idea of "collective-mode Monte Carlo" was proposed by M. Kalos, in Proceedings of the Brookhaven Conference on Monte Carlo Methods and Future Computer Architectures," May 1983 (unpublished). Closely related ideas have been proposed by G. Parisi, in *Progress in Gauge Field Theory,* edited by G.'t Hooft *et al.,* NATO Advanced Study Institute Series B, Vol. 115 (Plenum, New York, 1984) and by G. G. Batrouni *et al.,* Phys. Rev. D **32**, 2736 (1985). Parisi also gives a brief sketch of a multigrid Monte Carlo method, but we differ with him in important details.

[10]For the deterministic multigrid method, the corresponding perspective is the "unigrid" viewpoint: See S. F. McCormick and J. W. Ruge, Math. Comp. **41**, 43 (1983).

[11]E. Nelson, in *Constructive Quantum Field Theory,* edited by G. Velo and A. Wightman, Lecture Notes in Physics Vol. 25 (Springer, Berlin, 1973); B. Simon, *The $P(\phi)_2$ Euclidean (Quantum) Field Theory* (Princeton Univ. Press, Princeton, N. J., 1974), Chap. I.

[12]R. S. Varga, *Matrix Iterative Analysis* (Prentice-Hall, Englewood Cliffs, N. J., 1962).

[13]For a review of multigrid convergence theorems, see W. Hackbusch, in *Multigrid Methods,* edited by W. Hackbusch and U. Trottenberg, Lecture Notes in Mathematics Vol. 960 (Springer, Berlin, 1982), and *Multi-Grid Methods and Applications* (Springer, Berlin, 1985). A convergence theorem for the case of Galerkin with piecewise-constant injection can be found in Ref. 5.

[14]S. L. Adler, Phys. Rev. D **23**, 2901 (1981); C. Whitmer, Phys. Rev. D **29**, 306 (1984).

[15]It is this quantity which determines the statistical error bars: See Binder, Ref. 3, Eq. (1.35), or Berretti and Sokal, Ref. 3, Eqs. (4.4)–(4.6). Crudely speaking, $\tilde{\rho}(0) \approx 2\tau$.

[16]M. B. Priestley, *Spectral Analysis and Time Series,* (Academic, London, 1981), Chaps. 5–7. We used a truncation window of width $\approx 4\tau$ in all cases.

[17]It is known rigorously that the autocorrelation time of a single-site heat-bath algorithm is bounded below by a constant (presumably of order 1) times the susceptibility, and so the susceptibility curve can be used as a crude guide to the probable critical slowing down of the heat-bath algorithm. See, e.g., B. I. Halperin, Phys. Rev. B **8**, 4437 (1973), and references cited therein. Further results of this kind are proven in Ref. 5.

[18]Batrouni *et al.,* Ref. 9.

[19]R. H. Swendsen, private communication.

[20]K. E. Schmidt, Phys. Rev. Lett. **51**, 2175 (1983); M. Faas and H. J. Hilhorst, Physica (Amsterdam) A (to be published).

[21]A. J. Chorin, Commun. Math. Phys. **99**, 501 (1985).