PHYSICAL REVIEW
LETTERS

VOLUME 72 | 10 JANUARY 1994 | NUMBER 2

A General Limitation on Monte Carlo Algorithms of Metropolis Type

Sergio Caracciolo*
Scuola Normale Superiore and Istituto Nazionale di Fisica Nucleare–Sezione di Pisa, Piazza dei Cavalieri, Pisa 56100, Italy

Andrea Pelissetto† and Alan D. Sokal‡
Department of Physics, New York University, 4 Washington Place, New York, New York 10003
(Received 30 July 1993)

We prove that for any Monte Carlo algorithm of Metropolis type, the autocorrelation time of a suitable "energy"-like observable is bounded below by a multiple of the corresponding "specific heat." This bound does not depend on whether the proposed moves are local or nonlocal; it depends only on the distance between the desired probability distribution π and the proability distribution $\pi^{(0)}$ for which the proposal matrix satisfies detailed balance. We show, with several examples, that this result is particularly powerful when applied to nonlocal algorithms.

PACS numbers: 02.70.Lq, 02.50.Ga, 05.50.+q, 11.15.Ha

Forty years ago, Metropolis *et al.* [1] introduced a general method for constructing dynamic Monte Carlo algorithms (= Markov chains [2]) that satisfy detailed balance for a specified probability distribution π. In this Letter we would like to point out a general limitation on *all* algorithms of Metropolis type. We prove that the autocorrelation time of a suitable "energy"-like observable is bounded below by a multiple of the corresponding "specific heat." This bound does not depend on whether the proposed moves are local or nonlocal; it depends only on the distance between the desired probability distribution π and the probability distribution $\pi^{(0)}$ for which the proposal matrix satisfies detailed balance.

Let us begin by recalling the general Metropolis *et al.* [1] method, as slightly generalized by Hastings [3]. We use the notation of a discrete (finite or countably infinite) state space S, but the same considerations apply with minor modifications to a general measurable state space. Let $P^{(0)}=\{p_{xy}^{(0)}\}$ be an arbitrary transition matrix on S. We call $P^{(0)}$ the *proposal matrix,* and use it to generate proposed moves $x\to y$ that will then be accepted or rejected with probabilities a_{xy} and $1-a_{xy}$, respectively. If a proposed move is rejected, we make a "null transition" $x\to x$. The transition matrix $P=\{p_{xy}\}$ of the full algorithm is thus

$$p_{xy}=\begin{cases} p_{xy}^{(0)}a_{xy} & \text{for } x\neq y\,, \\ p_{xx}^{(0)}+\sum_{z\neq x} p_{xz}^{(0)}(1-a_{xz}) & \text{for } x=y\,, \end{cases} \tag{1}$$

where of course we must have $0\le a_{xy}\le 1$ for all x,y. It is easy to see that P satisfies detailed balance for π if and only if

$$\frac{a_{xy}}{a_{yx}}=\frac{\pi_y p_{yx}^{(0)}}{\pi_x p_{xy}^{(0)}} \tag{2}$$

for all pairs $x\neq y$. But this is easily arranged: Just set

$$a_{xy}=F\left(\frac{\pi_y p_{yx}^{(0)}}{\pi_x p_{xy}^{(0)}}\right), \tag{3}$$

where F: $[0,+\infty]\to[0,1]$ is any function satisfying

$$\frac{F(z)}{F(1/z)}=z \quad \text{for all } z\,. \tag{4}$$

The choice suggested by Metropolis *et al.* [1] is

$$F_{\text{Metr}}(z)=\min(z,1)\,. \tag{5}$$

Other choices of F are possible, but it is easy to see that they must all satisfy the *inequality*

$$F(z)\le\min(z,1)\,. \tag{6}$$

0031-9007/94/72(2)/179(4)$06.00

Of course, it is still necessary to check that P is irreducible (= ergodic); this is usually straightforward.

Note that if the proposal matrix $P^{(0)}$ happens to *already* satisfy detailed balance for π, then we have $\pi_y p_{yx}^{(0)}/\pi_x p_{xy}^{(0)} = 1$, so that $a_{xy} = 1$ (if we use the Metropolis choice of F) and $P = P^{(0)}$. On the other hand, no matter what $P^{(0)}$ is, we obtain a matrix P that satisfies detailed balance for π. So the Metropolis procedure can be thought of as a prescription for minimally modifying a given transition matrix $P^{(0)}$ so that it satisfies detailed balance for π.

Let us now assume that $P^{(0)}$ satisfies detailed balance for some probability measure $\pi^{(0)}$; in practice this is virtually always the case. We then define an energylike observable H by

$$H(x) = \begin{cases} -\log(\pi_x/\pi_x^{(0)}) & \text{if } \pi_x > 0\,, \\ +\infty & \text{if } \pi_x = 0\,. \end{cases} \tag{7}$$

The point is that H is the "energy" of the probability distribution π *relative* to $\pi^{(0)}$.

The heart of our argument is the following upper bound on the mean-square change in energy in a single step of the Metropolis algorithm.

Proposition.—In the situation described above, we always have

$$\langle (\Delta H)^2 \rangle \equiv \sum_{x,x'} \pi_x p_{xx'} [H(x') - H(x)]^2 \le \frac{8}{e^2} f_+ \le \frac{8}{e^2}\,, \tag{8}$$

where

$$f_+ \equiv \sum_{\substack{x,x' \\ H(x') > H(x)}} \pi_x p_{xx'}^{(0)} \le 1 \tag{9}$$

is the fraction (in equilibrium) of proposals that would strictly increase the energy.

Proof.—Since P satisfies detailed balance for π, the summand in (8) is symmetric under $x \leftrightarrow x'$. Therefore it suffices to consider the terms for which $H(x') > H(x)$, and to multiply the result by 2. [The terms having $H(x') = H(x)$ of course make no contribution to the sum.]

If $H(x') > H(x)$, we have $a_{xx'} \le e^{-[H(x') - H(x)]}$ by (3) and (6). Therefore

$$\sum_{\substack{x,x' \\ H(x') > H(x)}} \pi_x p_{xx'} [H(x') - H(x)]^2 = \sum_{\substack{x,x' \\ H(x') > H(x)}} \pi_x p_{xx'}^{(0)} a_{xx'} [H(x') - H(x)]^2 \le \sum_{\substack{x,x' \\ H(x') > H(x)}} \pi_x p_{xx'}^{(0)} e^{-[H(x') - H(x)]} [H(x') - H(x)]^2 \le \frac{4}{e^2} f_+ \tag{10}$$

since $z^2 e^{-z} \le 4/e^2$ for all $z \ge 0$.

The physical intuition behind this proof is simple: Proposed moves having a large energy change $\Delta H > 0$ have an exponentially small acceptance proability, so the mean-square energy increase $\langle (\Delta H)_+^2 \rangle$ in a single Metropolis step is at most of order 1. Proposed moves having an energy change $\Delta H < 0$ are connected to those with $\Delta H > 0$ by detailed balance: When proposed they are accepted, but if $|\Delta H|$ is large they are only rarely proposed. The result is that the mean-square energy change in either direction is at most of order 1.

Let us now recall the definitions of autocorrelation functions and autocorrelation times [4]: If A is a real-valued function defined on the state space S (i.e., a real-valued observable), we define its unnormalized autocorrelation function (in equilibrium) by

$$C_{AA}(t) \equiv \langle A_s A_{s+t} \rangle - \langle A \rangle_\pi^2 \tag{11a}$$

$$= \sum_{x,y} A(x) [\pi_x (P^{|t|})_{xy} - \pi_x \pi_y] A(y)\,. \tag{11b}$$

The corresponding normalized autocorrelation function is

$$\rho_{AA}(t) \equiv C_{AA}(t)/C_{AA}(0)\,. \tag{12}$$

The integrated and exponential autocorrelation times are then defined by

$$\tau_{\text{int},A} = \frac{1}{2} \sum_{t=-\infty}^{\infty} \rho_{AA}(t)\,, \tag{13}$$

$$\tau_{\text{exp},A} = \limsup_{t \to \infty} \frac{|t|}{-\log|\rho_{AA}(t)|}\,, \tag{14}$$

$$\tau_{\text{exp}} = \sup_A \tau_{\text{exp},A}\,. \tag{15}$$

Some simple identities are worth noting:

$$C_{AA}(0) = \langle A^2 \rangle_\pi - \langle A \rangle_\pi^2\,, \tag{16a}$$

$$C_{AA}(1) = C_{AA}(0) - \tfrac{1}{2} \sum_{x,x'} \pi_x p_{xx'} [A(x') - A(x)]^2\,. \tag{16b}$$

Also, from detailed balance combined with the spectral theorem one can deduce the following inequalities:

$$\tau_{\text{int},A} \ge \frac{1}{2} \frac{1 + \rho_{AA}(1)}{1 - \rho_{AA}(1)}\,, \tag{17}$$

$$\tau_{\text{exp}} \ge \tau_{\text{exp},A} \ge -1/\log|\rho_{AA}(1)| \tag{18}$$

(see, e.g., [5], Appendix A).

With these preliminaries, the following theorem is an immediate consequence of the proposition.

Theorem.—Under the preceding hypotheses, we have

$$\tau_{\text{int},H} \ge \frac{e^2}{4} \frac{\text{var}(H)}{f_+} - \frac{1}{2}\,, \tag{19a}$$

$$\tau_{\text{exp}} \ge -1/\log[1 - 4f_+/e^3 \text{var}(H)]\,, \tag{19b}$$

where $\mathrm{var}(H) \equiv \langle H^2 \rangle_\pi - \langle H \rangle_\pi^2$.

Proof.—From the proposition together with (16), we get

$$\rho_{HH}(1) \equiv \frac{C_{HH}(0)}{C_{HH(1)}} \geq 1 - \frac{4}{e^2 \mathrm{var}(H)} . \tag{20}$$

Now use (17) and (18).

Again the physical intuition is simple: The mean-square energy change per Metropolis step is at most of order 1. On the other hand, in order to sample adequately the probability distribution π, the Markov chain must traverse an energy distribution of width $\sim \mathrm{var}(H)^{1/2}$. This takes a time of order $[\mathrm{var}(H)^{1/2}]^2 \sim \mathrm{var}(H)$.

Single-site Metropolis algorithm.—Here $\pi^{(0)}$ is the *a priori* measure for the spins, and H is the full Hamiltonian. $P^{(0)}$ selects a spin at random and proposes to update it in some way that satisfies detailed balance for $\pi^{(0)}$. We have $\mathrm{var}(H) = VC_h$, where V is the volume and C_h is the specific heat. So the theorem shows that

$$\tau_{\mathrm{int},H}, \tau_{\mathrm{exp},H} \gtrsim VC_h , \tag{21}$$

where time is measured here in hits of a single site; or equivalently $\tau \gtrsim C_h$ when time is measured in "sweeps." This is a well-known result. However, it is a rather poor bound because the energy, being a *short*-distance observable, has a rather weak overalp with the slowest (*long*-wavelength) modes of this local dynamics. (A much stronger bound can be obtained by using the magnetization $\mathcal{M}$ rather than the energy as the trial function: one gets $\tau_{\mathrm{int},\mathcal{M}}, \tau_{\mathrm{exp},\mathcal{M}} \gtrsim V\chi$, where χ is the susceptibility [4,6].)

The real power of the theorem comes when it is applied to *nonlocal* algorithms: It still yields $\tau \gtrsim VC_h$, but now the unit of time (a "hit" of $P^{(0)}$) is a nonlocal move which costs a CPU time $\gg 1$. As a result, several algorithms which *a priori* look promising must in fact perform rather poorly:

q-state Potts model with mixed ferromagnetic/antiferromagnetic interaction [7].—The purely ferromagnetic Potts model can be simulated very efficiently by the Swendsen-Wang (SW) algorithm [8,9] or its single-cluster (1CSW) variant [10,11], but these algorithms do not extend easily to the mixed ferromagnetic/antiferromagnetic case. One might therefore try using the SW or 1CSW algorithm for the ferromagnetic part of the Hamiltonian as a Metropolis proposal for the full theory. Thus, let $\pi^{(0)}$ (π) be the Gibbs measure for the ferromagnetic (full) theory, so that H is the antiferromagnetic part of the Hamiltonian. Let $P^{(0)}$ be *any* algorithm that satisfies detailed balance for $\pi^{(0)}$ (for example, SW or 1CSW); and let P be the corresponding Metropolis algorithm for π. One expects $\mathrm{var}(H)$ to behave near criticality as $\sim J_{\mathrm{af}}^2 VC_h$, where J_{af} is the antiferromagnetic coupling. So the theorem shows that

$$\tau_{\mathrm{int},H}, \tau_{\mathrm{exp},H} \gtrsim J_{\mathrm{af}}^2 VC_h , \tag{22}$$

where time is measured here in hits of $P^{(0)}$. For SW (1CSW), each hit takes a CPU time of order V (χ). So the proposed algorithm must perform quite poorly, except when J_{af} is very small [12].

$d=3$ Heisenberg model with topological term [13].—The ferromagnetic Heisenberg model can be simulated very efficiently by the Wolff embedding algorithm [10,14] using either SW of 1CSW moves to update the induced Ising model [15]. The topological term seems difficult to incorporate into the cluster-algorithm framework, but one might try using the SW or 1CSW algorithm for the ferromagnetic two-body part of the Hamiltonian as a Metropolis proposal for the full theory. (The intuitive idea is that a 1CSW move is likely to make a modest change in the topological-charge field, so the acceptance rate should be reasonable.) Thus, let $\pi^{(0)}$ (π) be the Gibbs measure for the ferromagnetic (full) theory, so that H is the topological term. Let $P^{(0)}$ be *any* algorithm that satisfies detailed balance for $\pi^{(0)}$ (for example, SW or 1CSW); and let P be the corresponding Metropolis algorithm for π. One expects $\mathrm{var}(H)$ to behave near criticality as $\sim J_{\mathrm{top}}^2 VC_h$, where J_{top} is the topological coupling [16]; and it is known that $C_h \to \mathrm{const} > 0$ at criticality (since $\alpha < 0$). So the theorem shows that

$$\tau_{\mathrm{int},H}, \tau_{\mathrm{exp},H} \gtrsim J_{\mathrm{top}}^2 V , \tag{23}$$

where time is measured here in hits of $P^{(0)}$. For SW (1CSW), each hit takes a CPU time of order V (χ). So the proposed algorithm must perform quite poorly, except when J_{top} is very small.

Self-avoiding walk with nearest-neighbor interaction.—Fix an integer N, and let S be the space of all N-step self-avoiding walks on some specified lattice. Let $\pi^{(0)}$ be the probability measure that gives equal weight to each element of S. Then define the probability measure π by

$$\pi_\omega = Z(\epsilon)^{-1} e^{-\epsilon M(\omega)} \pi_\omega^{(0)} , \tag{24}$$

where $M(\omega)$ is the number of nonbonded nearest-neighbor contacts in the walk ω. Let $P^{(0)}$ be *any* algorithm that satisfies detailed balance for $\pi^{(0)}$ (e.g., the pivot algorithm [17,18]); and let P be the corresponding Metropolis algorithm for (24). Then the theorem shows that

$$\tau_{\mathrm{int},M}, \tau_{\mathrm{exp},M} \gtrsim \epsilon^2 \mathrm{var}_\pi(M)/f , \tag{25}$$

where f is the fraction of proposals $p_{\omega\omega'}^{(0)}$ with $\omega' \neq \omega$ (e.g., the fraction of proposed pivot moves that preserve self-avoidance). And we expect $\mathrm{var}_\pi(M) \approx NC(\epsilon)$, where the "specific heat per step" $C(\epsilon)$ is everywhere nonzero and diverges like $(\epsilon - \epsilon_\theta)^{-\alpha_\theta}$ at the theta (tricritical) point.

For the pivot algorithm, the bound (25) is a rather weak result: In fact we expect that $\tau_{\mathrm{int},M}, \tau_{\mathrm{exp},M} \sim N/f$ even for $\epsilon = 0$, because M is a "primarily local" observable [18]. But (25) does show that for $\epsilon \neq 0$ (and in particular for $\epsilon \to \epsilon_\theta$) the difficulties cannot be avoided by

using a different proposal $P^{(0)}$; they are inherent in the Metropolis method with this choice of $\pi^{(0)}$ [19].

We conclude by noting that the Metropolis *et al.* method is often applied indirectly: We define transition matrices $P_1, \ldots, P_n$ by the Metropolis method, and we then execute either $P = \sum_{i=1}^{n} \lambda_i P_i$ for some weights $\lambda_i \geq 0$ ("random updating") or else $P = P_1 \cdots P_n$ ("sequential updating"). The first case can easily be handled by our method. The second case is more subtle, because typically P does not satisfy detailed balance [20]; but the bound is almost certainly correct in order of magnitude, except in special situations like "successive overrelaxation" [21].

The authors wish to thank Giovanni Ferraro for a helpful discussion. One of us (A.D.S.) wishes to thank the Scuola Normale Superiore for hospitality while this work was being carried out. The authors' research was supported in part by the Consiglio Nazionale delle Ricerche (S.C. and A.P.), the Istituto Nazionale di Fisica Nucleare (S.C., A.P., and A.D.S.), U.S. Department of Energy Contract No. DE-FG02-90ER40581 (A.D.S.), U.S. National Science Foundation Grants No. DMS-8911273 and No. DMS-9200719 (A.D.S.), and NATO Collaborative Research Grant No. CRG 910251 (S.C. and A.D.S.).

---

*Electronic address (Internet): CARACCIO@UX1SNS.SNS.IT and caraccio@riscle1.le.infn.it

†Address until August 31, 1994 [Electronic address (Internet): PELISSET@MAFALDA.PHYSICS.NYU.EDU]. Permanent address: Dipartimento di Fisica and INFN-Sezione di Pisa, Università degli Studi di Pisa, Pisa 56100, Italy. Electronic address: (Internet): PELISSET@SUNTHPI1.DIFI.UNIPI.IT

‡Electronic address (Internet): SOKAL@ACF4.NYU.EDU

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

[2] J. G. Kemeny and J. L. Snell, *Finite Markov Chains* (Springer, New York, 1976); M. Iosifescu, *Finite Markov Processes and Their Applications* (Wiley, Chichester, 1980); K. L. Chung, *Markov Chains with Stationary Transition Probabilities* (Springer, New York, 1967), 2nd ed.; E. Nummelin, *General Irreducible Markov Chains and Non-Negative Operators* (Cambridge Univ. Press, Cambridge, 1984); D. Revuz, *Markov Chains* (North-Holland, Amsterdam, 1975).

[3] W. K. Hastings, Biometrika **57**, 97 (1970).

[4] A. D. Sokal, *Monte Carlo Methods in Statistical Mechanics: Foundations and New Algorithms*, Cours de Troisième Cycle de la Physique en Suisse Romande, Lausanne, June 1989 (unpublished).

[5] S. Caracciolo, A. Pelissetto, and A. D. Sokal, J. Stat. Phys. **60**, 1 (1990).

[6] K. Kawasaki, Phys. Rev. **148**, 375 (1966); R. Abe and A. Hatano, Prog. Theor. Phys. **41**, 941 (1969); B. I. Halperin, Phys. Rev. B **8**, 4437 (1973).

[7] F. Fucito and G. Parisi, J. Phys. A **14**, L499 (1981); F. Fucito and A. Vulpiani, Phys. Lett. A **89**, 33 (1982); S. Caracciolo, G. Parisi, and S. Patarnello, Europhys. Lett. **4**, 7 (1987); S. Caracciolo and S. Patarnello, Phys. Lett. A **126**, 233 (1988); L. A. Fernández *et al.*, Phys. Lett. B **217**, 309 (1989); L. A. Fernández, U. Marini Bettolo Marconi, and A. Tarancón, Phys. Lett. B **217**, 314 (1989); Physica (Amsterdam) **161A**, 284 (1989); F. R. Brown, Phys. Lett. B **224**, 412 (1989); M. Bernaschi *et al.*, Phys. Lett. B **231**, 157 (1989); R. V. Gavai and F. Karsch, Phys. Lett. B **233**, 417 (1989); M. Bernaschi *et al.*, Phys. Lett. B **240**, 419 (1990); A. Billoire, R. Lacaze, and A. Morel, Nucl. Phys. **B340**, 542 (1990); E. Marinari and R. Marra, Nucl. Phys. **B342**, 737 (1990).

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

[9] X.-J. Li and A. D. Sokal, Phys. Rev. Lett. **63**, 827 (1989).

[10] U. Wolff, Phys. Rev. Lett. **62**, 361 (1989).

[11] U. Wolff, Phys. Lett. B **228**, 379 (1989); P. Tamayo, R. C. Brower, and W. Klein, J. Stat. Phys. **58**, 1083 (1990); C. F. Baillie and P. D. Coddington, Phys. Rev. B **43**, 10617 (1991).

[12] In conventional CPU units of "hits per site" (equivalent to "sweeps" of a local algorithm), we have $\tau \gtrsim VC_h$ (χC_h). In terms of dynamic critical exponents, this means $z_{\text{CPU}} \geq d + [\max(\alpha,0)/\nu]$ ($[\gamma + \max(\alpha,0)]/\nu$).

[13] M-h. Lau and C. Dasgupta, Phys. Rev. B **39**, 7212 (1989).

[14] R. G. Edwards and A. D. Sokal, Phys. Rev. D **40**, 1374 (1989); M. Hasenbusch, Nucl. Phys. **B333**, 581 (1990); S. Caracciolo, R. G. Edwards, A. Pelissetto, and A. D. Sokal, Nucl. Phys. **B403**, 475 (1993).

[15] C. Holm and W. Janke, Phys. Lett. A **173**, 9 (1993); Phys. Rev. B **48**, 936 (1993).

[16] C. Holm and W. Janke, Freie Universität Berlin Report No. FUB-HEP 10/93, 1993, hep-lat@ftp.scri.fsu.edu #9306020.

[17] M. Lal, Mol. Phys. **17**, 57 (1969).

[18] N. Madras and A. D. Sokal, J. Stat. Phys. **50**, 109 (1988).

[19] In addition, we suspect—though we are at present unable to prove—that for $\epsilon \neq 0$ the slowness in M is "transmitted" to all other observables, so that $\tau_{\text{int},A}, \tau_{\text{exp},A} \gtrsim \epsilon^2 N/f$ even for *global* observables such as the radius of gyration.

[20] Of course, $P = P_1 \cdots P_n$ does leave π invariant, which is all that matters.

[21] S. L. Adler, Nucl. Phys. B (Proc. Suppl.) **9**, 437 (1989).