Journal of Statistical Physics, Vol. 40, Nos. 3/4, 1985

New Monte Carlo Method for the Self-Avoiding Walk

Alberto Berretti and Alan D. Sokal[1]

Received October 1, 1984; revised January 31, 1985

We introduce a new Monte Carlo algorithm for the self-avoiding walk (SAW), and show that it is particularly efficient in the critical region (long chains). We also introduce new and more efficient statistical techniques. We employ these methods to extract numerical estimates for the critical parameters of the SAW on the square lattice. We find

$$\mu = 2.63820 \pm 0.00004 \pm 0.00030$$

$$\gamma = 1.352 \pm 0.006 \pm 0.025$$

$$\nu = 0.7590 \pm 0.0062 \pm 0.0042$$

where the first error bar represents systematic error due to corrections to scaling (subjective 95% confidence limits) and the second bar represents statistical error (classical 95% confidence limits). These results are based on SAWs of average length ≈ 166, using 340 hours CPU time on a CDC Cyber 170-730. We compare our results to previous work and indicate some directions for future research.

KEY WORDS: Self-avoiding walk; polymer; lattice model; critical exponents; Monte Carlo; algorithm; maximum-likelihood estimation.

1. INTRODUCTION

The self-avoiding walk (SAW) was first proposed nearly half a century ago as a model of a polymer molecule with excluded volume.[1-3] Since then it has been studied extensively by chemical physicists.[4-6] More recently, the SAW has been shown to be equivalent to the $N = 0$ case of the N-vector

[1] Courant Institute of Mathematical Sciences, New York University, 251 Mercer Street, New York, New York 10012.

0022-4715/85/0800-0483$04.50/0

822/40/3-4-8

model,[7–11] a fact which has made it an important test case for the theory of critical phenomena.

In this paper we wish to make a modest contribution to the numerical study of the critical (i.e., long-chain) behavior of SAWs by the Monte Carlo method. Our aim is threefold: (1) to introduce a new and highly efficient algorithm for the Monte Carlo generation of SAWs, (2) to introduce new and more efficient statistical techniques for the analysis of the Monte Carlo data, and (3) to employ these methods to extract numerical estimates for the critical temperature and critical exponents of the SAW.

Let us explain each of these aspects in greater detail.

Our algorithm is a *dynamic* Monte Carlo algorithm which generates SAWs in a *grand canonical* ensemble. It is inspired by, and closely related to, both the "slithering-snake" (reptation) algorithm of Kron, Wall, and Mandel[12–15] and the "chain-deformation" algorithm of Berg, Foerster, Aragão de Carvalho, Caracciolo, and Fröhlich.[16,17,11] Nevertheless, it has important advantages over both of these previous algorithms: it is ergodic, unlike the slithering-snake algorithm[13,14]; and it has an autocorrelation time (measured in elementary computer operations) of order $\langle N \rangle^2$ (here $\langle N \rangle$ is the average number of steps in the chain), as opposed to order $\langle N \rangle^3$ or greater for all known algorithms of chain-deformation type[18,19] (see footnotes 2 and 3). It is this latter fact which makes the new algorithm vastly more efficient in the critical region (large N) than most previous SAW algorithms. In fact, as we explain in Section 3, the new algorithm makes feasible high-precision Monte Carlo studies of critical behavior for the SAW, with an efficiency exceeding that of comparable Ising-model studies by a factor which grows rapidly as one plunges deeper into the critical region. We thus argue that the SAW is a uniquely advantageous "laboratory" for Monte Carlo studies of critical phenomena. The numerical results for the two-dimensional SAW reported in the present paper are intended as an initial illustration of this assertion; some directions for future work are discussed in Section 6.2.

The algorithm introduced here is also closely related to an algorithm of Redner and Reynolds[20]; we discuss the rather subtle relation between the two algorithms in Section 3. The Redner–Reynolds algorithm is, like ours, a grand-canonical algorithm; the two algorithms appear to be of approximately equal efficiency.

[2] Beware: Much of the literature measures time in "bead cycles." One bead cycle equals N (or $N+1$) elementary computer operations.

[3] A more detailed comparison with previous algorithms will appear in a separate paper.[19]

Our second goal in this paper is to introduce new statistical techniques for the analysis of SAW Monte Carlo data. In particular, we show how *maximum-likelihood estimation* can be employed to provide estimates of the critical quantities μ and γ which are not only demonstrably optimal in a rigorous (or almost rigorous) statistical sense (i.e., they achieve the minimum possible mean-square error for a given quantity of Monte Carlo data), but which also provide *a priori* (or partially *a priori*) error estimates. This means that statistical errors can be computed reliably, *in advance* of performing the Monte Carlo simulation. Or to put it more strikingly: before performing the simulation, one cannot know what the final central estimates will be, but one *can* know the error bars! We believe that this is a significant advance. It is, to be sure, made possible by the particularly simple scaling form arising in the SAW [Eq. (2.1)], and so the method is not immediately applicable to Monte Carlo studies of other models; but we hope that this example will spur other workers to meditate more carefully on their statistical techniques and to invent clever new ones.

More generally, we have tried in this paper to take especial care in our statistical analysis, stressing those issues which we feel have been given short shrift in previous work. Such care, while obviously desirable in all cases,[4] is an absolute necessity in Monte Carlo studies as delicate as those of critical phenomena. In particular, we wish to emphasize two points:

(1) The importance, in any work employing a dynamic Monte Carlo algorithm, of a correct statistical treatment of *autocorrelations* and their effects. This should include a theoretical (i.e., pre-simulation) study of the autocorrelation function to the extent that this is possible; it should in any case include a thorough empirical (i.e., post-simulation) statistical analysis, including statistically valid error bars.[5] We hope that our brief discussion in Section 4.1 will make the statistics literature on time-series analysis more widely known to physicists engaged in Monte Carlo work.

(2) The distinction between *systematic errors* (errors resulting from misspecification of the mathematical model on which the data analysis is

[4] As Wood and Erpenbeck[21] note, "these [Monte Carlo] investigations share some of the features of ordinary experimental work, in that they are susceptible to both statistical and systematic errors. With regard to these matters, we believe that papers should meet much the same standards as are normally required for experimental investigations. We have in mind the inclusion of estimates of statistical error, descriptions of experimental conditions (i.e., parameters of the calculation), relevant details of apparatus (program) design, comparisons with previous investigations, discussion of systematic errors, etc. Only if these are provided will the results be trustworthy guides to improved theoretical understanding."

[5] For the new algorithm proposed here, the theoretical part of this analysis is carried out in Section 3 and Appendix A; the empirical part is carried out in Section 5.2, using the methods described in Section 4.1.

based, e.g., due to unincluded corrections to scaling) and *statistical errors* (the random fluctuation inherent in any probabilistic experiment). The two types of errors play very different roles and must be treated separately (something which naive least-squares curve-fitting fails to do). The treatment of corrections to scaling is difficult and subtle—and we ourselves have not been entirely successful (compare Sections 5.3 and 5.4)—but it is unavoidable in any quantitative study of critical behavior, whether by Monte Carlo or any other method.

The plan of this paper is as follows: In Section 2 we review briefly those facts and conjectures about the SAW that will be needed in the remainder of the paper. In Section 3 we introduce the new algorithm and analyze some of its properties. In Section 4 we describe the statistical techniques needed for the data analysis. In Section 5 we present our numerical results for the two-dimensional SAW. In Section 6 we compare our results with previous work and indicate some directions for future work. In Appendix A we present an exact solution of the dynamics of our algorithm for the case of ordinary random walk. In Appendix B we prove a spectral representation for the autocorrelation function of a reversible Markov chain. In Appendix C we discuss briefly some computer-programming considerations which arise in Monte Carlo studies of SAWs.

2. THE SELF-AVOIDING WALK (SAW)

In this section we review briefly the basic facts and conjectures about the SAW that will be used in the remainder of the paper. Let $\mathscr{L}$ be some regular d-dimensional lattice. Then an *N-step self-avoiding walk* (SAW) ω on $\mathscr{L}$ is a sequence of *distinct* points $\omega(0)$, $\omega(1)$,..., $\omega(N)$ in $\mathscr{L}$ such that each point is a nearest neighbor of its predecessor. Unless stated otherwise, we assume all walks to begin at the origin, i.e., $\omega(0) = 0$.

Let c_N [respectively, $c_N(x)$] be the number of N-step SAWs starting at the origin and ending anywhere (respectively, ending at x). These quantities are believed to have the asymptotic behavior

$$c_N \sim \mu^N N^{\gamma - 1} \tag{2.1}$$

$$c_N(x) \sim \mu^N N^{\alpha_{\text{sing}} - 2} \qquad (x \text{ fixed} \neq 0) \tag{2.2}$$

as $N \to \infty$.[6,7] Here μ is called the *connective constant* (or *effective coor-*

[6] On certain lattices, (2.2) requires a proviso: e.g., on the d-dimensional simple cubic lattice, N must be taken to have the same parity mod 2 that x does, since otherwise $c_N(x) = 0$. But this is a trivial restriction.

[7] The subscript "sing" is attached because the exponent defined in (2.2) is analogous to the exponent for the *singular part* of the specific heat in a spin model. For example, for $d > 4$ one expects $\alpha_{\text{sing}} = 2 - d/2 < 0$. See (2.9).

dination number) of the lattice $\mathscr{L}$, and is lattice-dependent; γ and α_{sing} are *critical exponents*, and are believed to be universal for lattices of a given dimension d.

The mean-square end-to-end distance

$$\langle \omega(N)^2 \rangle \equiv \frac{1}{c_N} \sum_x |x|^2 c_N(x) \tag{2.3}$$

is believed to scale as

$$\langle \omega(N)^2 \rangle \sim N^{2\nu} \tag{2.4}$$

as $N \to \infty$, where ν is another (universal) critical exponent. More generally, the full probability distribution of $\omega(N)$ is believed to scale as

$$\frac{c_N(x)}{c_N} \sim N^{-d\nu} f(x/N^{\nu}) \tag{2.5}$$

as $N \to \infty$, for a suitable *scaling function* f (also universal up to trivial changes of scale). Moreover, f is expected to be rotation-invariant. [Actually, (2.5) is only claimed to hold for $|x|$ on the order of N^{ν}. The precise statement of (2.5) is therefore that the limit

$$f(y) \equiv \lim_{N \to \infty} N^{d\nu} \frac{c_N(N^{\nu} y)}{c_N} \tag{2.6}$$

exists for each $y \neq 0$, with $0 < f(y) < \infty$.]

Finally, let c_{N_1,N_2} be the number of pairs (ω_1, ω_2) such that ω_1 is an N_1-step SAW starting at the origin, ω_2 is an N_2-step SAW starting *anywhere*, and ω_1 and ω_2 have at least one point in common (i.e., $\omega_1 \cap \omega_2 \neq \varnothing$). (This quantity is closely related to the osmotic second virial coefficient for polymer molecules.) Then is it believed that

$$c_{N_1,N_2} \sim \mu^{N_1+N_2} N_1^{2\Delta_4+\gamma-2} g(N_1/N_2) \tag{2.7}$$

as $N_1, N_2 \to \infty$, where Δ_4 is yet another (universal) critical exponent and g is a (universal) scaling function.

We remark that the above-defined critical exponents for the SAW are precise analogues of the critical exponents defined for spin systems.[22] Indeed, the "susceptibility," "two-point correlation function," and "connected four-point function at zero momentum" defined by

$$\chi(\beta) \equiv \sum_{N=0}^{\infty} \beta^N c_N \tag{2.8}$$

$$G(x;\beta) \equiv \sum_{N=0}^{\infty} \beta^N c_N(x) \tag{2.9}$$

$$\bar{u}_4(\beta) \equiv -\sum_{N_1,N_2=0}^{\infty} \beta^{N_1+N_2} c_{N_1,N_2} \tag{2.10}$$

are *equal* to the corresponding quantities in the N-vector model at inverse temperature β, analytically continued to $N=0$.[9-11] The critical inverse-temperature is $\beta_c = 1/\mu$. [Note, however, the clash of terminology: in the polymer language, (2.8)–(2.10) define a *grand canonical* ensemble at monomer activity β; in the spin-system (or field-theory) language, (2.8)–(2.10) define a *canonical* ensemble at inverse temperature β.]

Thus, one can pose for the SAW questions analogous to those posed in the theory of critical phenomena in lattice spin systems. In particular, one can study the conjectured hyperscaling relations

$$d\nu = 2 - \alpha_{\text{sing}} \tag{2.11}$$

and

$$d\nu - 2\Delta_4 + \gamma = 0 \tag{2.12}$$

The latter has a particularly clear heuristic interpretation: Let $p(N)$ be the probability that two N-step SAWs, one starting at the origin and the other starting a distance of order N^ν away from the origin, intersect each other; then (2.12) holds if $p(N) \geqslant c > 0$ as $N \to \infty$. Hyperscaling thus concerns the probability of intersection of two independent SAWs.[8] In this paper we shall not in fact study (2.11) or (2.12)—we limit attention to μ, γ, and ν—but we mention the hyperscaling problem because it is one of the most important open problems in the theory of critical phenomena, and because it was the main motivation for the present work. In future work we plan to apply our Monte Carlo method to study Δ_4 in dimension $d=3$.

Few rigorous results are known for the SAW. Hammersley[28,29] has proven a very weak form of (2.1)/(2.2), namely, that the limits

$$\mu = \lim_{N\to\infty} c_N^{1/N} = \lim_{N\to\infty} c_N(x)^{1/N} \tag{2.13}$$

exist and are equal. Slightly stronger results are due to Kesten.[30,31] Some rigorous upper and lower bounds on μ are known.[32] Aizenman's[33] proof

[8] The analogous problem for *ordinary* random walks has been solved by the combined work of Erdös and Taylor[23] and Lawler.[24] These results have recently been rederived by Felder and Fröhlich[25] and by Aizenman,[26] using an elegant rigorous version of the field-theoretic renormalization group. The intersection properties of M independent ordinary random walks can be related[27] to the $N \to 0$ limit of a model of M coupled N-component fields.

of hyperscaling for the two-dimensional Ising model extends immediately to prove the "difficult" half of (2.12), i.e., $d\nu - 2\Delta_4 + \gamma \leqslant 0$, for the two-dimensional SAW. (This is our reason for not undertaking a Monte Carlo study of Δ_4 in $d=2$.) One of the authors[34] has proven for the SAW a strengthening of the Aizenman–Fröhlich[33,35,11,36] correlation inequality; it follows from this result that

$$0 \leqslant -\bar{u}_4 \leqslant \left[\frac{\partial(\beta\chi)}{\partial\beta}\right]^2 \tag{2.14}$$

and hence that the hyperscaling relation (2.12) *fails* whenever $d\nu > 2$. [This makes rigorous one half of an appealing intuitive argument due to des Cloizeaux[37]: by (2.4)/(2.5) one expects a single long SAW to act like an object of "fractal dimension"[38] $1/\nu$; so two independent such objects should intersect "generically" if and only if the dimension d of the ambient space is less than or equal to $1/\nu + 1/\nu$, i.e., if $d\nu \leqslant 2$.] Very recently, Brydges and Spencer[39] have proven $\gamma = 1$ and $\nu = 1/2$ for a slightly modified SAW in dimension $d > 4$, using rigorous renormalization-group methods.

The hyperscaling problem for the SAW (like other models) in dimension $2 < d < 4$ appears, therefore, to be inaccessible to currently available rigorous methods. For this reason we turn to Monte Carlo.

3. NEW MONTE CARLO ALGORITHM FOR THE SELF-AVOIDING WALK

In this section we describe a new Monte Carlo algorithm for the self-avoiding walk, and discuss some of its properties. A detailed comparison with other SAW Monte Carlo algorithms can be found in a separate paper.[19]

Our algorithm generates SAWs with one endpoint anchored at the origin and the other endpoint free, in a *grand canonical* ensemble at monomer activity β. Thus each N-step SAW has probability $\beta^N/\Xi(\beta)$ of occurring in the ensemble, where

$$\Xi(\beta) = \sum_{N=0}^{\infty} \beta^N c_N \tag{3.1}$$

is the grand partition function.[9] The algorithm is a *dynamic* Monte Carlo algorithm: that is, it is a stochastic (in fact Markovian) dynamics which has the desired ensemble as its unique equilibrium probability distribution.

[9] Note that from the spin-system point of view $\Xi(\beta)$ is the susceptibility; cf. (2.8).

The algorithm's "elementary moves" are as follows: either one attempts to append a new bond to the walk, with equal probability in each of the q possible directions (here q is the coordination number of the lattice); or else one deletes the last bond from the walk. In the former case, one must check that the attempted new step does not violate the self-avoidance constraint; if it does, then the attempted move is rejected and the old configuration is counted again in the sample (a "null transition"). If an attempt is made to delete a bond from an already-empty walk, then a null transition is also made. The relative probabilities of $\Delta N = +1$ and $\Delta N = -1$ attempts are chosen to be

$$P(\Delta N = +1 \text{ attempt}) = \frac{q\beta}{1+q\beta} \tag{3.2}$$

$$P(\Delta N = -1 \text{ attempt}) = \frac{1}{1+q\beta} \tag{3.3}$$

It follows that the transition probability from a walk ω to a walk ω' is given by

$$p(\omega \to \omega') = \begin{cases} \dfrac{\beta}{1+q\beta}\, \chi_{\mathrm{SAW}}(\omega') & \text{if } \omega \prec \omega' \\ \dfrac{1}{1+q\beta} & \text{if } \omega' \prec \omega \text{ or } \omega = \omega' = \varnothing \\ \dfrac{\beta}{1+q\beta}\, A(\omega) & \text{if } \omega = \omega' \neq \varnothing \end{cases} \tag{3.4}$$

where

$$\chi_{\mathrm{SAW}}(\omega') = \begin{cases} 1 & \text{if } \omega' \text{ is an SAW} \\ 0 & \text{if } \omega' \text{ is not an SAW} \end{cases} \tag{3.5}$$

Here $\omega \prec \omega'$ denotes that the walk ω' is obtained by appending one bond onto the end of ω; and $A(\omega)$ is the number of *non-self-avoiding* walks ω' with $\omega \prec \omega'$. It is easily verified that (3.4) satisfies the detailed-balance condition for the grand canonical ensemble with monomer activity β: that is,

$$\pi(\omega)\, p(\omega \to \omega') = \pi(\omega')\, p(\omega' \to \omega) \tag{3.6}$$

for all walks ω, ω', where

$$\pi(\omega) = \frac{\beta^{|\omega|}}{\Xi(\beta)}\, \chi_{\mathrm{SAW}}(\omega) \tag{3.7}$$

and $|\omega|$ is the number of bonds in the walk ω. The detailed-balance condition (3.6) ensures that the probability distribution (3.7) is a stationary distribution for the Markov process defined by (3.4). To ensure that it is the *unique* stationary distribution, it suffices to verify that the process (3.4) is *ergodic*, i.e., that one can get from any SAW ω to any other SAW ω' by a finite sequence of allowed moves. But this is easy: one simply uses the $\Delta N = -1$ moves to "eat up" all the bonds of the walk ω until the empty walk $\varnothing$ is reached, and then uses the $\Delta N = +1$ moves to build up step-by-step the walk ω'. From the detailed-balance and ergodicity conditions it then follows, by the general theory of Markov chains,[40] that the occupation-time distribution over a long time interval converges with probability 1 to the equilibrium distribution π, irrespective of the initial state. In fact, since the process (3.4) is *aperiodic*, the probability distribution at any single time in the far future also converges to π. Thus, simulation of the Markov process defined by (3.4) provides a legitimate Monte Carlo method for estimating averages with respect to π.

The successive states in this Markov process are, of course, highly correlated; it takes a while for the process to "lose memory" of its current configuration. This means that the variance of Monte Carlo estimates produced by the dynamic algorithm may be much higher than would be the case if one could produce *independent* samples from the distribution π. Crudely speaking, successive blocks of data of width $\approx 2\tau$ can be considered to be "effectively independent", where τ is the autocorrelation time (in equilibrium) of the dynamic process. (A more rigorous discussion is given in Section 4.1.) It is thus of some importance to know how big τ is. We argue that, in the process defined by (3.4),

$$\tau \sim \langle N \rangle^2 \tag{3.8}$$

To see this, consider the quantity $N(t) = |\omega|(t)$, the number of bonds in the walk at time t. This quantity executes, crudely speaking, a random walk (with drift) on the nonnegative integers; the average time required to go from some point N to the point 0 (i.e., the empty walk) is of order N^2. Moreover, each time the empty walk is reached, all memory of the past is erased; future walks are then independent of past ones. Thus, the autocorrelation time ought to be of order $\langle N^2 \rangle$, or equivalently $\langle N \rangle^2$. This is borne out by an exact solution of the dynamics for the case of the ordinary random walk (Appendix A) and by our Monte Carlo results for the SAW (Section 5.2). The dynamic critical exponent in (3.8) is the same as in the slithering-snake (reptation) algorithm,[12–15] and less than in all other known dynamic Monte Carlo algorithms for the SAW[11,17,18]; for further discussion, see Ref. 19.

One may thus obtain an ensemble of SAWs of any desired average length by choosing β appropriately, with

$$0 \leqslant \beta < \beta_c \equiv \frac{1}{\mu} \tag{3.9}$$

For β near to β_c we have, by (2.1),

$$\langle N \rangle \equiv \frac{\sum_{N=0}^{\infty} N\beta^N c_N}{\sum_{N=0}^{\infty} \beta^N c_N} \approx \frac{\beta\gamma}{\beta_c - \beta} \tag{3.10}$$

In Section 4 we discuss how to obtain valid statistical estimates of μ, γ and ν from such an ensemble.

Remarks. (1) The foregoing algorithm can be made slightly more efficient by restricting the $\Delta N = +1$ attempts to those bond directions which are not immediate reversals of the preceding bond. See Ref. 19 for details. In the present paper we do not make use of this improvement.

(2) The foregoing algorithm is closely related, but *not* identical, to the algorithm of Redner and Reynolds.[20] Both algorithms generate SAWs in a grand canonical ensemble with one free end point. However, the conceptual foundations of the two algorithms are quite different. The present algorithm is of "dynamic" type, i.e., it generates a sample from an asymptotically stationary stochastic process over the space of all SAWs. By contrast, the Redner–Reynolds algorithm is most naturally thought of as a stochastic generalization of exact enumeration, in which each step in the "decision tree" is taken with probability $p < 1$ (for $p = 1$ it reduces to exact enumeration). Thus, run for an integer number of "full cycles," the Redner–Reynolds algorithm generates a sample consisting of independent batches of SAWs, with the correlations among the SAWs within a batch being quite complicated (and not necessarily those of an asymptotically stationary stochastic process). Such an algorithm is said to be of "quasistatic" type. In fact, the Redner–Reynolds algorithm can alternatively be viewed[19] as a variant of the $s = 1$ special case of the enrichment algorithm,[41] which is likewise of quasistatic type. Nevertheless (as we discovered after the completion of this work), it is possible to give the Redner–Reynolds and enrichment algorithms yet another (though less natural) interpretation, in this case as a dynamic algorithm *over a larger state space*; see Ref. 19 for details. The upshot is that all three algorithms are somewhat related; in particular, the Redner–Reynolds and enrichment algorithms probably share the order-N^2 computation time which is characteristic of the algorithm described here. We find the Redner–Reynolds algorithm quite interesting and deserving of further study. (We thank Sid Redner for helpful conversations and correspondence on these points.)

(3) It is perhaps worth comparing our algorithm for the SAW to the standard single-spin-flip Metropolis (or heat-bath) algorithm for the Ising model. Both algorithms have fairly simple, hence rapidly executed, elementary moves. Monte Carlo simulations of spin systems must contend with finite-size effects. To avoid severe systematic errors, one must take $L \gtrsim \xi$, where L is the linear size of the system and ξ is the correlation length in an infinite-volume system at the same temperature. One can then use finite-size scaling theory[42] to extrapolate the results to $L = \infty$, but this extrapolation is extremely delicate (one must allow, for example, for corrections to scaling) and significant systematic errors are likely to remain. Moreover, the standard dynamic algorithms exhibit critical slowing down: the autocorrelation time diverges as $\tau_{\text{phys}} \sim \min(L, \xi)^z$, where z is a dynamic critical exponent ($z \approx 2$ for the usual kinetic Ising model in all dimensions[43,44]). Thus, one must sweep a lattice of $\gtrsim \xi^d$ sites about ξ^z times in order to obtain one independent data point, i.e., $\tau \gtrsim \xi^{d+z}$ when measured in elementary computer operations. [The proportionality constant in this relation can be reduced by a factor of 10 or more by clever computational methods such as multispin coding,[45–48] but the underlying ξ-dependence remains the same.] In the self-avoiding walk, by contrast, there are no finite-size effects whatsoever: the simulation operates directly on the infinite-volume system.[10] (This is a general property of most SAW Monte Carlo algorithms, not only ours; it is inherent in the nature of the SAW as compared to spin systems.) There is, to be sure, critical slowing down; in our algorithm we have, by (3.8) and (2.4),

$$\tau \sim \langle N \rangle^2 \sim \xi^{2/\nu} \tag{3.11}$$

In Table I we compare the dynamic critical exponents for the Ising model and the SAW for dimensions $d = 2, 3, 4$. It is seen that the SAW has a significant advantage which grows rapidly as one plunges deeper into the critical region ($\xi \to \infty$)[11]: e.g., for $d = 3$ and $\xi \approx 50$, the advantage is a factor

[10] Of course, since computer words are of finite length, one is really simulating the SAW in some large but finite box (with, say, periodic boundary conditions). But it is easy to arrange for this box to be so large that its boundaries are *never* touched in the course of thousands of hours of simulation; in that case one has in fact simulated the infinite-volume SAW, with no finite-volume effects whatsoever.

[11] Such a statement requires, of course, a standard of comparison, between two different models, of "how deep one is into the critical region." What is ultimately of interest is how close one has come to attaining the true asymptotic critical behavior; but this depends on correction-to-scaling exponents and amplitudes for the two models, which are difficult to estimate *a priori*. As a naive first approximation, we believe that the correlation length ξ, measured in units of the interaction range (here equal to one lattice spacing for both models), is a reasonable standard of comparison.

Table I. Comparison of Autocorrelation Times τ for Ising-Model and Self-Avoiding-Walk Dynamic Monte Carlo Algorithms

	Ising model (Metropolis algorithm)[a] $\tau \sim \xi^{d+z}$	Self-avoiding walk (our algorithm)[a] $\tau \sim \xi^{2/\nu}$
$d=2$	$\sim \xi^{\approx 4.1}$	$\sim \xi^{\approx 2.7}$
$d=3$	$\sim \xi^{\approx 5.1}$	$\sim \xi^{\approx 3.4}$
$d=4$	$\sim \xi^{6}$	$\sim \xi^{4}$

[a] ξ is the correlation length, measured in lattice spacings.

of ≈ 1000. This is, to be sure, a very crude estimate which neglects the possibly differing proportionality constants. But the qualitative conclusion is, we believe, inescapable: *the SAW is a uniquely advantageous laboratory for undertaking Monte Carlo studies of critical phenomena.*[12]

4. STATISTICAL METHODS

In this section we discuss statistical procedures for analyzing the Monte Carlo data produced by the algorithm of Section 3. We assume that the reader has a good background in mathematical statistics; the topics used here include time-series analysis, maximum-likelihood estimation, least-squares estimation, and confidence sets. An excellent exposition of all these topics except time-series analysis can be found in the brief book of Silvey[52]; other good sources are the books of Cramér,[53] Fisz,[54] Wilks,[55] Lindgren,[56] and Kendall and Stuart.[57] An excellent exposition of time-series analysis can be found in the book of Priestley[58]; other good sources are the books of Anderson[59] and Jenkins and Watts.[60]

4.1. Autocorrelations

Suppose that we run the Monte Carlo algorithm of Section 3 for a time n, starting from an initial SAW configuration ω_0; we then obtain a random sequence of SAWs $\omega_1, \ldots, \omega_n$. These SAWs are, of course, highly correlated; in this section we show how to obtain quantitative statistical

[12] At least until someone invents a radically better algorithm for the Ising model. For some recent attempts in this direction, see Refs. 49–51.

estimates of this autocorrelation. Knowledge of the autocorrelation is an indispensable prerequisite for correct statistical analysis of the Monte Carlo data.

As noted in Section 3, it follows from the general theory of Markov chains[40] that the probability distribution of ω_t converges as $t \to \infty$ to the stationary distribution (3.7), irrespective of the initial state ω_0. Thus, the Markov chain is an *asymptotically stationary* stochastic process; and if we choose t_0 sufficiently large, then the observations $\omega_{t_0+1}, \ldots, \omega_n$ can be considered for all practical purposes to come from a *stationary* stochastic process (i.e., this "thermalizes" the system). We shall thus consider the general question of how to estimate the autocorrelation function of a stationary stochastic process; at the end of this section we shall return to the important question of how to choose t_0.

Let $A = A(\omega)$ be some real-valued observable, and let $A_t = A(\omega_t)$; then the sequence $A_1, \ldots, A_n$ is a sample from a real-valued stationary stochastic process $\{A_t\}$. That process has a *mean value*

$$\mu_A \equiv \langle A_t \rangle \tag{4.1}$$

and an *autocorrelation function*[13]

$$C_{AA}(s) \equiv \langle A_t; A_{t+s} \rangle \equiv \langle A_t A_{t+s} \rangle - \mu_A^2 \tag{4.2}$$

both of which we assume to be finite. (They are independent of t, by stationarity.) We shall consider the following statistical estimation problems:

(a) The mean μ_A is unknown to us, and we wish to estimate it using the sample data $A_1, \ldots, A_n$. The autocorrelation function $C_{AA}(s)$ may be either known or unknown.

(b) The autocorrelation function $C_{AA}(s)$ is unknown to us, and we wish to estimate it using the sample data $A_1, \ldots, A_n$. The mean μ_A is assumed to be known.

(c) Same as (b), but the mean μ_A is assumed to be unknown.

The analysis of problem (a) is well known in the physics literature (see, e.g., Refs. 61 and 62); we include it here for the sake of completeness. The analysis of problems (b) and (c) appears to be less well known in the physics literature, although some partial discussions have appeared.[63,64] Here we will only summarize the results of the analysis; more detailed

[13] In the statistics literature, $C_{AA}(s)$ is called the *autocovariance function*, and $\rho(s) \equiv C_{AA}(s)/C_{AA}(0)$ is called the autocorrelation function. We shall adhere to the physicists' terminology.

treatments can be found in the textbooks on statistical time-series analysis.[58–60]

The "natural" solution to problem (a) is to estimate the population mean μ_A by the sample mean

$$\bar{A} \equiv \frac{1}{n} \sum_{i=1}^{n} A_i \tag{4.3}$$

This estimator is unbiased, i.e., $\langle \bar{A} \rangle = \mu_A$ for any stationary stochastic process. Its variance is

$$\begin{aligned} \mathrm{Var}(\bar{A}) &= \frac{1}{n^2} \sum_{i,j=1}^{n} C_{AA}(i-j) \\ &= \frac{1}{n} \sum_{s=-(n-1)}^{n-1} \left(1 - \frac{|s|}{n}\right) C_{AA}(s) \qquad (4.4a) \\ &\approx \frac{1}{n} \sum_{s=-\infty}^{\infty} C_{AA}(s) \qquad (4.4b) \end{aligned}$$

[The approximation in (4.4b) is valid for $n \gg \tau_{AA}$, where τ_{AA} is a typical "decay time" of the function $C_{AA}(s)$.] Clearly, then, the accuracy of $\bar{A}$ as an estimator of μ_A depends on the autocorrelation function $C_{AA}(s)$; the variance of $\bar{A}$ is a factor

$$S \equiv \frac{\sum_{s=-\infty}^{\infty} C_{AA}(s)}{C_{AA}(0)} \tag{4.5}$$

larger than it would be if the $\{A_i\}$ were independent. For example, if $C_{AA}(s) = \text{const} \times \exp(-|s|/\tau_{AA})$ (this is often a close approximation to the actual behavior), then

$$\begin{aligned} S &= \frac{2}{1 - e^{-1/\tau_{AA}}} - 1 \qquad (4.6a) \\ &\approx 2\tau_{AA} \qquad \text{if} \quad \tau_{AA} \gg 1 \qquad (4.6b) \end{aligned}$$

Thus, even if we are interested only in the *static* quantity μ_A, it is necessary to estimate the *dynamic* quantity $C_{AA}(s)$ in order to determine valid error bars for μ_A.

The "natural" estimator for $C_{AA}(s)$ is

$$\hat{C}_{AA}(s) \equiv \frac{1}{n - |s|} \sum_{t=1}^{n-|s|} (A_t - \mu_A)(A_{t+|s|} - \mu_A) \tag{4.7}$$

if the mean μ_A is known, and

$$\hat{\hat{C}}_{AA}(s) \equiv \frac{1}{n-|s|} \sum_{t=1}^{n-|s|} (A_t - \bar{A})(A_{t+|s|} - \bar{A}) \tag{4.8}$$

if the mean μ_A is unknown. We emphasize the conceptual distinction between the autocorrelation function $C_{AA}(s)$, which for each s is a *number*, and the estimator $\hat{C}_{AA}(s)$ or $\hat{\hat{C}}_{AA}(s)$, which for each s is a *random variable*; as will become clear, this conceptual distinction is also of *practical* importance. $\hat{C}_{AA}(s)$ is an unbiased estimator of $C_{AA}(s)$, and $\hat{\hat{C}}_{AA}(s)$ is almost unbiased (the bias is of order $1/n$) (Ref. 59, p. 463). Their variances are (Ref. 59, pp. 464–471)

$$\begin{aligned} \mathrm{Var}(\hat{C}_{AA}(s)) = \frac{1}{n-|s|} \sum_{t=-\infty}^{\infty} & [C_{AA}(t)^2 + C_{AA}(t-s)\, C_{AA}(t+s) \\ & + \kappa(|s|, -t, |s|-t)] + o\left(\frac{1}{n-|s|}\right) \end{aligned} \tag{4.9}$$

and

$$\mathrm{Var}(\hat{\hat{C}}_{AA}(s)) = \text{same thing} + o\left(\frac{1}{n-|s|}\right) \tag{4.10}$$

where κ is the connected 4-point autocorrelation function

$$\begin{aligned} \kappa(r, s, t) \equiv & \langle (A_i - \mu_A)(A_{i+r} - \mu_A)(A_{i+s} - \mu_A)(A_{i+t} - \mu_A) \rangle \\ & - C_{AA}(r)\, C_{AA}(t-s) - C_{AA}(s)\, C_{AA}(t-r) - C_{AA}(t)\, C_{AA}(s-r) \end{aligned} \tag{4.11}$$

For example, if $C_{AA}(s) = \text{const} \times \exp(-|s|/\tau_{AA})$ and $\kappa(r, s, t) = 0$ (this is often a decent approximation to the actual behavior), then we have (assuming $\tau_{AA} \gg 1$ for simplicity)

$$\begin{aligned} \mathrm{Var}(\hat{C}_{AA}(s)) \approx & \frac{\tau_{AA}}{n-|s|} C_{AA}(0)^2 \left[1 + e^{-2|s|/\tau_{AA}} \left(1 + \frac{2\,|s|}{\tau_{AA}}\right)\right] \\ & + o\left(\frac{1}{n-|s|}\right) \end{aligned} \tag{4.12}$$

and hence

$$\frac{\tau_{AA}}{n} C_{AA}(0)^2 \lesssim \mathrm{Var}(\hat{C}_{AA}(s)) \lesssim \frac{2\tau_{AA}}{n} C_{AA}(0)^2 \tag{4.13}$$

Of course, these formulas give the variance of $\hat{C}_{AA}(s)$ or $\hat{\hat{C}}_{AA}(s)$ in terms of the unknown quantity which we are trying to estimate, namely, $C_{AA}(s)$ (or τ_{AA}), as well as the further unknown quantity $\kappa(r, s, t)$. This is not an entirely satisfactory situation. Nevertheless we may proceed *heuristically* as follows: Usually we will have some vague prior knowledge of the autocorrelation time τ_{AA} (or at least an upper bound for it), based on theoretical considerations or on past experience. Let us choose, then, a run length n which is at least, say, $100\tau_{AA}$; by (4.13) this will ensure that the error bars on $\hat{C}_{AA}(s)$ or $\hat{\hat{C}}_{AA}(s)$ are no more than about 20% of $C_{AA}(0)$, and hence that one can obtain reasonable estimates of $C_{AA}(s)$ out to at least $s \approx \tau_{AA}$ before falling into the noise. This estimate of $C_{AA}(s)$ in the region of greatest importance can then be plugged into (4.9)/(4.10) to get a more refined estimate of the error bars. This procedure is, of course, somewhat circular, but it seems to work well in practice *provided that sufficient data is available* (i.e., $n \gtrsim 100\tau_{AA}$). As for κ, we usually have little *a priori* knowledge of it: it is identically zero if the process $\{A_t\}$ happens to be Gaussian, but we are rarely so lucky. We may have some knowledge about the kurtosis of the stationary distribution, i.e., $\kappa(0, 0, 0)/C_{AA}(0)^2$; if so, we could make the plausible guess that $\kappa(r, s, t)$ decays as a function of r, s, t more or less like $C_{AA}(s)$ does, i.e., with decay time $\approx \tau_{AA}$, and thereby get a crude estimate for the contribution of the $\kappa(r, s, t)$ term to (4.9)/(4.10). Or we might just ignore it and hope for the best.

It is clear from (4.13) that an enormous quantity of data is required if one wishes to obtain precise estimates of τ_{AA} or of $C_{AA}(s)$ for $s \gtrsim \tau_{AA}$. Luckily, our primary interest is in *static* quantities such as μ_A. We study the dynamic quantity $C_{AA}(s)$ only as a preliminary step toward a correct statistical analysis of the static quantities; and for this purpose, fairly crude estimates of $C_{AA}(s)$ and τ_{AA} are sufficient. The foregoing formulas should make clear, however, the severe difficulties involved in Monte Carlo studies of *dynamic* critical phenomena.

One further remark: We do not claim that the estimators $\bar{A}$, $\hat{C}_{AA}(s)$ and $\hat{\hat{C}}_{AA}(s)$ are the *optimal* solutions to the estimation problems (a)–(c); we claim merely that they are *simple and satisfactory*. In fact they are not optimal; but it would be too long a digression to discuss that question here.

We also note that subroutines for time-series analysis are available in both the IMSL[65] and NAG[66] software libraries.

We now return to our original problem, in which $\{A_t\}$ is not merely a stationary stochastic process but is in fact a function of a *reversible Markov process*. [A Markov process is called *reversible*[67] if it satisfies the detailed-balance condition, in our case (3.6).] In this case the autocorrelation function $C_{AA}(s)$ satisfies a spectral representation reminiscent of the Källén–Lehmann representation in quantum field theory, namely,

$$C_{AA}(s) = \int_{-1}^{1} \lambda^{|s|}\, d\rho_1(\lambda) \tag{4.14a}$$

$$= \int_{0}^{\infty} e^{-a|s|}\, d\rho_2(a) \qquad \text{for} \quad s \text{ even} \tag{4.14b}$$

where the spectral weights $d\rho_1(\lambda)$ and $d\rho_2(a)$ are *positive* measures. We prove this fact in Appendix B. Here we merely note the following consequence of (4.14):

$$\begin{aligned} \tau_{AA} &\equiv \lim_{\substack{s \to \infty \\ s \text{ even}}} \frac{-|s|}{\log[C_{AA}(s)/C_{AA}(0)]} \\ &= \sup_{s \text{ even}} \frac{-|s|}{\log[C_{AA}(s)/C_{AA}(0)]} \\ &= [\inf \operatorname{supp} d\rho_2(a)]^{-1} \end{aligned} \tag{4.15}$$

Finally, we return to the problem of how much data one must discard at the beginning of the run, i.e., how to choose t_0. This question can only be answered *after* one has at least a crude estimate (or upper bound) for the autocorrelation time τ. (Here τ is, by definition, the *longest* relaxation time in the system, i.e., the supremum of τ_{AA} over all observables A.) In that case one argues as follows: The *bias* associated with using the sample mean

$$\bar{A}^{(t_0)} \equiv \frac{1}{n - t_0} \sum_{i = t_0 + 1}^{n} A_i \tag{4.16}$$

as an estimator of the stationary-distribution mean value

$$\mu_A \equiv \sum_{\omega} \pi(\omega)\, A(\omega) \tag{4.17}$$

is presumably of order

$$\text{bias} \approx a_1 \frac{\tau}{n - t_0} e^{-t_0/\tau} \tag{4.18}$$

with a_1 being most plausibly of order $|A(\omega_0) - \mu_A|$. On the other hand, the *standard deviation* of this estimator is, from (4.4b)/(4.6b),

$$\text{standard deviation} \approx \frac{a_2}{(n - t_0)^{1/2}} \tag{4.19}$$

where $a_2 \approx [2\tau_{AA} C_{AA}(0)]^{1/2}$. It then suffices to take t_0 large enough so that the bias is negligible compared to the standard deviation, without making t_0 so large that the standard deviation increases significantly. This is easily accomplished, even if $a_1 \tau / a_2 n^{1/2}$ is enormous: just take t_0 to satisfy, say,

$$20\tau \lesssim t_0 \lesssim \frac{n}{4} \tag{4.20}$$

Any value of t_0 in this range is just about as good as any other. We emphasize that the most serious problem is *not* the need to discard some of the data at the beginning of the run, i.e., before the process has "reached equilibrium"; rather, it is the autocorrelation of the stochastic process *in equilibrium*, which dramatically raises the variance of all estimators. This latter problem (unlike the former one) cannot be evaded by tinkering with the initial configuration; it is an inherent feature of dynamic Monte Carlo methods. It can, however, be ameliorated by inventing new Monte Carlo algorithms (i.e., new stochastic processes) with smaller autocorrelation time τ.

4.2. Estimation of μ and γ

The algorithm of Section 3 in its stationary distribution produces a (correlated) sequence of SAWs $\omega_1, \ldots, \omega_n$ each of which is distributed according to the grand canonical ensemble at monomer activity β. Thus, the corresponding walk lengths $N_1, \ldots, N_n$ are distributed according to

$$\text{Prob}(\text{length} = N) = \text{const} \times \beta^N c_N \tag{4.21}$$

From these data we can make inferences about the unknown constants $\{c_N\}$. In particular, our goal is to estimate the parameters μ and γ defined in (2.1). Of course, in a strict sense this is impossible, since (2.1) is an asymptotic statement valid as $N \to \infty$, while a Monte Carlo experiment concerns only a finite range of N. Thus it is necessary to adopt additional *assumptions* about the behavior of the $\{c_N\}$ if we are to proceed further. We begin by adopting the simplest reasonable assumption, namely, that the formula

$$c_N = \mu^N N^{\gamma - 1} a_0 \tag{4.22}$$

is *exact* whenever N is greater than or equal to some cutoff value $N_{\min}$ (which we can choose later); here μ, γ, and a_0 are unknown constants. Of course, this assumption is manifestly false: (4.22) is only an *approximation* which gets better and better as N gets large. Thus our estimates for μ and γ

will likewise be *approximations* whose accuracy is only as good as that of the fundamental assumption (4.22). This error is a *systematic* error induced by corrections to scaling; it adds to the purely *statistical* error inherent in any Monte Carlo experiment (and which we shall calculate forthwith). We return to the discussion of corrections to scaling at the end of this subsection.

For now we assume that (4.22) is exact for $N \geqslant N_{\min}$. It follows that the probability distribution of the walk length N, *conditional on it being* $\geqslant N_{\min}$, is

$$\begin{aligned} &\text{Prob}(\text{length} = N \mid \text{length} \geqslant N_{\min}) \\ &\qquad = \text{const}(\beta\mu, \gamma, N_{\min}) \times (\beta\mu)^N N^{\gamma-1} \end{aligned} \tag{4.23}$$

Note that a_0 has dropped out of this formula.

If the $N_1, \ldots, N_n$ were independent, then the likelihood of a given sequence $\{N_1, \ldots, N_n\}$ would be simply the product of the individual likelihoods (4.23):

$$\text{likelihood} = \prod_{\substack{1 \leqslant i \leqslant n \\ N_i \geqslant N_{\min}}} \text{const}(\beta\mu, \gamma, N_{\min}) \times (\beta\mu)^{N_i} N_i^{\gamma-1} \tag{4.24}$$

(Here the product is taken only over those i for which $N_i \geqslant N_{\min}$; the walks of length $< N_{\min}$ play no role in this analysis.) We would then have a parametric-estimation problem for which we could use the method of maximum likelihood[52–57] to make statistical estimates of μ and γ. Of course, the $N_1, \ldots, N_n$ are far from independent, so it is not immediately clear that this approach is justified. We return to this issue at the end of this subsection; for now we simply pretend that the $N_1, \ldots, N_n$ *are* independent. (This could be arranged, for example, by choosing the $N_1, \ldots, N_n$ to be data points spaced in time by, say, 4τ; this would ensure that they are "for practical purposes" independent. Please note that we are not *advocating* this technique of discarding some of the data; we are only saying that it is *one* possibility which would make our theoretical pretenses almost true.)

With these caveats, we may proceed to the maximum-likelihood analysis. The maximum-likelihood estimates $\hat{\mu}$ and $\hat{\gamma}$ are, by definition, those values of μ and γ which, for the given data $\{N_1, \ldots, N_n\}$, maximize the likelihood (4.24). Since (4.24) is an "exponential family,"[52-57] the likelihood equations are particularly simple: they say that $\hat{\mu}$ and $\hat{\gamma}$ are determined by the conditions

$$\langle N \rangle_{\hat{\mu},\hat{\gamma}} = \langle N \rangle_{\text{obs}} \tag{4.25a}$$

$$\langle \log N \rangle_{\hat{\mu},\hat{\gamma}} = \langle \log N \rangle_{\text{obs}} \tag{4.25b}$$

where we have defined the theoretical mean values

$$\langle f(N)\rangle_{\mu,\gamma} \equiv \frac{\sum_{N=N_{\min}}^{\infty} f(N)(\beta\mu)^N N^{\gamma-1}}{\sum_{N=N_{\min}}^{\infty} (\beta\mu)^N N^{\gamma-1}} \tag{4.26}$$

and the observed mean values

$$\langle f(N)\rangle_{\text{obs}} \equiv \frac{\sum_{1\le i\le n, N_i \ge N_{\min}} f(N_i)}{\sum_{1\le i\le n, N_i \ge N_{\min}} 1} \tag{4.27}$$

The equations (4.25a, b) are easily solved numerically for $\hat{\mu}$ and $\hat{\gamma}$, e.g., by Newton's method. By the general theory of maximum-likelihood estimation,[52-57] the bias of the estimators $\hat{\mu}$ and $\hat{\gamma}$ is of order $1/n$. Moreover, the probability distribution of $(\hat{\mu}, \hat{\gamma})$ is asymptotically Gaussian as $n \to \infty$, with covariance matrix given explicitly as

$$\begin{aligned} \text{Cov}(\hat{\mu}, \hat{\gamma}) &\equiv \begin{pmatrix} \langle \hat{\mu}; \hat{\mu}\rangle_{\mu,\gamma} & \langle \hat{\mu}; \hat{\gamma}\rangle_{\mu,\gamma} \\ \langle \hat{\mu}; \hat{\gamma}\rangle_{\mu,\gamma} & \langle \hat{\gamma}; \hat{\gamma}\rangle_{\mu,\gamma} \end{pmatrix} \\ &= \frac{1}{n'} I(\mu,\gamma)^{-1} + O\left(\frac{1}{n'^2}\right) \end{aligned} \tag{4.28}$$

where

$$n' \equiv \sum_{\substack{1\le i\le n \\ N_i \ge N_{\min}}} 1 \tag{4.29}$$

is the censored sample size, $I(\mu, \gamma)$ is Fisher's information matrix

$$I(\mu,\gamma) = \begin{pmatrix} \frac{1}{\mu^2}\langle N; N\rangle_{\mu,\gamma} & \frac{1}{\mu}\langle N; \log N\rangle_{\mu,\gamma} \\ \frac{1}{\mu}\langle N; \log N\rangle_{\mu,\gamma} & \langle \log N; \log N\rangle_{\mu,\gamma} \end{pmatrix} \tag{4.30}$$

and $I(\mu, \gamma)^{-1}$ is its matrix inverse. We have here used the notation

$$\langle A; B\rangle \equiv \langle AB\rangle - \langle A\rangle\langle B\rangle \tag{4.31}$$

Note that in principle $\text{Cov}(\hat{\mu}, \hat{\gamma})$ depends on the unknown "true" values μ and γ; but since this dependence is rather weak, and since $\hat{\mu}, \hat{\gamma}$ will be fairly close estimates of μ, γ (provided $n' \gg 1$), it suffices for our purposes to replace μ, γ by the estimated values $\hat{\mu}, \hat{\gamma}$ when attempting to compute error bars for $\hat{\mu}, \hat{\gamma}$.[14] These error bars are computed (for $n' \gg 1$) by taking the

[14] More rigorously, one would limit oneself to some region R of (μ, γ) space in which the true values are *assumed* to lie, compute the *worst* possible error bars subject to that assumption, and thereby derive a rigorous confidence set for (μ, γ) subject to the assumption that $(\mu, \gamma) \in R$.

joint distribution of $(\hat{\mu}, \hat{\gamma})$ to be Gaussian with mean (μ, γ) and covariance matrix given by (4.28).[15] A very important feature of this method is that *the statistical error bars can be estimated prior to performing the Monte Carlo experiment.* Moreover, in the large-sample limit $n \to \infty$ *it can be proven that the maximum-likelihood estimator is the optimal estimator*, in the sense that any other estimator (within a certain very broad class) has larger or equal mean-square error at leading order in $1/n$.[52–57] Thus, the maximum-likelihood method provides an optimal data analysis: it extracts from the Monte Carlo data $\{N_1, ..., N_n\}$ their full content as regards the parameters μ and γ.

We now return to the problem of corrections to scaling. Clearly, (2.1) is only the leading term in the asymptotic expansion of c_N for large N; the renormalization group predicts[68] that the actual behavior is

$$c_N \sim \mu^N N^{\gamma-1} \left[a_0 + \frac{a_1}{N} + \frac{a_2}{N^2} + \cdots + \frac{b_0}{N^{\Delta_1}} + \frac{b_1}{N^{\Delta_1+1}} + \cdots + \frac{c_0}{N^{\Delta_2}} + \frac{c_1}{N^{\Delta_2+1}} + \cdots \right] \tag{4.32}$$

Here, in addition to analytic corrections to scaling of the form a_k/N^k, there are non-analytic corrections to scaling of the form b_k/N^{Δ_1+k} and c_k/N^{Δ_2+k} as well as more complicated terms not shown in (4.32). [Please note that the correction-to-scaling exponents $\Delta_1 < \Delta_2 < \cdots$ have no relation whatsoever to the gap exponent Δ_4 defined in (2.7). Our notation here is standard but unfortunate; we hope that it does not lead to any confusion.] The exponents $\Delta_1, \Delta_2, ...$ are believed to be universal among lattices of a given dimension d. The amplitudes $a_1, a_1, ..., b_0, b_1, ..., c_0, c_1, ...$ are lattice-dependent.

The maximum-likelihood analysis described above is based on the assumption that (4.22) is exact for $N \geqslant N_{\min}$; if (4.32) is correct, then this assumption is in error by an amount of order $1/N_{\min}$ (or $1/N_{\min}^{\Delta_1}$ if $\Delta_1 < 1$). Thus we expect that the estimates of μ and γ derived using (4.22) have likewise a systematic error of this order (as well as higher-order corrections). A useful procedure would then be to perform the analysis for a variety of values of $N_{\min}$; to plot $\hat{\mu}$ and $\hat{\gamma}$, together with their purely statistical error bars, as a function of $N_{\min}$ (or $1/N_{\min}$); and finally to attempt an extrapolation to $N_{\min} = \infty$. Of course, such an extrapolation is difficult: the data are "noisy" (in fact, the statistical error bars grow rapidly with $N_{\min}$), and one must contend with higher-order corrections. Thus,

[15] For $n' \gg 1$ we can neglect the bias (which is of order $1/n'$) since it is much smaller than the standard deviation [which is of order $1/(n')^{1/2}$].

some uncertainty about the "correct" $N_{\min} = \infty$ limit of the central values $\hat{\mu}$ and $\hat{\gamma}$ will inevitably remain. A crude (subjective) estimate of this residual uncertainty should be made; it should be reported as a possible *systematic error* induced by unaccounted-for corrections to scaling; it adds to the purely *statistical error* computed from (4.28) and embodied in the statistical error bars at each fixed $N_{\min}$.

As a further consistency check, one can perform the foregoing analysis with (4.22) replaced by

$$c_N = \mu^N N^{\gamma-1}\left(1+\frac{\tilde{a}_1}{N}\right) a_0 \tag{4.33}$$

(obviously $\tilde{a}_1 = a_1/a_0$). Here $\tilde{a}_1$ is considered, for the purpose of the maximum-likelihood analysis, to be a fixed constant. Since (4.33) is equivalent to (4.22) at leading order in $1/N$, any small value of $\tilde{a}_1$ is as reasonable *a priori* as any other; there is nothing sacred about zero. So one should perform the analysis for a variety of values of $\tilde{a}_1$, and check for consistency. The estimates extrapolated to $N_{\min} = \infty$ should ideally be independent of $\tilde{a}_1$; the discrepancy between the estimates obtained using different values of $\tilde{a}_1$ should be included in the reported systematic error. It is heuristically plausible that the "best" value of $\tilde{a}_1$ is that one which makes the estimates $\hat{\mu}$ and $\hat{\gamma}$ as "flat" as possible as a function of $N_{\min}$; though we are unable to offer any strong theoretical justification for this belief, we give some encouraging empirical results in Section 5.3. Finally, one may replace (4.33) by any functional form which is equivalent to it through order $1/N$; for example, one might use

$$c_N = \mu^N N^{\gamma-1} e^{\tilde{a}_1/N} a_0 \tag{4.34}$$

or

$$c_N = \mu^N (N+\tilde{\tilde{a}}_1)^{\gamma-1} a_0 \tag{4.35}$$

[here $\tilde{\tilde{a}}_1 = a_1/(\gamma-1)\,a_0$, and we must take $\tilde{\tilde{a}}_1 > -N_{\min}$; but (4.35) is not a good choice if γ is near 1].

The foregoing procedures for dealing with corrections to scaling are, of course, very crude and *ad hoc*. A more systematic approach would be to perform the maximum-likelihood analysis on (4.33) [or (4.34) or (4.35)] *with* $\tilde{a}_1$ *considered as an unknown parameter to be estimated along with* μ *and* γ. [(4.34) is particularly convenient for this purpose, since it is an exponential family.] This three-parameter maximum-likelihood estimation can be analyzed by the methods described previously. We have not carried out the details, but it seems clear that the variance of the estimator for $\tilde{a}_1$ will be very high, because $\langle N^{-1}; N^{-1}\rangle$ is so small. The variance of the

estimators for μ and γ will be slightly higher than in the two-parameter MLE in which $\tilde{a}_1$ is considered fixed; but the potential systematic error will be lower. In fact, the three-parameter MLE is probably roughly equivalent in practice to the more *ad hoc* approach described in the preceding paragraph; it is likely to be superior only if the available sample size is enormous. However, these questions deserve further investigation.

We emphasize that the foregoing analysis is based on the implicit assumption that the leading correction-to-scaling term in (4.32) is the a_1/N term, i.e., that $\Delta_1 \geqslant 1$. This assumption may or may not be correct; a more precise analysis would use a variable Δ_1. Unfortunately, however, such a subtle distinction among correction-to-scaling terms is currently invisible due to *statistical* error (see Section 5.3); any effect of $\Delta_1 \neq 1$ would simply be absorbed into a slightly changed correction-to-scaling amplitude. As more powerful computers become available, and the statistical errors are correspondingly reduced, the more precise correction-to-scaling analysis will become important. In any case, the errors resulting from the approximation $\Delta_1 \gtrsim 1$, as well as from the neglected higher-order terms in (4.32), are included in the quoted systematic error.

Finally, we return to the problem of autocorrelations in the data $\{N_1, ..., N_n\}$. The autocorrelation time τ can be estimated by the methods of Section 4.1. One then has several options: One may, as noted previously, choose to discard enough data so that the remaining data are essentially independent; the maximum-likelihood analysis then retains (more or less) its rigorous statistical justification. But this is a waste: the discarded data do contain useful information (albeit not as much as they would if they were independent). So an alternative procedure is to use *all* the data, computing the estimators $\hat{\mu}$ and $\hat{\gamma}$ *as if* these data were independent, and then to adjust the *error bars* to correct for the nonindependence. A conservative adjustment would be to multiply the variance by 4τ; by analogy with (4.4)–(4.6), one might guess that the "correct" factor is $\approx 2\tau$. A rigorous statistical justification for this procedure can be found in Refs. 69–71. A third alternative is to divide the Monte Carlo run into several blocks, each of which is large enough so that (a) distinct blocks are essentially independent, (b) the maximum-likelihood estimators $\hat{\mu}$ and $\hat{\gamma}$ for each block are essentially Gaussian-distributed, and (c) the bias of the block maximum-likelihood estimators is much smaller than the overall standard error. [Conditions (a) and (b) may typically be satisfied by taking the block size to be, say, $\gtrsim 50\tau$; but condition (c) is often more delicate, and may require block sizes of order 1000τ or even larger.] The block maximum-likelihood estimates are then independent Gaussian random variables of mean approximately (μ, γ); the standard t-test[52–57] can thus be used to derive confidence intervals for μ and γ.

We have used all three of these procedures on our Monte Carlo data, as a consistency check; in Section 5.3 we give the results.

4.3. Estimation of ν

Let $\omega_1, \ldots, \omega_n$ be the sequence of SAWs produced by the algorithm of Section 3 in its stationary distribution; let $N_1, \ldots, N_n$ be the corresponding walk lengths and $\mathbf{r}_1, \ldots, \mathbf{r}_n$ the end-to-end vectors. Then, for each i, the probability distribution of ω_i conditional on N_i (*but averaging over* $\{N_j\}_{j \neq i}$) gives equal probability to each N_i-step SAW.

Our goal is to estimate the critical exponent ν defined in (2.4)–(2.6). As in Section 4.2, we begin by assuming that (2.4)–(2.6) are *exact* for $N \geqslant N_{\min}$. Actually, the situation is here somewhat more subtle, because the validity of (2.5) requires not only $N \gg 1$ but also $|x| \gg 1$ (actually $|x| \sim N^{\nu}$). Of course, for large N the probability distribution of x is indeed concentrated around $|x| \sim N^{\nu}$, so the use of (2.5) also for small $|x|$ will presumably entail a systematic error which is proportional to some inverse power of N, like other corrections to scaling; such an error can be neglected at the present level of analysis. We are therefore justified in assuming that

$$E(\log(r_i^2 + a_1) \mid N_i) = 2\nu \log(N_i + a_2) + b \tag{4.36}$$

is *exact* whenever $N_i \geqslant N_{\min}$. Here a_1 and a_2 are arbitrary but fixed small numbers, which play a role analogous to $\tilde{a}_1$ in (4.33)[16]; b is an unknown constant which will be estimated along with ν; and $E(Y \mid X)$ denotes the conditional expectation of the random variable Y given the values of the random variable X [in the physicists' notation it might be written $\langle Y \rangle_X$, but this is cumbersome].

We are thus in the following situation: Let $X_i \equiv \log(N_i + a_2)$ and $Y_i \equiv \log(r_i^2 + a_1)$. Then we know that $\{(X_i, Y_i)\}$ form a stationary stochastic process satisfying

$$E(Y_i \mid X_i) = 2\nu X_i + b \tag{4.37}$$

Let the means of this stochastic process be

$$\mu_X \equiv \langle X_i \rangle \tag{4.38a}$$

$$\mu_Y \equiv \langle Y_i \rangle \tag{4.38b}$$

[16] Of course, we must demand that $a_1 > -1$ ($a_1 > 0$ if $N_{\min} = 0$) and $a_2 > -N_{\min}$, in order that the logarithms be always well defined.

and the covariances be

$$C_{XX}(i-j) \equiv \langle X_i X_j \rangle - \mu_X^2 \tag{4.39a}$$

$$C_{YY}(i-j) \equiv \langle Y_i Y_j \rangle - \mu_Y^2 \tag{4.39b}$$

$$C_{XY}(i-j) \equiv \langle X_i Y_j \rangle - \mu_X \mu_Y \tag{4.39c}$$

Then we can combine (4.37) with (4.38)/(4.39) to solve for ν and b in terms of the means and covariances:

$$\nu = \frac{C_{XY}(0)}{2C_{XX}(0)} \tag{4.40}$$

$$b = \frac{C_{XX}(0)\,\mu_Y - C_{XY}(0)\,\mu_X}{C_{XX}(0)} \tag{4.41}$$

If we now replace μ_X, μ_Y, $C_{XX}(0)$, and $C_{XY}(0)$ by their natural *estimators* $\bar{X}$, $\bar{Y}$, $\hat{\bar{C}}_{XX}(0)$, and $\hat{\bar{C}}_{XY}(0)$ [see (4.3) and (4.8)], we obtain the corresponding *estimators* $\hat{\nu}$ and $\hat{b}$. These estimators for ν and b are in general *biased*, because they involve a quotient of two random variables; but under very general assumptions on the stochastic process it can be shown that the bias is of order $1/n$, and thus negligible compared to the standard error. The variance of $\hat{\nu}$ and $\hat{b}$ (more generally, their covariance matrix) can be computed from the variances and covariances of $\bar{X}$, $\bar{Y}$, $\hat{\bar{C}}_{XX}(0)$, and $\hat{\bar{C}}_{XY}(0)$, which are in turn given by simple generalizations of (4.4) and (4.10).

We emphasize that our stochastic process does *not* satisfy the condition

$$E(Y_i \mid \{X_j\}) = 2\nu X_i + b \tag{4.42}$$

which is stronger than (4.37); this is because knowledge of the lengths of predecessor or successor SAWs, i.e., of $\{N_j\}_{j \neq i}$, does affect the probability distribution of ω_i. Thus, this process is *not* a "general linear model",[52–57] and the classical least-squares theory does *not* apply. Although the estimators $\hat{\nu}$ and $\hat{b}$ turn out to have the same form as in the classical least-squares theory, they have very different properties: the bias is nonzero, and the variance is much larger than in the classical case (one would guess by a factor $\approx 2\tau$).

Finally, the use made here of least-squares estimation *as a valid statistical method* should be distinguished from the common use of least-squares formulas *as a mere curve-fitting technique*. The latter is improper in the context of Monte Carlo data analysis: it confounds the roles played by

statistical errors (random fluctuations due to the inherent probabilistic nature of the Monte Carlo experiment) and *systematic errors* (errors due to misspecification of the model on which the data analysis is founded, e.g., due to corrections to scaling).

5. RESULTS

5.1. Preliminary Tests

Before starting the main Monte Carlo run, we performed several shorter runs at various values of β in order to test the correctness of our implementation of the algorithm and to test our conjecture (3.8) on the autocorrelation time.

Our algorithm is supposed to produce walks with the probability distribution (4.21). For the square lattice, the coefficients c_N are known by exact enumeration for $N \leqslant 24$.[72] We thus performed several runs at low values of β ($0.2 \leqslant \beta \leqslant 0.3$) in order to produce primarily short walks; we then compared the observed and expected distributions of N (conditional on $N \leqslant 15$) using the χ^2 test. We obtained good agreement with theory, both for ordinary random walks and for self-avoiding walks. We also observed the fraction of walks ending in each quadrant, and compared by the χ^2 method with the exact values of 1/4, again with good results.

We tested our statistical-analysis programs (time-series analysis to estimate autocorrelations, maximum-likelihood analysis to estimate μ and γ, least-squares analysis to estimate ν) by analyzing ordinary random walks produced by our algorithm. Here the exact answers $\mu = 4$, $\gamma = 1$, $\nu = 1/2$ are well known; the exact solution for the autocorrelations is given in Appendix A.

Our pseudo-random-number generator was the multiplicative congruential generator[73]

$$x_{n+1} = a x_n \bmod m \tag{5.1}$$

with modulus $m = 2^{48}$ and multiplier $a = 11^{13}$. This generator is due to Kalos, and has been used extensively in Monte Carlo studies at the Courant Institute. In particular, it has passed all the standard tests for randomness.[74]

Our main run was performed at $\beta = 0.376$, corresponding to

$$\langle N \rangle \approx \frac{\beta\mu\gamma}{1 - \beta\mu} \approx 166 \tag{5.2}$$

The random-number-generator seed was $x_0 = 1$. We took the initial configuration to be the empty walk, and then performed 45×10^9 Monte Carlo

steps; this took about 340 hours CPU time on the CDC Cyber 170-730. Data was taken once every 50×10^3 MC steps. In doing the statistical analysis we always skipped the data from the first 50×10^6 MC steps; since this is $\approx 300\tau$ (see Section 5.2), the system has clearly reached equilibrium.

5.2. Autocorrelations

We estimated the autocorrelation functions for the following observables:

$N(t)$: number of bonds in the walk
$\log[N(t)+1]$
$\mathbf{r}(t)$: end-point vector
$r(t)^2$: squared end-to-end distance
$\log[r(t)^2+1]$
$\varepsilon(t) \equiv \log[r(t)^2+1] - 2\nu_0 \log[N(t)+1]$:
error term in classical least-squares analysis;
here ν_0 is a guess for ν (we used $\nu_0 = 3/4$)

These autocorrelation functions are known to satisfy the spectral representation (4.14). Our main aim was to estimate the autocorrelation time τ defined in (4.15). The difficulty is that our estimates of the autocorrelation function $C(s)$ are afflicted by a statistical error [cf. (4.9)/(4.10)] which grows rapidly [relative to $C(s)$ itself] as $s \to \infty$, which is exactly the limit in which τ is defined. We therefore proceed heuristically as follows: we define a function

$$\tau(s) \equiv \frac{-s}{\log[C(s)/C(0)]} \tag{5.3}$$

and note that, by (4.14), $\tau(s)$ is increasing and its limit as $s \to \infty$ is the autocorrelation time τ. We then compute estimates $\hat{\tau}(s)$ to $\tau(s)$ by using our $\hat{C}(s)$ [or $\hat{\hat{C}}(s)$] defined in (4.7)/(4.8); we take as our estimate $\hat{\tau}$ of τ the *maximum* value achieved by $\hat{\tau}(s)$ before it begins to decrease due to noise. (This procedure probably tends to give a slight overestimate of τ.) The error bars on $\hat{\tau}$ are computed as usual by propagating the estimated error bars of $\hat{C}(s)$ [cf. (4.12)].

The results of this analysis for the main run are summarized in Table II. They are consistent with all observables other than ε having the same autocorrelation time, about 150 000 MC steps. Note also that the autocorrelation time for ε is almost exactly half that for the other quantities. We have no explanation for this curious result, but it seems vaguely reminiscent of a "two-particle threshold" in quantum field theory. Apparen-

Table II. Estimated Autocorrelation Times τ for Selected Observables, from Main Run at $\beta = 0.376$[a]

	τ
N	$(157 \pm 17) \times 10^3$
$\log(N+1)$	$(130 \pm 36) \times 10^3$
$\mathbf{r}$	$(154 \pm 6) \times 10^3$
r^2	$(161 \pm 16) \times 10^3$
$\log(r^2+1)$	$(138 \pm 13) \times 10^3$
ε	$(72 \pm 4) \times 10^3$

[a] Error bars are $\pm$ two standard deviations, and include statistical error only.

tly, ε acting on the "vacuum" yields a state which is orthogonal to the slowest mode, and the next-slowest mode (to which ε does couple) has exactly twice the "mass" of the slowest mode.

We also performed shorter runs at lower β in order to test the conjecture (3.8). The results of the time-series analysis for these runs are shown in Table III; for simplicity we give τ only for the observable N. We see that over a wide range of values of $\langle N \rangle$ and τ, the ratio $\tau/\langle N \rangle^2$ is constant within error bars. This supports our conjecture (3.8), with a proportionality constant $c \simeq 5$. Note, however, the extreme difficulty of getting high-precision estimates of dynamic quantities.

5.3. Estimates of μ and γ

We performed a maximum-likelihood estimation of μ and γ as described in Section 4.2, using the scaling form (4.35):

$$c_N = \mu^N (N + k_0)^{\gamma - 1} a_0 \qquad (N \geqslant N_{\min}) \tag{5.4}$$

Table III. Estimated Autocorrelation Time τ for the Observable N, for Selected Values of β[a]

β	τ	$\langle N \rangle$	$\tau/\langle N \rangle^2$
0.335	494 ± 41	10.22 ± 0.09	4.7 ± 0.5
0.355	1710 ± 170	19.7 ± 0.2	4.4 ± 0.5
0.369	$(16.2 \pm 4.1) \times 10^3$	49.0 ± 0.6	6.7 ± 1.9
0.376	$(157 \pm 17) \times 10^3$	165.7 ± 0.7	5.7 ± 0.7

[a] Error bars are $\pm$ two standard deviations, and include statistical error only.

Table IV. Maximum-Likelihood Estimates of μ, Assuming (5.4)[a]

k_0 \ $N_{\min}$	1	20	50	100
0	2.63853 ± 0.00017	2.63833 ± 0.00022	2.63826 ± 0.00030	2.63823 ± 0.00042
1	2.63831 ± 0.00017	2.63829 ± 0.00022	2.63824 ± 0.00030	2.63822 ± 0.00042
1.5	2.63822 ± 0.00018	2.63826 ± 0.00023	2.63823 ± 0.00029	2.63822 ± 0.00043
2	2.63815 ± 0.00018	2.63824 ± 0.00022	2.63822 ± 0.00030	2.63821 ± 0.00043
2.5	2.63808 ± 0.00018	**2.63822** ± 0.00023	**2.63821** ± 0.00030	**2.63821** ± 0.00043
3	2.63801 ± 0.00018	**2.63820** ± 0.00023	**2.63820** ± 0.00030	**2.63820** ± 0.00043
3.5	2.63795 ± 0.00018	**2.63818** ± 0.00023	**2.63819** ± 0.00030	**2.63819** ± 0.00043
4	2.63789 ± 0.00019	2.63815 ± 0.00023	2.63818 ± 0.00030	2.63819 ± 0.00043
5	2.63778 ± 0.00018	2.63811 ± 0.00023	2.63815 ± 0.00030	2.63818 ± 0.00043

[a] Error bars are ± two standard deviations, and include statistical error only; they are the values computed from (4.28), multiplied by $(2\tau)^{1/2}$.

[Here and in the following we write k_0 for $\tilde{a}_1$. We also remind the reader of the approximation $\Delta_1 \gtrsim 1$ implicit in (4.35).] We tried a wide range of values of k_0 and $N_{\min}$. The main results of this analysis are summarized in Tables IV and V.

The estimates obtained by taking $N_{\min} = 1$ are clearly biased by strong systematic error due to higher-order corrections to scaling not accounted for in (5.4); we thus feel it is safest to disregard these results. Restricting ourselves to the cases $N_{\min} = 20$, 50, 100 we see that for $k_0 = 2.5$, 3, 3.5

Table V. Maximum-Likelihood Estimates of γ, Assuming (5.4)[a]

k_0 \ $N_{\min}$	1	20	50	100
0	1.318 ± 0.009	1.333 ± 0.015	1.340 ± 0.024	1.344 ± 0.045
1	1.337 ± 0.009	1.340 ± 0.015	1.344 ± 0.025	1.346 ± 0.046
1.5	1.346 ± 0.010	1.343 ± 0.015	1.346 ± 0.025	1.348 ± 0.046
2	1.353 ± 0.010	1.346 ± 0.015	1.348 ± 0.025	1.349 ± 0.046
2.5	1.360 ± 0.010	**1.349** ± 0.015	**1.350** ± 0.025	**1.350** ± 0.046
3	1.367 ± 0.010	**1.352** ± 0.016	**1.352** ± 0.025	**1.352** ± 0.046
3.5	1.374 ± 0.010	**1.355** ± 0.016	**1.354** ± 0.025	**1.353** ± 0.046
4	1.380 ± 0.011	1.358 ± 0.016	1.356 ± 0.025	1.355 ± 0.047
5	1.393 ± 0.011	1.365 ± 0.016	1.360 ± 0.026	1.354 ± 0.047

[a] Error bars are ± two standard deviations, and include statistical error only; they are the values computed from (4.28), multiplied by $(2\tau)^{1/2}$.

(values printed in boldface) the dependence on $N_{\min}$ of the estimates of μ and γ is very weak; we consider these to be the "best" estimates. We take the arithmetic mean of the highest and lowest among the boldface values as our central estimate; we take the difference between these highest and lowest values as the systematic error (subjective 95% confidence limits)[17]; and we take the error bars at $N_{\min} = 50$ (which are 95% confidence limits in the classical statistical sense) as the purely statistical error. We thus obtain

$$\mu = 2.63820 \pm 0.00004 \pm 0.00030 \tag{5.5a}$$

$$\gamma = 1.352 \pm 0.006 \pm 0.025 \tag{5.5b}$$

as our final estimates; here the format is central estimate $\pm$ systematic error $\pm$ statistical error. Note that the data in Tables IV and V exhibit a very smooth and monotonic behavior as a function of k_0 and $N_{\min}$; our heuristic "flatness criterion" *appears* (based on internal evidence alone) to work well. For this reason, the estimated systematic error is considerably smaller than the statistical error.

We would have liked to do a further check on our method of dealing with corrections to scaling by using it on ordinary random walks; but since $\gamma = 1$ in this case, varying k_0 in (5.4) has no effect. This check could be done using (4.33) or (4.34) instead of (4.35)/(5.4), but we have not yet had time to do so.

The estimates in Tables IV and V are obtained using our preferred approach for dealing with the effects of autocorrelations, namely, using all the data as if it were independent and then multiplying the error bars by $(2\tau)^{1/2}$ to adjust for the nonindependence. In order to test this approach, we also performed a more traditional kind of analysis: we split the main Monte Carlo run into 50 blocks, each large enough ($\simeq 6000\tau$) to be considered effectively independent. We then computed maximum-likelihood estimates of μ and γ for each block (still using all the data as if it were independent), and defined final estimates as the arithmetic average of the 50 block estimates. The statistical error is a 95% confidence interval computed by the standard t test. Typical results are

[17] Note that this choice makes the confidence interval *twice* as wide as the minimal interal which contains all of the boldface values; we make this choice because we feel we do not completely understand the systematic error, and we therefore wish to be conservative. Our error bars are also *conservative* 95% confidence limits in the technical statistical sense: we are willing to wager any small sum of money (say, $1 U.S.), giving 19:1 odds, that our confidence interval covers the true value; but we are *not* willing to bet the contrary at only 19:1 odds.

$$k_0 = 1,\ N_{\min} = 50: \quad \mu = 2.63821 \pm 0.00023 \qquad \gamma = 1.346 \pm 0.014 \tag{5.6}$$

$$k_0 = 1,\ N_{\min} = 100: \quad \mu = 2.63816 \pm 0.00033 \qquad \gamma = 1.352 \pm 0.028 \tag{5.7}$$

The central estimates agree reasonably well with the corresponding ones obtained by a single maximum-likelihood analysis of the whole Monte Carlo run; the difference is probably attributable to the differing biases of the maximum-likelihood estimators for the whole run and for the blocks (since the bias is inversely proportional to the sample size). For this reason, we consider the central estimates (5.5) to be more accurate. However, (5.6)/(5.7) do indicate that the statistical error bars shown in Tables IV and V are probably a bit too conservative.

As a final check, we performed a maximum-likelihood analysis based on data taken only once every 500 000 MC steps; since this spacing is $\simeq 3\tau$, such data can be considered to be effectively independent, and so the standard maximum-likelihood theory applies without adjustment. Typical results are

$$k_0 = 1,\ N_{\min} = 50: \quad \mu = 2.63799 \pm 0.00040 \qquad \gamma = 1.356 \pm 0.033 \tag{5.8}$$

$$k_0 = 1,\ N_{\min} = 100: \quad \mu = 2.63803 \pm 0.00057 \qquad \gamma = 1.349 \pm 0.061 \tag{5.9}$$

The results are again consistent with those in Tables IV and V, but with larger error bars; this latter fact is hardly surprising, since some of the data have been thrown away.

5.4. Estimates of ν

We estimated ν as described in Section 4.3, using the scaling form (4.36):

$$\langle \log(r^2 + k_1) \rangle = 2\nu \log(N + k_2) + b \qquad (N \geqslant N_{\min}) \tag{5.10}$$

We tried a wide range of values of k_1, k_2, and $N_{\min}$. The main results of this analysis are summarized in Table VI.

In the present case the problem of systematic error due to corrections to scaling appears to be quite serious. As before, we disregard the estimates corresponding to $N_{\min} = 1$. Even so, we are unable to get a truly flat depen-

Table VI. Least-Squares Estimates of v, Assuming (5.10)[a]

k_1, k_2 \ $N_{\min}$	1	20	50	70	100
0, 0	0.7340 ± 0.0023	0.7434 ± 0.0030	0.7472 ± 0.0041	0.7495 ± 0.0050	0.7512 ± 0.0063
1, 1	0.7441 ± 0.0023	0.7483 ± 0.0030	0.7506 ± 0.0041	0.7525 ± 0.0050	0.7537 ± 0.0063
2, 2	0.7530 ± 0.0023	0.7533 ± 0.0030	0.7542 ± 0.0042	0.7555 ± 0.0050	0.7563 ± 0.0063
2.25, 2.25	0.7551 ± 0.0023	0.7546 ± 0.0030	0.7551 ± 0.0042	0.7563 ± 0.0050	0.7569 ± 0.0063
2.5, 2.5	0.7567 ± 0.0023	**0.7559** ± 0.0030	**0.7560** ± 0.0042	**0.7571** ± 0.0050	**0.7575** ± 0.0063
2.75, 2.75	0.7592 ± 0.0023	**0.7571** ± 0.0030	**0.7568** ± 0.0042	**0.7578** ± 0.0050	**0.7582** ± 0.0064
3, 3	0.7612 ± 0.0023	**0.7584** ± 0.0030	**0.7577** ± 0.0042	**0.7586** ± 0.0050	**0.7588** ± 0.0064
3.25, 3.25	0.7632 ± 0.0024	**0.7596** ± 0.0030	**0.7586** ± 0.0042	**0.7594** ± 0.0050	**0.7594** ± 0.0064
3.5, 3.5	0.7651 ± 0.0024	**0.7609** ± 0.0030	**0.7595** ± 0.0042	**0.7601** ± 0.0050	**0.7601** ± 0.0064
3.75, 3.75	0.7670 ± 0.0024	**0.7621** ± 0.0030	**0.7604** ± 0.0042	**0.7609** ± 0.0050	**0.7607** ± 0.0064
4, 4	0.7689 ± 0.0024	0.7634 ± 0.0030	0.7613 ± 0.0042	0.7616 ± 0.0050	0.7614 ± 0.0064
4.5, 4.5	0.7725 ± 0.0024	0.7659 ± 0.0031	0.7631 ± 0.0042	0.7632 ± 0.0050	0.7626 ± 0.0064
5, 5	0.7761 ± 0.0024	0.7684 ± 0.0031	0.7649 ± 0.0042	0.7647 ± 0.0050	0.7639 ± 0.0064

[a] Error bars are ± two standard deviations, and include statistical error only; they are the values computed from classical least-squares theory, multiplied by $(2\tau)^{1/2}$.

dence on $N_{\min}$ no matter how we choose k_1 and k_2. On the other hand, for a rather large range of values of k_1 and k_2 we can obtain a fairly weak dependence on $N_{\min}$ (for $N_{\min} \geqslant 20$); but this leads to a rather large range of estimates of ν, all of which appear equally plausible. More specifically, we find the following: for $k_1 \not\approx k_2$ the plot of $\hat{\nu}$ versus $N_{\min}$ is always very far from flat, so we restrict ourselves henceforth to the case $k_1 = k_2$. Then, for $2.5 \leqslant k_1 = k_2 \leqslant 3.75$ (values printed in boldface) the dependence of $\hat{\nu}$ on $N_{\min}$ is fairly weak; proceeding as in Section 5.3 we obtain the final estimate

$$\nu = 0.7590 \pm 0.0062 \pm 0.0042 \tag{5.11}$$

where the format is again central value $\pm$ systematic error $\pm$ statistical error. Note that here the systematic error is *larger* than the statistical error; clearly there are strong corrections to scaling which we do not understand and which are not captured in our assumption (5.10).

As a check on our method of dealing with corrections to scaling, we applied it to ordinary random walks. We produced ordinary random walks at $\beta = 0.244$, corresponding to $\langle N \rangle \simeq 40$, and performed 10^9 MC steps (about 5 hours CPU time). The resulting estimates of ν are summarized in Table VII. There is a quite wide range of values of k_1 and k_2 yielding reasonably flat plots; considering only these values of k_1, k_2 (and $N_{\min} \geqslant 10$), we are led to the estimates

$$\nu = 0.511 \pm 0.020 \pm 0.016 \tag{5.12}$$

where the statistical error is taken at $N_{\min} = 25$.

This is compatible with the exact answer $\nu = 1/2$. However, there seems to be in both (5.11) and (5.12) a consistent tendency to *overestimate* ν; this may indicate the inappropriateness of our form (5.10) for representing corrections to scaling. Perhaps a more suitable form would be

$$\langle \log(r^2 + k_1 N^{2\nu_0}) \rangle = 2\nu \log(N + k_2) + b \qquad (N \geqslant N_{\min}) \tag{5.13}$$

where ν_0 is a guess for ν; this form seems to reflect more accurately the expected scaling behavior (2.6). We hope in the near future to reanalyze our data using (5.13) in place of (5.10).

The error bars in Tables VI and VII are computed, for simplicity, by taking the standard least-squares error estimates[52–57] and multiplying them by $(2\tau)^{1/2}$, rather than by the more correct (but more complicated) formula mentioned in Section 4.3. As a check on this method, we also did a

Table VII. Least-Squares Estimates of ν for Ordinary Random Walk, Assuming (5.10)[a]

k_1, k_2 \ $N_{\min}$	1	10	25	50
0.25, 1.00	0.500 ± 0.006	0.499 ± 0.010	0.503 ± 0.016	0.509 ± 0.031
0.25, 1.50	0.518 ± 0.006	**0.506** ± 0.010	**0.507** ± 0.016	**0.512** ± 0.031
0.25, 2.00	0.534 ± 0.006	**0.513** ± 0.010	**0.512** ± 0.016	**0.515** ± 0.031
0.25, 2.50	0.550 ± 0.006	**0.520** ± 0.010	**0.516** ± 0.016	**0.518** ± 0.032
0.25, 3.00	0.564 ± 0.006	0.527 ± 0.010	0.520 ± 0.017	0.520 ± 0.032
0.50, 2.00	0.512 ± 0.006	0.501 ± 0.010	0.504 ± 0.016	0.509 ± 0.031
0.50, 2.50	0.527 ± 0.006	**0.508** ± 0.010	**0.508** ± 0.016	**0.512** ± 0.031
0.50, 3.00	0.541 ± 0.006	**0.515** ± 0.010	**0.512** ± 0.016	**0.515** ± 0.031
0.50, 3.50	0.554 ± 0.006	**0.521** ± 0.010	**0.516** ± 0.016	**0.517** ± 0.031
0.50, 4.00	0.567 ± 0.006	0.528 ± 0.010	0.520 ± 0.017	0.520 ± 0.031
0.75, 2.50	0.510 ± 0.006	0.498 ± 0.009	0.501 ± 0.016	0.507 ± 0.030
0.75, 3.00	0.524 ± 0.006	**0.505** ± 0.009	**0.505** ± 0.016	**0.510** ± 0.030
0.75, 3.50	0.536 ± 0.006	**0.512** ± 0.010	**0.509** ± 0.016	**0.512** ± 0.031
0.75, 4.00	0.549 ± 0.006	**0.518** ± 0.010	**0.513** ± 0.016	**0.515** ± 0.031
0.75, 4.50	0.560 ± 0.006	0.524 ± 0.010	0.517 ± 0.016	0.518 ± 0.031
1.00, 3.00	0.509 ± 0.006	0.496 ± 0.009	0.499 ± 0.016	0.505 ± 0.030
1.00, 3.50	0.521 ± 0.006	**0.503** ± 0.009	**0.503** ± 0.016	**0.508** ± 0.030
1.00, 4.00	0.533 ± 0.006	**0.509** ± 0.009	**0.507** ± 0.016	**0.510** ± 0.030
1.00, 4.50	0.545 ± 0.006	**0.515** ± 0.010	**0.511** ± 0.016	**0.513** ± 0.030
1.00, 5.00	0.556 ± 0.006	0.522 ± 0.010	0.515 ± 0.016	0.516 ± 0.031
1.25, 3.50	0.508 ± 0.006	0.495 ± 0.009	0.497 ± 0.015	0.504 ± 0.030
1.25, 4.00	0.520 ± 0.006	**0.501** ± 0.009	**0.501** ± 0.015	**0.506** ± 0.030
1.25, 4.50	0.531 ± 0.006	**0.508** ± 0.009	**0.505** ± 0.015	**0.509** ± 0.030
1.25, 5.00	0.542 ± 0.006	**0.514** ± 0.009	**0.509** ± 0.016	**0.512** ± 0.030

[a] Error bars are ± two standard deviations, and include statistical error only; they are the values computed from classical least-squares theory, multiplied by $(2\tau)^{1/2}$.

"block" analysis similar to the one done for μ and γ; we obtained for the SAW

$$k_1 = k_2 = 2.25,\ N_{\min} = 50: \quad \nu = 0.7550 \pm 0.0025 \tag{5.14}$$

$$k_1 = k_2 = 2.25,\ N_{\min} = 100: \quad \nu = 0.7565 \pm 0.0044 \tag{5.15}$$

The results agree closely with the corresponding entries in Table VI, and show that the error bars shown in that table are probably somewhat conservative.

Finally, we performed an analysis based on data taken only once every 500 000 MC steps, for which the standard (independent-errors) least-squares theory applies without adjustment. Typical results are

$$k_1 = k_2 = 2.25,\ N_{\min} = 50: \quad \nu = 0.7570 \pm 0.0055 \tag{5.16}$$

$$k_1 = k_2 = 2.25,\ N_{\min} = 100: \quad \nu = 0.7599 \pm 0.0084 \tag{5.17}$$

As expected, the results are consistent with the corresponding entries of Table VI but with larger error bars.

6. DISCUSSION

6.1. Comparison with Previous Work

In Table VIII we compare our Monte Carlo estimates of μ, γ, and ν with previously published estimates obtained by a variety of methods. Our estimates are consistent with most previous ones; in particular, they are consistent with Nienhuis'[79,80] exact (but nonrigorous) results $\nu = 3/4$ and $\gamma = 43/32$.[18]

Our error bars are larger than those in many of the cited works. However, at the risk of being considered spoilsports, we would like to explain why we believe that the error bars in several previous studies may have been underestimated.

Exact enumeration/extrapolation studies are notoriously difficult and subjective. They are extremely sensitive to the assumed form of corrections to scaling.[87] Apparently these difficulties were sufficiently serious that neither Guttmann[75] nor Adler[77] attempted to specify error bars for γ (naively one would guess from their data about ± 0.004). Moreover, Grassberger's[81] error bars for ν are clearly too small by a factor of 3, if for no other reason than that they exclude the probable exact value $\nu = 3/4$. Our inclination is therefore to be conservative in interpreting exact-enumeration/extrapolation data; we would prefer to multiply the error bars of Guttmann,[75] Adler,[77] and Majid *et al.*[82,83] by a factor of 1.5 or 2. Even so, we believe that exact enumeration/extrapolation is still the most accurate method available for estimating critical exponents. This will change, however, in the future, as more powerful computers become available: the error bars for Monte Carlo decrease much more rapidly with computer time than those for exact enumeration/extrapolation.[19] In any

[18] In Nienhuis' original article,[79] $\nu = 3/4$ was supported by a renormalization-group argument but $\gamma = 43/32$ had only the status of a promising numerological conjecture. Subsequently,[80] Nienhuis gave the result for γ a comparable renormalization-group foundation.

[19] The labor involved in exact enumeration grows exponentially with the order, while the extrapolation error is proportional to some inverse power of the order (the power depends on the details of the correction-to-scaling terms and the extrapolation method). Thus the error decreases only as an inverse power of the *logarithm* of the computer time. For Monte Carlo, by contrast, the behavior is a plain inverse power; see Section 6.2 and Table IX.

Table VIII. Comparison with Previous Estimates of μ (Square Lattice) and γ and ν (all Two-Dimensional Lattices)[a]

	Exact enumeration/extrapolation	Renormalization group	Previous Monte Carlo	Present paper	Exact values(?)
μ	2.6381 ± 0.0002[b]	2.63817 ± 0.00021[c]		$2.63820 \pm 0.00004 \pm 0.00030$	
	2.638155 ± 0.000025[l]				
γ	$\simeq 1.34$[b]	1.327,		1.352	1.34375[f]
	$\simeq 1.344$[d]	1.352[e]		$\pm 0.006 \pm 0.025$	
ν	0.747 ± 0.001[g]	0.7503 ± 0.0002[b]			
	0.7500 ± 0.0025[h]		0.753 ± 0.004[k]	0.7590 $\pm 0.0062 \pm 0.0042$	0.75[f]
		0.71 ± 0.10, 0.737, 0.7516[e]			
		0.756 ± 0.004[i]			
		0.74 ± 0.01[j]			

[a] Error bars are those claimed by the respective authors.
[b] Guttmann.[75]
[c] Derrida.[76]
[d] Adler.[77]
[e] Kolb *et al.* [78]
[f] Nienhuis.[79,80]
[g] Grassberger.[81]
[h] Majid *et al.*[82,83]
[i] Redner and Reynolds.[20]
[j] Kremer *et al.*[84]
[k] Havlin and Ben-Avraham.[85]
[l] Enting and Guttmann.[86]

case, the two methods have complementary advantages and drawbacks; it is important to pursue both. (After the completion of this work, Enting and Guttmann[86] took a significant step forward in exact enumeration: using a new algorithm, they extended the square-lattice self-avoiding-polygon series to order 46. *Assuming* the exact critical exponent $\alpha_{\text{sing}} = 1/2$ [which follows from $\nu = 3/4$ by the hyperscaling relation (2.11)], they estimated

$\mu = 2.638155 \pm 0.000025$—a value consistent with, but much more precise than, the present work.)

Derrida's[76] estimates $\mu = 2.63817 \pm 0.00021$ and $\nu = 0.7503 \pm 0.0002$ are based on a finite-size-scaling (phenomenological-renormalization) analysis of the transfer matrix on strips of width $n \leqslant 11$, extrapolated to $n = \infty$. However, the extrapolation of such short sequences is fraught with serious dangers, especially since our *theoretical* understanding of the convergence properties of the finite-size-scaling method is rather imperfect at present.[88–90] In particular, Derrida's periodic-b.c. data seem to converge anomalously fast (for one possible explanation of this phenomenon, involving a cancellation between leading and non-leading correction-to-scaling terms, see Refs. 89 and 90); moreover, the data are monotonic and go *beyond* the presumed exact value $\nu = 3/4$. For both these reasons, one may question whether the asymptotic regime (in which the leading correction-to-scaling term dominates all others) has yet been reached.[20] Derrida's periodic-b.c. estimates for ν and μ do in fact come remarkably close to the values obtained several years later by Nienhuis[79,80] and Enting and Guttmann,[86] respectively; but it is far from clear whether this reflects the systematic convergence rate of the finite-size-scaling method, or merely a fortuitous (and in fact misleading) cancellation among competing correction-to-scaling terms. In the latter case, Derrida's error bars would need to be increased rather drastically. Finally, we note that Derrida's data for free b.c. suggest a much higher value for ν, around 0.757. It is true that free b.c. are expected on theoretical grounds to be less reliable than periodic b.c., since they are subject to surface effects as well as finite-size effects; moreover, the free-b.c. estimates for ν in this model give internal evidence of their nonreliability by being non-monotonic. For these reasons the free-b.c. data should be given less credence, but in our opinion (as well as that of other authors[20,91]) they cannot simply be ignored: they indicate that we do not yet completely understand the convergence properties of the finite-size-scaling method—certainly not to the accuracy claimed. For further discussion, see Refs. 89–91. (We thank Michael Barber and Bernard Derrida for discussions and correspondence on these points; they are not, however, responsible for the views expressed here.)

Kolb *et al.*[78] performed a similar phenomenological–renormalization analysis of the "quantum Hamiltonian" version of the N-vector model (with periodic b.c. only), analytically continued to $N = 0$. By three distinct methods, they obtained $\nu = 0.71 \pm 0.10$, $\nu = 0.737$ and $\nu = 0.7516$, where no error bars are indicated on the latter two estimates. Kolb *et al.* also quoted estimates for the critical exponent η which, using the scaling law

[20] This possibility was stated clearly already by Derrida.[76]

$\gamma = (2-\eta)\nu$, imply the estimates $\gamma = 1.327$ and $\gamma = 1.352$. In our opinion the foregoing criticisms regarding extrapolation methods and boundary conditions apply here as well.

Redner and Reynolds[20] performed a large-cell position-space renormalization group analysis of the SAW on the square lattice, using a Monte Carlo method to generate the requisite data. Kremer *et al.*[84] carried out an analogous analysis of continuum polymer chains. Both of these works make a careful distinction between systematic and statistical errors; their error bars appear reasonable. (However, in retrospect one can see that they are at least a bit too small, since they exclude (or almost exclude) the probable exact value $\nu = 0.75$.)

Havlin and Ben-Avraham[85,92] analyze their Monte Carlo data using an interesting concept which they call "local fractal dimensionality." Their approach makes scaling assumptions on the *internal* distances of the SAW; as a result (and provided that their scaling assumptions are correct), they are able to utilize information which is simply thrown away in conventional analyses such as our own. One would therefore expect their method to be more accurate than otherwise-identical conventional Monte Carlo studies (how much more accurate is an open question). Unfortunately, however, Havlin and Ben-Avraham have not indicated clearly their statistical techniques or distinguished between systematic and statistical errors; further clarification in this regard would be desirable.

(After the completion of this work, we received an announcement of interesting Monte Carlo work by Rapaport.[93])

Finally, we note the existence of a number of position-space renormalization-group studies of self-avoiding walks[94–98,20,91,99]: their results are qualitatively reasonable but do not appear to have yet converged. For this reason we do not cite them in Table VIII.

6.2. Future Directions

We believe we have demonstrated in this paper that pure Monte Carlo[21] is a viable method for the high-precision numerical determination of critical exponents in the self-avoiding walk. We would like now to indicate some possible directions for future work:

(1) *Brute force.* Our use of computing power has been rather modest; one obvious way of improving on our results would be to apply

[21] By "pure" Monte Carlo we mean to distinguish our approach from "Monte Carlo renormalization group" methods.[100,20,84] In these latter methods, Monte Carlo is used as a *technical tool*, but the *conceptual foundation* is some form of the real-space renormalization group; in particular, the exponent estimates satisfy hyperscaling *automatically*. Of course, even our method does require *some* assumption on the form of scaling laws and corrections to scaling [cf. (4.32)].

Table IX. Scaling of Statistical and Systematic Error with Average Walk Length $\langle N \rangle$ and Computer Budget B

Quantity	Statistical error	Systematic error[a]	Optimal value of $\langle N \rangle$	Total error, using optimal $\langle N \rangle$
μ	$\sim B^{-1/2}$	$\sim \langle N \rangle^{-(1+\Delta)}$	∞ (or anything $\gtrsim B^{1/(2+2\Delta)}$)	$\sim B^{-1/2}$
γ	$\sim \langle N \rangle B^{-1/2}$	$\sim \langle N \rangle^{-\Delta}$	$\sim B^{1/(2+2\Delta)}$	$\sim B^{-\Delta/(2+2\Delta)}$
ν	$\sim \langle N \rangle B^{-1/2}$	$\sim \langle N \rangle^{-\Delta}$	$\sim B^{1/(2+2\Delta)}$	$\sim B^{-\Delta/(2+2\Delta)}$

[a] Here $\Delta \equiv \min(\Delta_1, 1)$ is the leading correction-to-scaling exponent.

more computer time and faster machines. The choice of β (and thus of $\langle N \rangle$) is a tradeoff between statistical error and systematic error. The statistical error can be computed from (4.28)–(4.30) multiplied by $(2\tau)^{1/2} \sim \langle N \rangle$. The order of magnitude of the systematic error can be determined by computing the effect on $\hat{\mu}$, $\hat{\gamma}$, and $\hat{\nu}$ of the correction-to-scaling terms in (4.32); it is controlled by the leading correction-to-scaling exponent $\Delta \equiv \min(\Delta_1, 1)$. Table IX shows how the errors scale with $\langle N \rangle$ and with the computer budget; it assumes that the data analysis is performed at $N_{\min} = \text{const} \times \langle N \rangle$.

(2) *One-parameter maximum-likelihood estimation.* If either μ or γ is known exactly, then the other one may be estimated by a one-parameter maximum-likelihood method; the estimates obtained in this way are about 3 times as accurate as the corresponding two-parameter MLEs described in Section 4.2. (Note that the error bars can here be computed *a priori* from (4.28)–(4.30) and their analogues.) For example, on the honeycomb lattice Nienhuis[79,80] has derived the exact value $\mu = (2+\sqrt{2})^{1/2}$; this can be employed to improve the estimate of γ. On the other hand, if one assumes the correctness of Nienhuis'[80] exact value $\gamma = 43/32$—which by universality ought to be valid for all two-dimensional lattices—then one can improve the estimate of μ on those two-dimensional lattices where it is not known exactly. We intend soon to reanalyze our square-lattice data in this way. For the triangular lattice, there is a long-standing conjecture[101] that $\mu_{\text{triangular}} + \mu_{\text{honeycomb}} = 6$; Guttmann, Osborn and Sokal[102] have recently undertaken a test of this conjecture using the Monte Carlo algorithm of the present paper.

(3) *Radius of gyration.* The radius of gyration of a SAW is expected to scale with the same exponent ν as the end-to-end distance; but its variance is likely to be smaller, since it is a more "global" measure of the behavior of the walk. The radius of gyration of an N-step walk can be com-

puted in a time of order N; but since this computation need be executed only when one takes data, i.e., once every $\sim\tau\sim\langle N\rangle^2$ Monte Carlo iterations, the extra computation involved is negligible. This is therefore a feature worth adding to future work.

(4) *Higher dimensions.* The methods of this paper extend immediately to SAWs on higher-dimensional lattices (see Appendix C for some programming considerations). The most important application, in our opinion, would be a direct test of hyperscaling in the $d=3$ SAW. (One point which requires further thought is the devising of a good scheme for the statistical estimation of the exponent Δ_4.) Another application would be the estimation of logarithmic correction-to-scaling exponents in $d=4$.

(5) *Monte Carlo renormalization group.* As we have emphasized, pure (non-RG) Monte Carlo methods have the advantage that they provide an independent determination of all critical exponents, and thus allow a *test* of scaling laws. In MCRG methods,[100,20,84] by contrast, all scaling laws (including hyperscaling) are built in from the beginning. On the other hand, if the scaling laws are in fact correct (as they probably are for $d<4$), MCRG provides a much more sophisticated—and, one would expect, correspondingly more accurate—method for the numerical determination of critical exponents. The raw data for a MCRG calculation can be supplied by any Monte Carlo algorithm, including our own. Ideally the computation should be arranged so as to permit *both* a "pure" and an MCRG analysis of the same Monte Carlo data. The concept of "local fractal dimensionality"[85,92] provides yet another interesting approach to analyzing the data.

APPENDIX A. EXACT SOLUTION OF THE DYNAMICS FOR THE ORDINARY RANDOM WALK

In this appendix we give the exact solution for the autocorrelation functions of our Monte Carlo algorithm for the case of ordinary random walk (i.e., with self-avoidance checking ignored).

Consider first those observables which are functions of $N(t)$, the number of bonds in the walk at time t. Now $N(t)$ executes a random walk with drift on the nonnegative integers: it is a Markov chain with one-step transition probabilities

$$p_{jk}=p(j\to k)=\begin{cases} w & \text{if } k=j+1 \\ 1-w & \text{if } k=j-1,\ j\geqslant 1 \\ 1-w & \text{if } k=j=0 \\ 0 & \text{otherwise} \end{cases} \tag{A1}$$

where

$$w = 2d\beta/(1+2d\beta) \tag{A2}$$

and $0 \leqslant w < 1/2$. This Markov chain has the unique stationary distribution

$$\pi_j = \left(\frac{w}{v}\right)^j \left(1 - \frac{w}{v}\right) \tag{A3}$$

where we have written $v \equiv 1 - w$. The autocorrelation function of $f(N)$ is then

$$\langle f(N(t)); f(N(t+s)) \rangle = \sum_{j,k=0}^{\infty} \pi_j f(j) [p_{jk}^{(s)} - \pi_k] f(k) \tag{A4}$$

where $p_{jk}^{(s)}$ are the s-step transition probabilities.

The transition probabilities $p_{jk}^{(s)}$ have been computed by Shenton[103] for a somewhat more general random walk, using a generating-function formalism. Specializing his equation (8) to our case,[22] we find after some algebra

$$p_{jk}^{(s)} = \pi_k + \frac{2v}{\pi} \left(\frac{w}{v}\right)^k \int_0^{\pi} \frac{g_j(\phi)\, g_k(\phi) [2(vw)^{1/2} \cos\phi]^s}{1 - 2(vw)^{1/2} \cos\phi} \, d\phi \tag{A5}$$

where we have written

$$g_j(\phi) \equiv \left(\frac{w}{v}\right)^{-j/2} \left[\sin j\phi - \left(\frac{w}{v}\right)^{1/2} \sin(j+1)\phi \right] \tag{A6}$$

Equation (A5) is the explicit form of the spectral representation whose existence is guaranteed by general theory (see Appendix B and Ref. 104, p. 59, exercise 3.21). The leading exponential behavior as $s \to \infty$ can be read off from (A5); the autocorrelation time is

$$\begin{aligned} \tau &= \{ -\log[2(vw)^{1/2}] \}^{-1} \\ &= -2/\log[1 - 1/(1 + 2\langle N \rangle)^2] \\ &= 8\langle N \rangle^2 + O(\langle N \rangle) \end{aligned} \tag{A7}$$

where we have used $\langle N \rangle = w/(1-2w)$. This confirms our qualitative argument leading to (3.8).

[22] A prefactor $1/\pi$ was inadvertently omitted in that equation (but not in subsequent ones).

We now consider observables which depend on the spatial conformation of the walk. Let $\mathbf{a}_j(t)$ be the jth bond of the walk, i.e., $\mathbf{a}_j(t)$ equals some unit coordinate vector if $j \leqslant N(t)$, and $\mathbf{a}_j(t) = 0$ if $j > N(t)$. Then

$$\mathbf{r}_k(t) = \sum_{j=1}^{k} \mathbf{a}_j(t) \tag{A8}$$

is the end point of the kth bond, and

$$\mathbf{r}_\infty(t) = \sum_{j=1}^{\infty} \mathbf{a}_j(t) = \sum_{j=1}^{N(t)} \mathbf{a}_j(t) \tag{A9}$$

is the end point of the chain. Clearly

$$\langle \mathbf{a}_j(t) \rangle = 0 \tag{A10}$$

by isotropy. Now we claim that

$$\langle \mathbf{a}_j(t_0) \cdot \mathbf{a}_k(t_0 + s) \rangle = \delta_{jk} \operatorname{Prob}(N(t) \geqslant j \quad \text{for all} \quad t_0 \leqslant t \leqslant t_0 + s) \tag{A11}$$

Here is why:

(a) If $j \neq k$, then $\mathbf{a}_j(t_0)$ and $\mathbf{a}_k(t_0 + s)$ are uncorrelated. [They are *not* independent, since $\mathbf{a}_j(t_0) \neq 0$ makes it more likely that $\mathbf{a}_k(t_0 + s) \neq 0$; but the *direction* of $\mathbf{a}_j(t_0)$ has no effect on that of $\mathbf{a}_k(t_0 + s)$.]

(b) If $j = k$ and $N(t) \geqslant j$ for all $t \in [t_0, t_0 + s]$, then $\mathbf{a}_j(t_0) = \mathbf{a}_k(t_0 + s)$ and hence $\mathbf{a}_j(t_0) \cdot \mathbf{a}_k(t_0 + s) = 1$.

(c) If $j = k$ and $N(t_0) < j$, then $\mathbf{a}_j(t_0) \cdot \mathbf{a}_k(t_0 + s) = 0$. Likewise if $N(t_0 + s) < j$.

(d) Finally, if $j = k$ and $N(t) < j$ for some $t \in (t_0, t_0 + s)$, then the jth bond was destroyed at time t. When it is later recreated, it will be done in a random direction, independent of the past history. Thus, the expectation value of $\mathbf{a}_j(t_0) \cdot \mathbf{a}_k(t_0 + s)$, *conditional* on $\{N(t) < j$ for some $t \in (t_0, t_0 + s)\}$, is zero.

Now the probability in (A11), conditional on $N(t_0)$, is just the probability of nonabsorption through time s in a random walk with drift on the nonnegative integers with absorbing barrier at the origin, given the initial position $N(t_0) - j + 1$. This probability has also been calculated by Shenton [Ref. 103, Eq. (18)]: it is zero if $N(t_0) < j$, and otherwise it is

$$\frac{[2(vw)^{1/2}]^{s+1}}{\pi} \left(\frac{v}{w}\right)^{m/2} \int_0^\pi \frac{\sin\phi \sin m\phi \, (\cos\phi)^s}{1 - 2(vw)^{1/2} \cos\phi} \, d\phi \tag{A12}$$

where $m \equiv N(t_0) - j + 1$. This quantity then has to be averaged with respect to the stationary distribution (A3) for $N(t_0)$. The result is

$$\langle \mathbf{a}_j(t_0) \cdot \mathbf{a}_k(t_0+s) \rangle = \delta_{jk} \frac{[2(vw)^{1/2}]^{s+1}}{\pi} \left(\frac{w}{v}\right)^{j-1/2} (v-w) \times \int_0^\pi \frac{\sin^2\phi \, (\cos\phi)^s}{[1-2(vw)^{1/2}\cos\phi]^2} \, d\phi \tag{A13}$$

The leading exponential behavior as $s \to \infty$ is given again by (A7).

APPENDIX B. PROOF OF THE SPECTRAL REPRESENTATION (4.14)

Let $\{X_0, X_1, \ldots\}$ be a Markov chain with finite or countably infinite state space S, and transition probabilities $p_{ij} \equiv p(i \to j)$. We assume that this Markov chain is *ergodic*, i.e., for each pair $i, j \in S$ there exists $n > 0$ such that $p_{ij}^{(n)} > 0$. [Here $p_{ij}^{(n)}$ are the matrix elements of P^n, the nth power of the matrix $P = \{p_{ij}\}$.] We further assume that the Markov chain is *reversible*, i.e., there exists a probability vector $\pi = \{\pi_i\}$ such that

$$\pi_i p_{ij} = \pi_j p_{ji} \tag{B1}$$

for all $i, j \in S$. Such a probability vector π is necessarily a stationary distribution for P [to see this, just sum (B1) over i]; by ergodicity, it is the *unique* stationary distribution for P, and is strictly positive. We assume henceforth that the Markov chain is started in this stationary distribution.

Now let A be a real-valued function defined on the state space S, and let $A_t \equiv A(X_t)$; by assumption, $\{A_t\}$ is a stationary stochastic process. Its mean value is

$$\mu_A \equiv \langle A_t \rangle = \sum_{i \in S} \pi_i A(i) \tag{B2}$$

Its second moment is

$$\langle A_t A_{t+s} \rangle = \sum_{i,j \in S} \pi_i A(i) \, p_{ij}^{(s)} A(j) \qquad [s \geqslant 0] \tag{B3}$$

and hence its autocorrelation function is

$$C_{AA}(s) \equiv \langle A_t; A_{t+s} \rangle \equiv \langle A_t A_{t+s} \rangle - \mu_A^2 = \sum_{i,j \in S} A(i)[\pi_i p_{ij}^{(s)} - \pi_i \pi_j] \, A(j) \qquad [s \geqslant 0] \tag{B4}$$

We assume that μ_A and $C_{AA}(s)$ are finite. Define now the matrix $Q = \{q_{ij}\}$ by

$$q_{ij} = \pi_i^{1/2} p_{ij} \pi_j^{-1/2} \tag{B5}$$

Q obviously has nonnegative entries; it is *irreducible* [by ergodicity] and *symmetric* [by (B1)]. Q has the strictly positive eigenvector $v = \{v_i\} \equiv \{\pi_i^{1/2}\}$ with eigenvalue 1. Note that v lies in the Hilbert space $l^2(S)$ and has norm 1, i.e., $\|v\| \equiv (\sum_i |v_i|^2)^{1/2} = 1$. Moreover, for any vector $x \in l^2(S)$ we have

$$\begin{aligned} \|Qx\|^2 = \|Q^T x\|^2 &= \sum_j \left| \sum_i x_i \pi_i^{1/2} p_{ij} \pi_i^{-1/2} \right|^2 \\ &\leqslant \sum_j \left(\sum_i |x_i|^2 p_{ij} \right) \left(\sum_l \pi_l \pi_j^{-1} p_{lj} \right) \\ &= \sum_{i,j} |x_i|^2 p_{ij} \\ &= \sum_i |x_i|^2 \\ &= \|x\|^2 \end{aligned} \tag{B6}$$

by the Schwarz inequality, hence Q is a *contraction* operator on $l^2(S)$. Let now E be the orthogonal projection in $l^2(S)$ onto the one-dimensional subspace spanned by v; its matrix elements are

$$e_{ij} = v_i v_j = \pi_i^{1/2} \pi_j^{1/2} \tag{B7}$$

Note that $EQ = QE = E$. Thus, the operator

$$R^{(s)} \equiv (1 - E)\, Q^s (1 - E) = Q^s - E \tag{B8}$$

has matrix elements

$$r_{ij}^{(s)} = q_{ij}^{(s)} - \pi_i^{1/2} \pi_j^{1/2} = \pi_i^{1/2} p_{ij}^{(s)} \pi_j^{-1/2} - \pi_i^{1/2} \pi_j^{1/2} \tag{B9}$$

Combining (B4) and (B9) and applying the spectral theorem to the self-adjoint contraction operator Q, we conclude that

$$\begin{aligned} C_{AA}(s) &= \sum_{i,j \in S} [\pi_i^{1/2} A(i)]\, r_{ij}^{(s)} [\pi_j^{1/2} A(j)] \\ &= \sum_{i,j \in S} \{\pi_i^{1/2} [A(i) - \mu_A]\}\, q_{ij}^{(s)} \{\pi_j^{1/2} [A(j) - \mu_A]\} \\ &= \int_{-1}^{1} \lambda^{|s|}\, d\rho_1(\lambda) \end{aligned} \tag{B10}$$

where $d\rho_1(\lambda)$ is a *positive* measure on $[-1, 1]$. [Note that the vector $w = \{w_i\} \equiv \{\pi_i^{1/2}[A(i) - \mu_A]\}$ must lie in l^2 if we are to have $C_{AA}(0) < \infty$.] This proves (4.14a); the change of variables $a = -\log|\lambda|$ yields (4.14b).[23]

It is interesting to note the following additional facts:

(a) The matrix P (and hence also Q) is either *aperiodic* or else of *period* 2.

(b) 1 is a *nondegenerate* eigenvalue of Q [considered always as an operator on $l^2(S)$].

(c) -1 is an eigenvalue of Q if and only if Q has period 2. In that case the eigenvalue is nondegenerate.

(d) If Q is aperiodic and the state space S is *finite*, then the measure $d\rho_1(\lambda)$ is actually supported in some interval $[-\lambda^*, \lambda^*]$ with $\lambda^* < 1$. If S is infinite, then $\lambda^* < 1$ and $\lambda^* = 1$ are both possible.

Part (a) is an immediate consequence of (B1). (b) and (c) are consequences of the Perron–Frobenius theory for countable stochastic (or more generally, nonnegative) matrices; see Ref. 105, Theorem 1; Ref. 106, Lemma 5.5 and Theorem 7.1; or Ref. 107, Theorem 2.2. (d) is an immediate consequence of (b) and (c).

The ideas in this Appendix go back at least to Nelson[108] and Kendall,[109,110] and probably much farther; see also Ref. 104, p. 58, exercise 3.19. Further references and generalizations can be found in Ref. 111.

APPENDIX C. PROGRAMMING CONSIDERATIONS

The only nonobvious aspect of our computer program is the data structure employed to store the walk configuration and to check the self-avoidance constraint.

If the walk configuration were stored only as a sequentially allocated linear list, then this list would have to be searched every time a $\Delta N = +1$ move is proposed, in order to check the self-avoidance constraint. But searching a linear list of N elements requires an average time of order N. For our purposes this is a disaster: it would nullify almost all of our algorithm's advantage as shown in Table I.

Our solution is the following: We maintain *two* data structures to store the current walk configuration: a linear list, and a "bit map." Both are updated at each iteration of the algorithm. The linear list contains the coordinates of points visited by the walk, in order; a pointer to the end of this list (i.e., to the walk endpoint) is maintained at all times. The bit map

[23] Here we define $0^0 = e^{-\infty \cdot 0} \equiv 1$. There could be a contribution in (B10)/(4.14) from $\lambda = 0$ ($a = \infty$), if 0 is an eigenvalue of Q.

is a large block of memory in which each site of a large spatial box (in our case 1024×1024) is assigned one bit: that bit is set to 1 if the site is visited by some step of the walk, and 0 otherwise. Thus, the checking of the self-avoidance constraint is rapid: one simply calculates the location of the relevant bit and examines it. In particular, the time is of order 1, i.e., independent of N. (We further speeded up this process by using in the bit map only 32 bits each 60-bit Cyber word: then calculation of the word and bit addresses requires only shifts and logical ANDs, i.e., no divisions.) This bit map does in principle impose periodic boundary conditions for the purpose of the self-avoidance checking; but in fact the box is so large that its boundary is *never* touched during the entire simulation, so the result is *rigorously equivalent* to simulating SAWs on an infinite lattice.

In dimension $d > 2$, the bit map would become prohibitively large (unless virtual memory is used). In that case one must use instead a "hash-coding" method.[112] The search time is still of order 1, provided that the hash table does not become nearly full.

The bit-map method has been used previously in several SAW MC studies[113–115]; likewise for the hash-coding method.[116,117]

All our programs are written in FORTRAN 5 ($\simeq$ FORTRAN 77) for the Cyber 170–730 with NOS 2.1 operating system, and are compiled at optimization level OPT = 2. The SAW-generation program required approximately 27 μs CPU time per MC step. (This has recently been improved to approximately 22 μs.) Our programs are available on request.

ACKNOWLEDGMENTS

We have benefited from conversations and correspondence with numerous people, among them Carlos Aragão de Carvalho, Michael Barber, Marvin Bishop, Bernard Derrida, Jürg Fröhlich, Paul Gans, Tony Guttmann, Mal Kalos, Sid Redner, Tom Spencer, Paula Whitlock, and George Williams. In addition, we would like to thank Erika Epstein, Ed Friedman, Max Goldstein, Barbara Kissner, Eleanor Kolchin, Terry Moore and the rest of the staff of the NYU Academic Computing Facility for helping us not to despair at persistent error messages. Finally, but not least, we would like to thank Judith Anderson and Connie Engle for their excellent typing (and retyping!) of the manuscript.

This research was supported in part by NSF grant No. DMS-8400955.

REFERENCES

1. W. Kuhn and H. Kuhn, *Helv. Chim. Acta* **26**:1394 (1943).
2. W. J. C. Orr, *Trans. Faraday Soc.* **43**:12 (1947).
3. E. W. Montroll, *J. Chem. Phys.* **18**:734 (1950).

4. C. Domb, *Adv. Chem. Phys.* **15**:229 (1969).
5. D. S. McKenzie, *Phys. Rep.* **27**:35 (1976).
6. S. G. Whittington, *Adv. Chem. Phys.* **51**:1 (1982).
7. P. G. deGennes, *Phys. Lett.* **38A**:339 (1972).
8. J. des Cloizeaux, *J. Phys. (Paris)* **36**:281 (1975).
9. M. Daoud *et al.*, *Macromolecules* **8**:804 (1975).
10. V. J. Emery, *Phys. Rev. B***11**:239 (1975).
11. C. Aragão de Carvalho, S. Caracciolo, and J. Fröhlich, *Nucl. Phys.* **B215**[FS7]:209 (1983).
12. A. K. Kron, *Vysokomol. Soyed.* **7**:1228 (1965) [*Polymer Sci. USSR* **7**:1361 (1965)].
13. A. K. Kron *et al.*, *Molek. Biol.* **1**:576 (1967) [*Molec. Biol.* **1**:487 (1967)].
14. F. T. Wall and F. Mandel, *J. Chem. Phys.* **63**:4592 (1975).
15. F. Mandel, *J. Chem. Phys.* **70**:3984 (1979).
16. B. Berg and D. Foerster, *Phys. Lett.* **106B**:323 (1981); see also J. Fröhlich, in *Recent Developments in Gauge Theories* (1979 Cargèse lectures), G. 't Hooft *et al.*, eds. (Plenum Press, New York, 1980), p. 65.
17. C. Aragão de Carvalho and S. Caracciolo, *J. Phys. (Paris)* **44**:323 (1983).
18. P. H. Verdier, *J. Chem. Phys.* **52**:5512 (1970).
19. A. D. Sokal, Comparative analysis of Monte Carlo methods for the self-avoiding walk (in preparation).
20. S. Redner and P. J. Reynolds, *J. Phys.* **A14**:2679 (1981); and S. Redner, private communication.
21. W. W. Wood and J. J. Erpenbeck, *Ann. Rev. Phys. Chem.* **27**:319 (1976).
22. H. E. Stanley, *Introduction to Phase Transitions and Critical Phenomena* (Oxford University Press, New York, 1971).
23. P. Erdös and S. J. Taylor, *Acta Math. Acad. Sci. Hung.* **11**:231 (1960).
24. G. F. Lawler, *Commun. Math. Phys.* **86**:539 (1982).
25. G. Felder and J. Fröhlich, *Commun. Math. Phys.* **97**: 111 (1985).
26. M. Aizenman, *Commun. Math. Phys.* **97**: 91 (1985).
27. D. C. Brydges, J. Fröhlich and A. D. Sokal (unpublished).
28. J. M. Hammersley, *Proc. Cambridge Philos. Soc.* **53**:642 (1957).
29. J. M. Hammersley, *Proc. Cambridge Philos. Soc.* **57**:516 (1961).
30. H. Kesten, *J. Math. Phys.* **4**:960 (1963).
31. H. Hesten, *J. Math. Phys.* **5**:1128 (1964).
32. A. J. Guttmann, *J. Phys.* **A16**:2233 (1983).
33. M. Aizenman, *Commun. Math. Phys.* **86**:1 (1982).
34. A. D. Sokal, unpublished.
35. J. Fröhlich, *Nucl. Phys.* **B200**[FS4]:281 (1982).
36. M. Aizenman and R. Graham, *Nucl. Phys.* **B225**[FS9]:261 (1983).
37. J. des Cloizeaux, private communication cited by E. Brézin, in *Order and Fluctuations in Equilibrium and Nonequilibrium Statistical Mechanics (17th Solvay Conference, 1978)*, G. Nicolis, G. Dewel, and J. W. Turner, eds. (Wiley-Interscience, New York, 1981).
38. B. Mandelbrot, *Fractals: Form, Chance and Dimension* (Freeman, San Francisco, 1977).
39. D. Brydges and T. Spencer, *Commun. Math. Phys.* **97**: 125 (1985).
40. K. L. Chung, *Markov Chains with Stationary Transition Probabilities*, 2nd ed. (Springer-Verlag, New York, 1967).
41. F. T. Wall and J. J. Erpenbeck, *J. Chem. Phys.* **30**:634 (1959).
42. M. N. Barber, in *Phase Transitions and Critical Phenomena*, Vol. 8, C. Domb and J. L. Lebowitz, eds. (Academic Press, London, 1983).
43. G. F. Mazenko and O. T. Valls, *Phys. Rev.* B **24**:1419 (1981).

44. R. Bausch *et al.*, *Phys. Rev. Lett.* **47**:1837 (1981).
45. R. Friedberg and J. E. Cameron, *J. Chem. Phys.* **52**:6049 (1970).
46. L. Jacobs and C. Rebbi, *J. Comput. Phys.* **41**:203 (1981).
47. C. Kalle and V. Winkelmann, *J. Stat. Phys.* **28**:639 (1982); S. Wansleben, J. G. Zabolitzky, and C. Kalle, *J. Stat. Phys.* **37**:271 (1984).
48. G. O. Williams and M. H. Kalos, *J. Stat. Phys.* **37**:283 (1984).
49. M. H. Kalos, in Proceedings of the Brookhaven Conference on Monte Carlo Methods and Future Computer Architectures, May 1983 (unpublished).
50. K. E. Schmidt, *Phys. Rev. Lett.* **51**:2175 (1983).
51. M. Creutz, *Phys. Rev. Lett.* **50**:1411 (1983).
52. S. D. Silvey, *Statistical Inference* (Chapman and Hall, London, 1975).
53. H. Cramér, *Mathematical Methods of Statistics* (Princeton University Press, Princeton, New Jersey, 1946).
54. M. Fisz, *Probability Theory and Mathematical Statistics* (Wiley, New York, 1963).
55. S. S. Wilks, *Mathematical Statistics* (Wiley, New York, 1962).
56. B. W. Lindgren, *Statistical Theory*, 3rd ed. (Macmillan, New York, 1976).
57. M. Kendall and A. Stuart, *The Advanced Theory of Statistics*, Vols. 1 and 2, 4th ed. (Macmillan, New York, 1977, 1979).
58. M. B. Priestley, *Spectral Analysis and Time Series*, 2 vols. (Academic Press, London, 1981).
59. T. W. Anderson, *The Statistical Analysis of Time Series* (Wiley, New York, 1971).
60. G. M. Jenkins and D. G. Watts, *Spectral Analysis and Its Applications* (Holden-Day, San Francisco, 1968).
61. H. Müller-Krumbhaar and K. Binder, *J. Stat. Phys.* **8**:1 (1973).
62. K. Binder, in *Monte Carlo Methods in Statistical Physics*, K. Binder, ed. (Springer-Verlag, Berlin, 1979), Chapter 1.
63. R. Zwanzig and N. K. Ailawadi, *Phys. Rev.* **182**:280 (1969).
64. G. J. Daniell, A. J. G. Hey, and J. E. Mandula, *Phys. Rev. D* **30**:2230 (1984).
65. *IMSL Library Reference Manual*, Edition 9 (IMSL, Houston, 1982), Chapter F.
66. *NAG* FORTRAN *Library Manual, Mark 10* (Numerical Algorithms Group, Oxford, 1983), Chapter G13.
67. J. G. Kemeny and L. J. Snell, *Finite Markov Chains* (Van Nostrand, Princeton, New Jersey, 1960).
68. F. J. Wegner, *Phys. Rev. B* **5**:4529 (1972).
69. I. V. Basawa and B. L. S. Prakasa Rao, *Statistical Inference for Stochastic Processes* (Academic Press, London, 1980), pp. 127–134, 162, 164.
70. Y. Ogata, *J. Appl. Prob.* **17**:59 (1980).
71. A. Azzalini, *Biometrika* **70**:381 (1983).
72. M. F. Sykes *et al.*, *J. Phys.* **A5**:653 (1972).
73. D. E. Knuth, *The Art of Computer Programming*, Vol. 2, 2nd ed. (Addison-Wesley, Reading, Massachusetts, 1981).
74. M. H. Kalos, private communication.
75. A. J. Guttmann, *J. Phys.* **A17**:455 (1984).
76. B. Derrida, *J. Phys.* **A14**:L5 (1981).
77. J. Adler, *J. Phys.* **A16**:L515 (1983).
78. M. Kolb, R. Jullien, and P. Pfeuty, J. Phys. **A15**:3799 (1982).
79. B. Nienhuis, *Phys. Rev. Lett.* **49**:1062 (1982).
80. B. Nienhuis, *J. Stat. Phys.* **34**:731 (1984).
81. P. Grassberger, *Z. Phys.* **B48**:255 (1982).
82. I. Majid, Z. V. Djordjevic, and H. E. Stanley, *Phys. Rev. Lett.* **51**:143 (1983).

83. Z. V. Djordjevic *et al.*, *J. Phys.* **A16**:L519 (1983).
84. K. Kremer, A. Baumgärtner, and K. Binder, *Z. Phys.* **B40**:331 (1981).
85. S. Havlin and D. Ben-Avraham, *Phys. Rev. A* **27**:2759 (1983).
86. I. G. Enting and A. J. Guttmann, *J. Phys. A* (1985), to appear.
87. B. G. Nickel, in *Phase Transitions* (1980 Cargèse lectures), M. Lévy, J.-C. LeGuillou, and J. Zinn-Justin, eds. (Plenum Press, New York, 1982).
88. B. Derrida and L. De Seze, *J. Phys. (Paris)* **43**:475 (1982), Appendix A.
89. V. Privman and M. E. Fisher, *J. Phys.* **A16**:L295 (1983).
90. V. Privman, *Physica* **123A**:428 (1984).
91. H. E. Stanley *et al.*, in *Real-Space Renormalization*, T. W. Burkhardt and J. M. J. van Leeuwen, eds. (Springer-Verlag, Berlin, 1982), Chapter 7.
92. S. Havlin and D. Ben-Avraham, *Phys. Rev. A* **26**:1728 (1982).
93. D. C. Rapaport, *J. Phys.* **A18**:L39 (1985); *J. Phys.* **A18**:113 (1985); *Phys. Rev. B* **30**:2906 (1984).
94. B. Shapiro, *J. Phys.* **C11**:2829 (1978).
95. M. Napiórkowski, E. H. Hauge, and P. C. Hemmer, *Phys. Lett.* **72A**:193 (1979).
96. A. Malakis, *Physica* **104A**:427 (1980).
97. S. L. A. de Quieroz and C. M. Chaves, *Z. Phys.* **B40**:99 (1980).
98. F. Family, *J. Phys.* **A13**:L325 (1980).
99. K. Kremer and M. N. Barber, *J. Phys.* **A17**:L215 (1984).
100. R. H. Swendsen, in *Real-Space Renormalization*, T. W. Burkhardt and J. M. J. van Leeuwen, eds. (Springer-Verlag, Berlin, 1982), Chapter 3.
101. A. J. Guttmann and M. F. Sykes, *Austral. J. Phys.* **26**:207 (1973).
102. A. J. Guttmann, T. Osborn, and A. D. Sokal, in preparation.
103. L. R. Shenton, *Proc. Cambridge Philos. Soc.* **51**:442 (1955).
104. D. Revuz, *Markov Chains* (North-Holland, Amsterdam, 1975).
105. Z. Šidák, *Czechoslovak Math. J.* **14**:438 (1964).
106. D. Vere-Jones, *Pacific J. Math.* **22**:361 (1967).
107. D. Vere-Jones, *Pacific J. Math.* **26**:601 (1968).
108. E. Nelson, *Duke Math. J.* **25**:671 (1958).
109. D. G. Kendall, in *Probability and Statistics* (*Cramér Memorial Volume*), U. Grenader, ed. (Amlqvist and Wiskell, Stockholm/New York, 1959).
110. D. G. Kendall, *Proc. London Math. Soc.* **9**:417 (1959).
111. E. Seneta, *Non-Negative Matrices and Markov Chains*, 2nd ed. (Springer-Verlag, New York, 1981), pp. 195, 218.
112. D. E. Knuth, *The Art of Computer Programming*, Vol. 3 (Addison-Wesley, Reading, Massachusetts, 1973).
113. F. L. McCrackin, *J. Chem. Phys.* **47**:1980 (1967).
114. J. Mazur and F. L. McCrackin, *J. Chem. Phys.* **49**:648 (1968).
115. P. H. Verdier, *J. Comput. Phys.* **4**:204 (1969).
116. P. C. Jurs and J. E. Reissner, *J. Chem. Phys.* **55**:4948 (1971).
117. R. Grishman, *J. Chem. Phys.* **58**:220 (1973).