Journal of Statistical Physics, Vol. 63, Nos. 5/6, 1991

Statistical and Systematic Errors in Monte Carlo Sampling

Alan M. Ferrenberg,[1] D. P. Landau,[1] and K. Binder[2]

Received May 1, 1990; final October 12, 1990

We have studied the statistical and systematic errors which arise in Monte Carlo simulations and how the magnitude of these errors depends on the size of the system being examined when a fixed amount of computer time is used. We find that, depending on the degree of self-averaging exhibited by the quantities measured, the statistical errors can increase, decrease, or stay the same as the system size is increased. The systematic underestimation of response functions due to the finite number of measurements made is also studied. We develop a scaling formalism to describe the size dependence of these errors, as well as their dependence on the "bin length" (size of the statistical sample), both at and away from a phase transition. The formalism is tested using simulations of the $d=3$ Ising model at the infinite-lattice transition temperature. We show that for a $96\times96\times96$ system noticeable systematic errors (systematic underestimation of response functions) are still present for total run lengths of 10^6 Monte Carlo steps/site (MCS) with measurements taken at regular intervals of 10 MCS.

KEY WORDS: Monte Carlo; statistical errors; systematic errors; Ising model.

1. INTRODUCTION

When planning a Monte Carlo (MC) investigation of a statistical mechanical model within a fixed computational budget one is faced with a difficult choice between performing long simulations of small systems or shorter simulations of larger systems.[1] In order to use the available computer time as efficiently as possible it is important to know the sources of errors, both

This paper is dedicated to Jerry Percus on the occasion of his 65th birthday.

[1] Center for Simulational Physics, University of Georgia, Athens, Georgia 30602.
[2] Institut für Physik, Universität Mainz, 6500 Mainz, Germany.

0022-4715/91/0600-0867$06 50/0

statistical and systematic, and how they depend on the size of the system and the number of updates performed.

Statistical errors occur because of the necessarily finite number of measurements which one can make during an MC simulation. The statistical uncertainty in the measured value of a quantity is proportional to $n^{-1/2}$, where n is the number of independent measurements made. {In general, successive measurements made during an MC simulation are not independent. This effect is well known and has been studied in detail (see ref. 2 for a recent review). Müller-Krumbhaar and Binder[3] have shown that the number of "effectively" independent measurements n obtained from N correlated MC measurements taken at intervals of Δt MCS is approximately $n = N/[2(\tau/\Delta t) + 1]$, where τ is the integrated correlation time measured in MCS.} For thermodynamic quantities the proportionality constant between the statistical uncertainty and $n^{-1/2}$, which is simply the standard deviation of the measured quantity, depends on the thermodynamic parameters (temperature, pressure, magnetic field, etc.) as well as on the size of the system. This size dependence is related to the degree of *self-averaging* the quantity exhibits.[4] For self-averaging quantities, the proportionality constant decreases as the system size increases.

An obvious way to reduce the statistical error is to increase the number of measurements made. Unfortunately, for large systems this is often impossible due to constraints on available computer resources. A second approach to reducing statistical errors is to make use of the size dependence of the variance. If the variance of a self-averaging quantity decreases sufficiently rapidly with increasing system size, it may be possible to reduce the statistical errors by studying large systems even though the number of measurements which can be made is small. Milchev *et al.* (MBH)[5] have addressed this question, but have not presented Monte Carlo data to test their predictions at a phase transition.

Systematic errors are more difficult to discuss because there are primarily two different kinds of systematic errors involved in MC sampling: those due to the finite size of systems studied and those due to the finite number of measurements made. (Additional errors which may arise due to correlations between the pseudorandom numbers generated for the simulation are not discussed here.) Finite-size effects arise when the characteristic length scale associated with some process in a physical system is comparable to or even larger than the linear dimension of the system being simulated. When this occurs, the behavior of the simulated system will systematically differ from the behavior of the infinite system. To obtain reliable information about the infinite-system behavior from simulations of finite systems, the linear dimension of the systems studied must be much larger than the correlation length ξ. This provides a model-dependent

lower bound on the system sizes studied in an MC simulation. At a phase transition, where ξ becomes infinite, it is impossible to overcome finite-size effects. However, using finite-size scaling techniques,[6,7] one can make use of the finite-size effects to predict the properties of the infinite system. The system sizes must then be chosen large enough that corrections to finite-size scaling are not large compared to the other statistical and systematic errors. Although an understanding of finite-size effects is important for planning an MC study, we will concentrate our effort on the other kind of systematic errors in the present paper.

Systematic errors due to the finite number of measurements made are particularly important in the MC sampling of response functions (susceptibility, specific heat, etc.) which are calculated from the measured variance of thermodynamic quantities. (These errors are also important in Monte Carlo renormalization group—MCRG[8]—calculations where the covariance of different operators is measured.) From elementary statistics (see, e.g., ref. 9) it is known that the measured variance of a distribution estimated using a finite number of independent samples is systematically lower than the true variance of the distribution. This systematic underestimation of the variance leads to systematically low estimates of response functions. Because the correlation time in an MC simulation can depend on the system size, the importance of this systematic error can be different for different system sizes. This introduces an additional size-dependent systematic error in the MC sampling of response functions. Error analysis for importance sampling MC methods has been considered by others,[10] but a coherent treatment at T_c is still lacking.

In this paper, we build on the work of MBH by considering the problem of the size dependence of statistical errors when the system being studied undergoes a phase transition. In addition, we consider the systematic errors in the MC sampling of response functions as mentioned above. In Section 2 we develop a formalism for estimating the statistical and systematic errors involved in the MC sampling of thermodynamic densities (energy and magnetization, for example) and their associated response functions (specific heat and susceptibility). In Section 3 we present a quantitative test of the formalism developed in Section 2 in a high-precision MC study of the $d=3$ simple-cubic Ising model. In Section 4 we summarize the results obtained in Sections 2 and 3.

2. THEORY

In the introduction to this paper, we presented a qualitative overview of the different sources of statistical and systematic errors which arise in MC simulations. In this section we provide a more quantitative discussion

of how the magnitude of these errors depends on the size of the system being simulated and the number of measurements made during the simulation. The first part of this section deals with the size dependence of statistical errors and its relation to the degree of self-averaging of the quantity measured. In the second part we discuss the systematic errors in the MC sampling of response functions.

2.1. Statistical Errors and Self-Averaging

We begin by defining the statistical error in a measured quantity. In our discussion we will consider the magnetization per particle m, although the arguments presented are valid for any thermodynamic quantity (energy per particle, magnetic susceptibility, etc.). This problem has been considered in detail by Müller-Krumbhaar and Binder,[3] who showed that the square of the error can be expressed, for large enough N, as

$$(\delta m)^2 = \frac{2(\tau/\Delta t)+1}{N}\frac{\chi_m}{L^d} \tag{1}$$

where L^d is the volume of the system, N is the number of measurements, made at regular time intervals Δt, and χ_m is the susceptibility defined by $\chi_m = L^d(\langle m^2\rangle - \langle m\rangle^2)$. The thermal average of a quantity is represented by $\langle\cdot\rangle$. The correlation time τ is defined as the sum from $t=1$ to $t=\infty$ of the time-displaced autocorrelation function for the magnetization

$$\varphi_{mm}(t) = \frac{\langle m(0)\, m(t)\rangle - \langle m\rangle^2}{\langle m^2\rangle - \langle m\rangle^2}$$

We will discuss the size dependence of statistical errors in terms of the *relative* error r given by

$$r = \frac{\delta m}{\langle m\rangle} = \left[\frac{2(\tau/\Delta t)+1}{N}\frac{\chi_m}{L^d\langle m\rangle^2}\right]^{1/2} \tag{2}$$

This differs from the approach taken by MBH, who worked with the absolute statistical errors. Away from a phase transition, where they concentrated their effort, the magnetization is fairly independent of system size, so that the absolute errors and the relative errors have the same size dependence. At a phase transition, however, some intensive quantities, such as an order parameter that vanishes, will have a strong size dependence, so that absolute and relative errors are no longer equivalent.

The term $\chi_m/L^d\langle m\rangle^2$ in (2) can be rewritten as $(\langle m^2\rangle - \langle m\rangle^2)/\langle m\rangle^2$, which is the expression for the relative fluctuations of m. We can relate the size dependence of the relative fluctuations to the degree of self-

averaging of the quantity in question. Assuming that the relative fluctuations vary as a power of the system size L^{-x}, we find:

1. If $x = d$, the quantity is strongly self-averaging.
2. If $0 < x < d$, the quantity is self-averaging.
3. If $x = 0$, the quantity exhibits a lack of self-averaging.

We can then express (2) in terms of the exponent x describing the degree of self-averaging,

$$r \sim \left[\frac{2(\tau/\Delta t) + 1}{N} L^{-x} \right]^{1/2} \tag{3}$$

An important concept needed for this discussion is that of fixed computational effort. We wish to determine whether the statistical errors from a simulation of a large system are larger than, smaller than, or the same as those obtained from a simulation of a smaller system using the same amount of computer time. The number of MCS which can be performed using a computer time B is $N = B \cdot u(L) \cdot L^{-d}$, where $u(L)$ is the number of updates performed per unit time. On a scalar computer, $u(L)$ is very nearly independent of L. However, on vector and parallel computers the speed at which updates are performed depends on the size of the system. For example, the speed of a vectorized program will increase with increasing system size until the size of the vectors used reaches the pipe length of the computer. With an efficient vectorized program, $u(L)$ can vary by an order of magnitude or more as L is increased, although eventually $u(L)$ becomes independent of L. In Fig. 1, we show the size dependence of $u(L)$ for the vectorized multispin coding program used in this study.[11] This program packs $q = 2^l$ ($l = 1,..., 6$) spin variables per word with the requirement that the system size L be an integer multiple (greater than two) of q. For fixed q, $u(L)$ increases monotonically with L, but jumps in $u(L)$ occur when

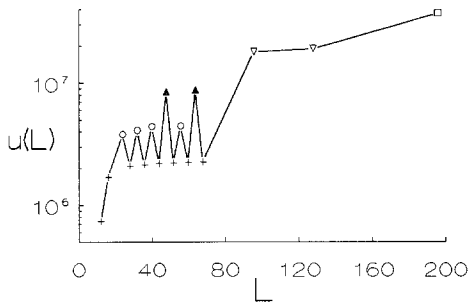

Fig. 1. Speed of the simulation algorithm used for this study on the CDC Cyber 205 (in updates per second) as a function of system size. The symbols correspond to the different numbers of spins packed per word (+ for 4 spins; ○ for 8; ▲ for 16; ▽ for 32; □ for 64).

adjacent system sizes have different q values. For a fixed computer time, the size dependence of the relative statistical error is given by

$$r \sim \left[\frac{2(\tau/\Delta t)+1}{u(L)} L^{d-x} \right]^{1/2} \tag{4}$$

Finally, we must include the size dependence of the correlation time τ. At a phase transition, the correlation time increases with increasing system size as L^z, where z is the dynamic critical exponent. Away from a transition, where the correlation time is independent of system size, we use $z=0$. Assuming that $\tau/\Delta t \gg 1$, we have

$$r \sim \left[\frac{L^{d+z-x}}{u(L)} \right]^{1/2} \tag{5}$$

With this formula for the relative statistical error we are now ready to discuss how this error varies with system size for different kinds of quantities with different degrees of self-averaging.

2.1.1. Strongly Self-Averaging Quantities. For strongly self-averaging quantities, $x=d$ in (5), so that the size dependence of the relative statistical errors is given by $[L^z/u(L)]^{1/2}$. An example of a strongly self-averaging quantity is the magnetization per particle of a spin system below its transition temperature. For this situation, the correlation time is also fairly independent of $L(z=0)$, so that the relative error varies like $u(L)^{-1/2}$. For scalar programs, where $u(L)$ is constant, this means that the relative error, for fixed computational effort, is independent of system size, so that simulations of large and small systems are equally efficient as far as statistical errors are concerned. On a vector computer, $u(L)$ increases with increasing L over some range of system sizes, so that for L within this range, the relative statistical error for fixed computational effort decreases as the system size is increased. Therefore, simulations of larger systems are more efficient than simulations of smaller systems within this range. However, for sufficiently large systems $u(L)$ is constant even for vector computers, so that the statistical errors are unaffected by increases in the system size. But if $z>0$, it becomes inefficient to increase L.

2.1.2. Self-Averaging Quantities. The energy per particle at a second-order phase transition is an example of a self-averaging quantity. The relative fluctuations in $\langle E \rangle$ decrease like $L^{-(d-\alpha/\nu)}$, with α and ν being the critical exponents for the specific heat and correlation length, respectively. Because $d-x$ in Eq. (5) is positive in this case, the size dependence of the relative error for self-averaging quantities is qualitatively different from that for strongly self-averaging quantities. Asymptotically, for $u(L)$

constant, the statistical error increases with increasing system size like $L^{(d-x)/2}$, even when the correlation time is constant, so that it is inefficient to increase the system size. At a phase transition the situation is even worse because $z \neq 0$, so the errors increase even more rapidly. For moderate system sizes, it will still be efficient to increase the system size provided that the program speed $u(L)$ increases more rapidly than L^{d+z-x}. While this may be possible in special cases, in general the speed of a vectorized program will not increase this rapidly. (For most vectorizable simulation algorithms, $z \simeq 2$, so that the errors increase very rapidly.)

2.1.3. Non-Self-Averaging Quantities. The analysis of statistical errors for quantities which exhibit a lack of self-averaging is qualitatively the same as that for self-averaging, but not strongly self-averaging quantities. Asymptotically it is inefficient to increase the system size, because the statistical errors increase like $L^{(d+z)/2}$ for fixed computational effort. For moderate system sizes it is even less likely that $u(L)$ will increase rapidly enough to make increasing the system size efficient.

The case of non-self-averaging quantities is important because response functions, such as the magnetic susceptibility χ, exhibit a lack of self-averaging. MBH have shown that the relative fluctuations in χ can be expressed in terms of the fourth-order cumulant of m, U_m, as

$$\frac{\langle \chi^2 \rangle - \langle \chi \rangle^2}{\langle \chi \rangle^2} = 2 - 3U_m$$

where

$$U_m = 1 - \frac{\langle m^4 \rangle - 4\langle m^3 \rangle \langle m \rangle + 6\langle m^2 \rangle \langle m \rangle^2 - 3\langle m \rangle^4}{3(\langle m^2 \rangle^2 - 2\langle m^2 \rangle \langle m \rangle^2 + \langle m \rangle^4)}$$

In situations where $\langle m \rangle = 0$, this reduces to the familiar result

$$U_m = 1 - \frac{\langle m^4 \rangle}{3\langle m^2 \rangle^2}$$

Away from a phase transition, U_m tends to zero as L is increased while at the transition U_m tends toward a universal constant U_m^*.

2.2. Systematic Errors in the MC Sampling of Response Functions

The goal of an MC simulation is to calculate the expectation value of thermodynamic quantities using the probability distribution $P(X) = (1/Z)\, e^{-\beta \mathcal{H}(X)}$, where Z is the partition function for the system, $\beta = 1/k_B T$, and X represents a system configuration. $\mathcal{H}(X)$ is the Hamiltonian of

the system. However, a finite-length MC simulation generates only a finite population of samples of the distribution $P(X)$. It is well known that certain quantities calculated from a finite population suffer from systematic errors due to the finite number of samples. For example, when estimating the variance s^2 of a probability distribution using n independent samples, the expectation value of the variance obtained $E(s^2)$ is systematically lower than the true variance of the distribution σ^2. It can easily be shown that the relationship between the two is given by

$$E(s^2) = \sigma^2 \left(1 - \frac{1}{n}\right) \tag{6}$$

Because the susceptibility χ is simply the measured variance of the magnetization, it is important to keep this systematic error in mind when the number of updates is small. Because an MC simulation does not provide statistically independent measurements, the relationship between the measured and correct values of χ is more complicated than that given in Eq. (6). Using the fact[3] that the number of independent measurements obtained from N correlated measurements is $n = N/[2(\tau/\Delta t) + 1]$, we have

$$\chi_N = \chi_\infty \left[1 - \frac{2(\tau/\Delta t) + 1}{N}\right] \tag{7}$$

where χ_N is the expectation value of the susceptibility when N measurements are made and χ_∞ is the true expectation value of the susceptibility. This effect becomes very important at a phase transition where one uses the values of χ from different system sizes to determine the critical exponent γ/ν. Because the correlation time depends on the system size, we see that the systematic error in χ will be different for different system sizes. The value of the exponent obtained using finite-size scaling methods will depend on the number of updates performed. In the next section we will present a study of this effect for the $d = 3$ Ising model.

3. MC SIMULATIONS

To test the formalism developed in Section 2, we chose to study the $d = 3$ nearest-neighbor Ising model. Simulations were performed at the infinite-lattice transition temperature $T_c^{-1} = 0.221654$ (see footnote 3) on

[3] The value of T_c^{-1} used for the simulations in this study was taken from ref. 12. More recent work using MCRG calculations,[13] reanalysis of series expansion results,[14] and high-resolution MC simulations[15] have provided results of comparable accuracy, although the results do not agree within the quoted error bars. The results presented in this paper are not of high enough precision to be affected by such small deviations from the true T_c.

system sizes ranging from $L=16$ to $L=96$ using an ultrafast multispin coding algorithm written for the CYBER 205 at the University of Georgia.[11] The peak speed of this program is 2×10^7 spin flip trials per second. Between 3 and 5×10^6 MCS were performed on all system sizes, except $L=16$, where 1.2×10^7 updates were performed. Data were taken at intervals of $\Delta t=10$ MCS. The long-time exponential relaxation time τ_0 for each lattice size has been measured and presented elsewhere.[16] Data for times $<10\tau_0$ were discarded for equilibrium.

In Fig. 2, we plot time-displaced autocorrelation functions for the energy E, energy squared E^2, magnetization m, and magnetization squared m^2 measured for the $L=16$ lattice. It is necessary to measure these autocorrelation functions to calculate the integrated correlation times needed for this analysis. From the plots in Fig. 2, we see that the magnetization squared, energy, and energy squared all have the same long-time relaxation time ($\tau_0\simeq63$ MCS). However, the integrated correlation times are different for the three quantities ($\tau=51$ MCS for the magnetization squared; $\tau=35$ MCS for E and E^2). The long-time relaxation time for the magnetization ($\tau_0\simeq421$ MCS) is considerably longer than that of the other three quantities. Our estimates for the correlation times were extracted from analyses which included data for much larger values of time displacement than are shown in Fig. 2, where for clarity we only show the relaxation functions for a restricted region of time displacement. Note that a simple estimate of relaxation times following Müller-Krumbhaar and Binder[3] suggests that $\tau_{m^2}=\frac{1}{2}\tau_m$. This approach clearly fails. It is important to note that for the same bin length N, the number of independent measurements $n=N/[2(\tau/\Delta t)+1]$ differs for different quantities. In Table I we show the integrated relaxation times for the specific heat (τ_C) and magnetic susceptibility (τ_χ). Note that τ_χ is approximately 11 times as large

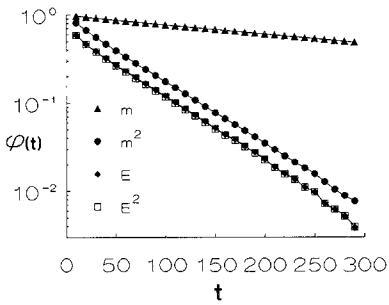

Fig. 2. Semilog plot of time-displaced autocorrelation functions for E, E^2, m, and m^2 for the $L=16$ lattice at T_c. The time t is measured in MCS.

Table I. Estimates for the Integrated Relaxation Times, in MCS, for the Specific Heat, τ_C, and the Magnetic Susceptibility, τ_χ

L	τ_C	τ_χ
16	35	395
32	130	1640
48	420	3745
64	545	6935
96	1305	15480

as τ_C, so that for a given bin size N, the statistical and systematic errors will be more pronounced for the susceptibility than for the specific heat. If the decay of the autocorrelation function is a pure exponential, as is nearly the case for the magnetization, the exponential (τ_0) and integrated (τ) correlation times will be equal, while for multiexponential decays, $\tau < \tau_0$.

As we saw in Section 2, the relative fluctuations in χ are given by $(2-3U_m)$. The relative statistical error in χ, from Eq. (2), is then

$$r = [(2-3U_m)/n]^{1/2} \tag{8}$$

This will be valid when the number of measurements made is large enough that systematic errors in the response function are small. To test this, we divided our data into M bins of different lengths N (N ranging from 5 to

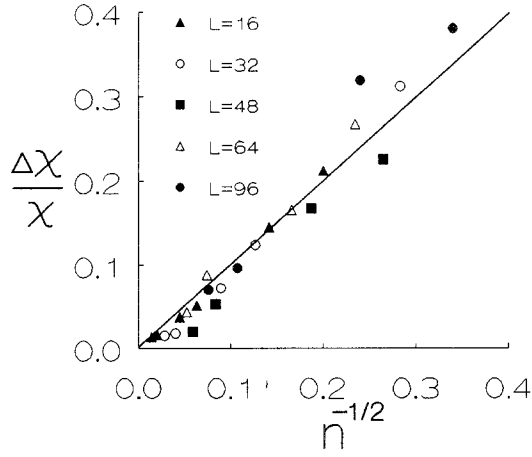

Fig. 3. Plot of scaled relative error in χ vs. $n^{-1/2}$, where $n = N/[2(\tau/\Delta t)+1]$ is the number of independent measurements. The straight line has slope = 1 and passes through the origin.

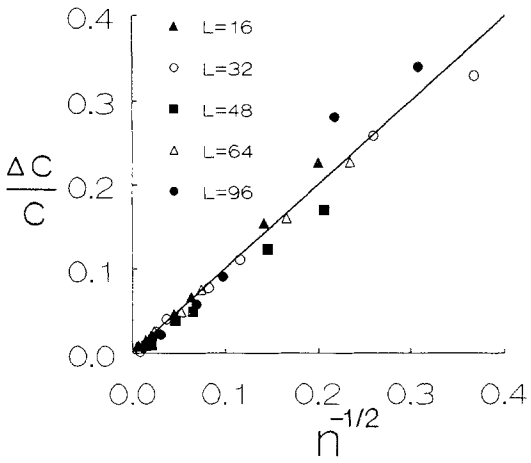

Fig. 4. Plot of scaled relative error in C vs. $n^{-1/2}$.

10^5). The response functions χ and C (specific heat) were measured over each bin and the statistical error in the response function was estimated from the variance over the M different bin values. In Fig. 3 we show a plot of the relative error in the response function χ scaled by $(2-3U_m)^{1/2}$ as a function of $n^{-1/2}$. This plot should give a straight line of slope unity passing through the origin. The corresponding plot for C is given in Fig. 4. In both plots there are fluctuations about the straight line, but no systematic deviations. From this, we conclude that the formalism presented in Section 2 is applicable and that response functions exhibit a lack of self-averaging.

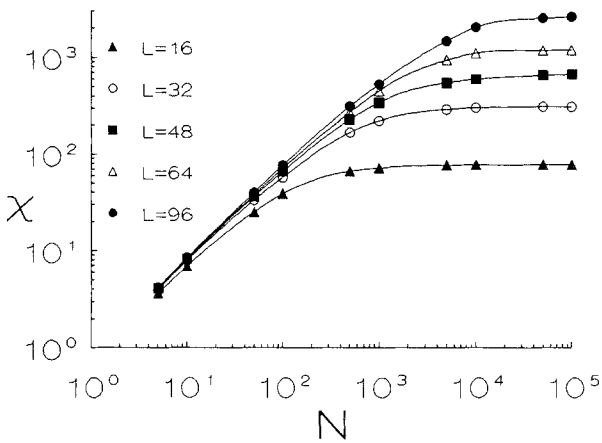

Fig. 5. Systematic deviations in the values of χ obtained using M bins of N data points each.

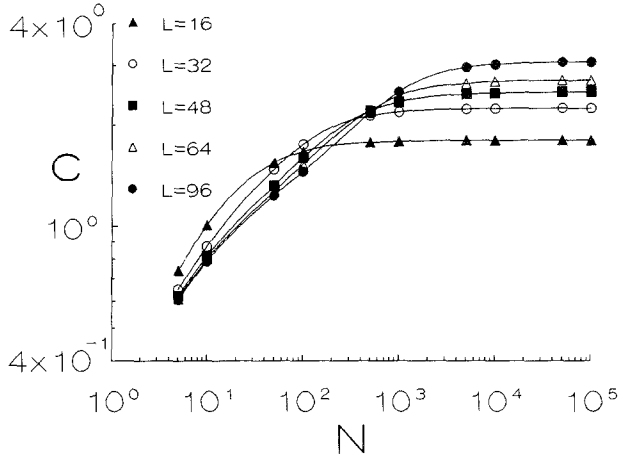

Fig. 6. Systematic deviations in the values of C obtained using M bins of N data points each.

While these statistical errors dominate for long MC runs, for short runs both systematic and statistical errors are important. In Figs. 5 and 6 we demonstrate this by plotting the measured response functions χ and C as a function of bin size N. The systematic effect of having a finite bin size can clearly be seen, especially for χ. (The correlation time corresponding to χ is approximately 11 times larger than that for C.) For the largest lattice, $L = 96$, the systematic error in χ is still noticable, even for the largest bin size $N = 10^5$ (which corresponds to 10^6 MCS!).

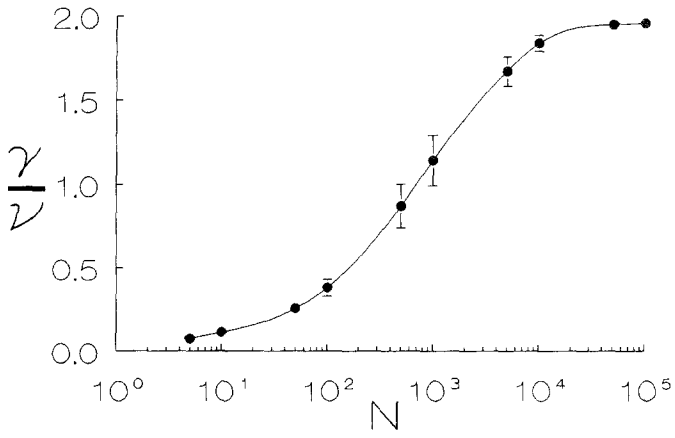

Fig. 7. Variation of γ/ν, determined by finite-size scaling, with the number of samples N.

As mentioned in Section 2.2, this systematic error in χ can affect the value of γ/ν measured using finite-size scaling. To demonstrate this effect, we calculated the exponent γ/ν by fitting the measured values of χ for each lattice size to $\chi = aL^{\gamma/\nu}$. The calculation of γ/ν was repeated for several different values of the bin size N. The results obtained are shown in Fig. 7. Even for a bin size $N = 10^4$, which corresponds to 10^5 MC updates/spin, the measured value differs from the accepted value $\gamma/\nu = 1.97$ by more than 6%. In some MC studies this effect could be even worse because one often uses shorter runs for the larger lattices! (For this calculation we used the same N for each lattice size.) This effect should be a particular problem for spin glasses and related random systems where z is very large.

The systematic error in χ depends only on the number of independent measurements n made. Therefore plots of scaled susceptibility $\chi L^{-\gamma/\nu}$ vs. scaled bin size n for different lattice sizes should collapse onto a single curve. This plot for the susceptibility is shown in Fig. 8. The solid curve comes from Eq. (7). Such a scaling plot for the specific heat is more difficult, as the slow divergence in C is superimposed on a relatively large background term. If this background is subtracted off, the specific heat also scales nicely using a value of $\nu = 0.629$ (or $\alpha/\nu = 0.18$), in good agreement with recent estimates for ν (see footnote 3 and refs. 17 and 18). This scaling curve is given in Fig. 9.

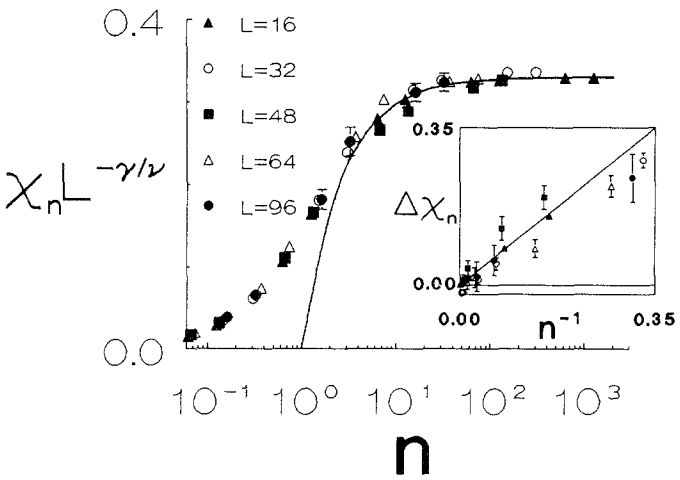

Fig. 8. Scaled susceptibility vs. scaled bin length n. The solid line is the scaling curve predicted by Eq. (7). In the inset the reduced systematic error $\Delta\chi_n = (\chi_\infty - \chi_n)/\chi_\infty$ is plotted vs. n^{-1} to highlight the large-bin-length behavior. The solid line, with slope = 1, is predicted by Eq. (7).

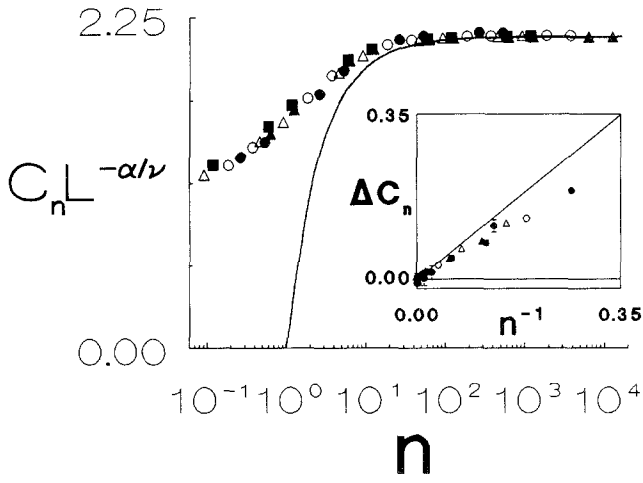

Fig. 9. Scaled specific heat (singular portion) vs. scaled bin length n. A background value of $C_0 = -1.6$ has first been subtracted off. The solid line is the scaling curve predicted by Eq. (7). The inset shows the reduced systematic error $\Delta C_n = (C_\infty - C_n)/C_\infty$ plotted as a function of n^{-1} as in Fig. 8.

4. CONCLUSIONS

From the results obtained, we can draw several general conclusions concerning ways to minimize systematic and statistical errors in functions measured during an MC simulation.

For response functions, it is clear that the only way to decrease systematic errors is to perform simulations that are very long compared to the correlation time. Finite-size effects (i.e., corrections to finite-size scaling) provide a lower bound on the system sizes to be studied, while the upper bound depends upon the amount of computer time available. Because response functions are not self-averaging, the statistical errors are not decreased by increasing the system size at fixed computational effort. Therefore, from the point of view of both statistical and systematic errors, it is never advantageous to perform short simulations on large lattices to measure response functions. Of course, systems which are too small also cannot be used to extract critical exponents due to unknown corrections to finite-size scaling.

If it is not necessary to measure response functions during an MC study, the conclusions are not as clear. When the quantities measured are not strongly self-averaging, for example at a phase transition, the above argument holds and a simulation of a larger system is less efficient than a simulation of a smaller system. Away from a phase transition, where the

thermodynamic densities are strongly self-averaging and the correlation time does not increase with increasing system size, we found that the statistical error remains constant, or even decreases, as the system size is increased, depending on the simulation algorithm and computer type used.

Although we have presented results for only a single kinetic model, the formalism is quite general and can be applied to different models and simulation techniques. We have seen that many factors must be taken into account in order to choose optimally the size of systems simulated in an MC study. Some of these are directly related to the model being studied, such as the finite-size effects which determine the lower bound for the system sizes, and the degree of self-averaging exhibited by the quantities to be measured. The variation of the speed of the simulation with system size also plays a role. For systems undergoing a phase transition, the size dependence of the correlation time, determined by the model *and* the simulation algorithm used, must also be taken into account.

We suggest that for high-precision estimates of critical exponents from finite-size scaling, the scaling analysis of the data as a function of the scaled bin length (as shown in Figs. 8 and 9) should be carried out.

ACKNOWLEDGMENTS

We would like to thank R. H. Swendsen for helpful and interesting discussions and S. Wansleben and L. Zha for their help during the preliminary phases of this project. This work was supported in part by NSF grant DMR-8715740.

REFERENCES

1. D. Stauffer, *J. Appl. Phys.* **53**:7980 (1982).
2. D. P. Landau, S. Tang, and S. Wansleben, *J. Phys. (Paris)* **49**:C8-1525 (1988), and references therein.
3. H. Müller-Krumbhaar and K. Binder, *J. Stat. Phys.* **8**:1 (1973).
4. A. Sadiq and K. Binder, *J. Stat. Phys.* **35**:617 (1984).
5. A. Milchev, K. Binder, and D. W. Heermann, *Z. Phys. B* **63**:521 (1986).
6. M. E. Fisher, in *Critical Phenomena*, M. S. Green, ed. (Academic Press, New York, 1971).
7. M. N. Barber, in *Phase Transitions and Critical Phenomena*, Vol. 8, C. Comb and J. Lebowitz, eds. (Academic Press, New York, 1983); V. Privman (ed.), *Finite-Size Scaling and Numerical Simulation* (World Scientific, Singapore, 1990).
8. R. H. Swendsen, *Phys. Rev. Lett.* **42**:859 (1979).
9. H. W. Alexander, *Elements of Mathematical Statistics* (Wiley, New York, 1961).
10. W. W. Wood, *Physics of Simple Liquids* (Wiley, New York, 1968), Chapter 5; D. P. Landau, *Phys. Rev. B* **13**:2997 (1976); N. Madras and A. Sokal, *J. Stat. Phys.* **50**:109 (1988); H. Flybjerg and H. G. Petersen, *J. Chem. Phys.* **91**:461 (1989); D. J. Wallace, Private communication.

11. S. Wansleben, *Comp. Phys. Commun.* **43**:315 (1987).
12. G. S. Pawley, R. H. Swendsen, D. J. Wallace, and K. G. Wilson, *Phys. Rev. B* **29**:4030 (1984).
13. H. W. J. Blöte, J. de Bruin, A. Compagner, J. H. Croockewit, Y. T. J. C. Fonk, J. R. Heringa, A. Hoogland, and A. L. van Willigen, *Europhys. Lett.* **10**:105 (1989); H. W. J. Blöte, A. Compagner, J. H. Croockewit, Y. T. J. C. Fonk, J. R. Heringa, T. S. Smit, and A. L. van Willigen, *Physica A* **161**:1 (1989).
14. A. J. Liu and M. E. Fisher, *Physica A* **156**:35 (1989).
15. A. M. Ferrenberg and D. P. Landau, *Phys. Rev. B* (in press).
16. S. Wansleben and D. P. Landau, *J. Appl. Phys.* **61**:3968 (1987); *Phys. Rev. B* **43**:6006 (1991).
17. G. Bhanot, D. Duke, and R. Salvador, *Phys. Rev. B* **33**:7841 (1986).
18. N. A. Alves, B. A. Berg, and R. Villanova, *Phys. Rev. B* **41**:383 (1990).