INTERNATIONAL ATOMIC ENERGY AGENCY
UNITED NATIONS EDUCATIONAL, SCIENTIFIC AND CULTURAL ORGANIZATION
INTERNATIONAL CENTRE FOR THEORETICAL PHYSICS
I.C.T.P., P.O. BOX 586, 34100 TRIESTE, ITALY, CABLE: **CENTRATOM TRIESTE**

H4.SMR/854-18

College on Computational Physics

15 May - 9 June 1995

Various Publications

R. Swendsen

Carnagie Mellon University
Pittsburgh, USA

Physica A 194 (1993) 53–62
North-Holland

PHYSICA A

Modern methods of analyzing Monte Carlo computer simulations

Robert H. Swendsen

Department of Physics, Carnegie-Mellon University, Pittsburgh, PA 15213, USA and Weizmann Institute of Science, Electronics Department, Rehovot 76100, Israel

In the past few years, new approaches to Monte Carlo simulations have produced substantial improvements in the efficiency of both simulation techniques and data analysis. This paper will focus on the recent renewal of interest in histogram methods and the new developments in this field. This approach to data analysis has proven very effective in improving the efficiency and ultimate accuracy of Monte Carlo calculations. The methods will be described along with several new applications.

1. Introduction

In recent years it has been recognized that the analysis of data from Monte Carlo computer simulations can be greatly enhanced by the use of histograms. Although the title of this paper refers to "modern methods", the use of histograms in this context actually has a history going back to at least 1959 [1–13]. Since this history has been discussed in detail elsewhere [14], I will concentrate on the methods themselves and their recent applications.

The basic single-histogram method will be reviewed briefly in section 2. The multiple-histogram method will be presented in section 3 along with a discussion of the essential assumptions and various alternatives that have been suggested. Section 4 will present applications of the histogram method, with emphasis on the use of Binder's cumulants [15,16]. Section 5 will discuss the new analysis methods for identifying first-order transitions [17,18], along with the closely related multicanonical simulation method [19].

2. The single-histogram method

Consider a Monte Carlo simulation of a thermodynamic model. During the simulation each configuration generated provides values for the energy, E,

magnetization, M, and other quantities of interest. We will restrict the discussion of the equations to histograms of the energy, since extensions to multiple observables are straightforward. Although the most usual procedure has been to simply record the averages of the quantities and their squares, we could also record a histogram, $N(E)$, of the number of times each value (or set of values) was generated in a run of length n. (For a model with a continuous energy spectrum it is convenient to discretize the value by setting up a large number of bins.) The expectation values of the histogram, $N(E)$, averaged over an infinite number of runs, is just

$$\overline{N(E)} = nW(E)\exp(-\beta E + f)\,, \tag{1}$$

where $f = \beta F(\beta)$ gives the free energy at the inverse temperature $\beta = 1/k_B T$ where the simulation was performed. Solving this equation for $W(E)$ shows that the histogram $N(E)$ provides a direct approximation for the density of states up to the constant factor $\exp(-f)$. Given the density of states, it is then straightforward to compute the expectation value of any function of E at any other temperature by the usual equations,

$$\langle E \rangle = \sum_E EW(E)\exp(-\beta E)/Z(\beta)\,, \tag{2}$$

where the partition function is given by

$$Z(\beta) = \sum_E W(E)\exp(-\beta E)\,. \tag{3}$$

Although these equations are quite old [1], the recent observation that they provide the most accurate method for determining the location and height of peaks in thermodynamic properties [13] added new interest to the application of this approach to Monte Carlo studies of phase transitions. Another particularly striking recent advance is that these equations allow the investigation of complex values of the temperature, to which no real experiment has access [6–12]. The locations of the zeros of the partition function can be determined accurately, providing yet another tool for investigating phase transitions. These two features of the histogram approach have proven extremely valuable in improving the ease and accuracy of the determination of critical properties with remarkably little additional computational effort.

Naturally, the accuracy of the estimate for $W(E)$ is limited by the accuracy of the histogram. This will be taken up again in the next section, but the basic observation is that the square of the error in an entry in the histogram is proportional to the expectation value of the entry. This may be approximated

by $N(E)$ so that the relative accuracy of $W(E)$ is greatest where $N(E)$ is largest. When eq. (2) is used at a temperature far from where the simulation is performed, the important values of $W(E)$ come from the wings of the histogram, which have few (or no) entries, and hence large errors. Clearly, the largest acceptable shift in temperature, ΔT_{max}, corresponds to moving the peak of the histogram (that is the peak of $W(E)\exp(-\beta E)$) by an amount proportional to the width of the distribution. The magnitude of ΔT_{max} has been shown to scale as $L^{-1/\nu}$ at a second-order phase transition (where l is the linear dimension of the system), which just coincides with the finite-size scaling region [13]. Away from a transition the width scales as $L^{-d/2}$.

Before introducing the multiple-histogram method, I would like to call attention to the very interesting work by Rickman and Phillpot [20], who showed that away from the critical region certain systems show much better results from two or three terms in a Monte Carlo series expansion about the simulation temperature than they do from the full histogram analysis. This offers a simple and effective way of covering large temperature ranges in which the variation of thermodynamic variables is expected to be fairly smooth.

3. The multiple-histogram method

Although the single-histogram method is sufficient for many useful applications, it was recognized long ago [4] that to cover a wide range of temperature (or other parameters) it would be necessary to somehow combine the results of several simulations. The first attempt to do this involved a simple matching of the temperature-shifted histograms in the overlap region where neighboring histograms both provide data [4]. Note from eq. (1) that the essential information needed is the difference in the free energies at the temperatures of the two simulations. This method of patching together neighboring histograms does indeed work, but a much more accurate method has been proposed by Ferrenberg and Swendsen, which optimizes the combination of information from an arbitrary number of Monte Carlo simulations [21].

From eq. (1) it is clear that we can form an estimate for the density of states from a linear combination of the estimates from R histograms $N_i(E)$ measured in simulations at inverse temperatures β_i,

$$W(E) = \sum_{i=1}^{R} p_i(E)\, N_i(E)\, n_i^{-1} \exp(\beta_i E - f_i)\,, \tag{4}$$

where the weights, $p_i(E)$, are normalized to one,

$$\sum_{i=1}^{R} p_i(E) = 1 . \tag{5}$$

A key feature of eq. (4) is that it allows the weights $p_i(E)$ to be optimized separately for each value of E. This is extremely important, since the accuracy of the information from each histogram depends strongly on E. Since the individual simulations are independent, it is easy to compute the error in $W(E)$ if we know the error in the histograms and choose the weights $p_i(E)$ to minimize that error. Ferrenberg and Swendsen [21] chose to treat the individual entries in a particular histogram as if they were independent of each other and used

$$\delta^2 N_i(E) = g_i \overline{N_i(E)} = g_i W(E)\, n_i \exp(-\beta_i E + f_i) , \tag{6}$$

where

$$g_i = 1 + 2\tau_i \tag{7}$$

and τ_i is the integrated correlation time [22].

In fact, the individual entries are not independent. There is a strong correlation between neighboring entries that contributes significantly to the error in expectation values of the energy and other thermodynamic variables. On the other hand, the actual error in an individual entry is smaller than that given by eqs. (6) and (7). This can be seen by considering that the probability

$$q_i(E) = \frac{N_i(E)}{\Sigma_{E'} N_i(E')} \tag{8}$$

of a given step in the ith Monte Carlo simulation producing the particular energy E is relatively small. Any particular value of E occurs at relatively long time intervals, during which the correlations decay. Instead of the value of g_i in eqs. (6) and (7), the true factor would be of the order $1 + 2\tau_i q_i(E)$, which is usually rather close to unity.

Therefore, the purpose of using eqs. (6) and (7) is to compensate for the neglect of correlations between the entries in the histograms by magnifying the error attributed to the individual entries. For the calculation of the energy at the temperature of the original simulation, it can be shown that this compensation is exact [22]. For other quantities it is clearly in approximation. A comforting observation is that the final results of using eqs. (6) and (7) are very weakly dependence on the specific values of g_i used.

Using eq. (6), it is easy to find the weights, $p_i(E)$, and derive the optimized

estimate for the density of states

$$W(E) = \frac{\sum_{i=1}^{R} g_i^{-1} N_i(E)}{\sum_{j=1}^{R} g_j^{-1} n_j \exp(f_j - \beta_j E)} \,. \tag{9}$$

The f_i's, which give the free energies at the simulation temperatures, are then found self-consistently from

$$\exp(-f_i) = \sum_E W(E) \exp(-\beta_i E) \,. \tag{10}$$

The solution can be found simply by iteration, although in some cases it is useful to apply acceleration techniques [23]. Since the equations are invariant on adding a constant to all of the f_i's, only $R - 1$ independent constants need be determined for the full set of histograms.

These equations are very closely related to those derived by Bennett in 1976 for a more restricted application [24]. Bennett was concerned with the question of finding the optimal estimate of the free energy difference between two temperatures given a Monte Carlo simulation at each temperature. If eqs. (9) and (10) are applied to this question (setting $g_1 = g_2$), they reduce exactly to Bennett's result[#1].

Ferrenberg and Swendsen also found a very simple expression for the relative error in the density of states [21],

$$\frac{\delta W(E)}{W(E)} = \left(\sum_{i=1}^{R} g_i^{-1} N_i(E) \right)^{-1/2} \,. \tag{11}$$

This result has two nice features. First, it demonstrates that the accuracy of the method is always improved by adding more data, which is not necessarily true of other methods. Secondly, plotting this expression as a function of E makes it very easy to plan a sequence of simulations to cover the entire energy (and temperature) range with uniform accuracy. A peak in this plot shows immediately where the next simulation should be performed, and the height of the peak indicates in advance how long the simulation should be.

Once the approximate density of states has been calculated, everything that can be done with a single histogram can also be done with multiple histograms. The additional advantage is, of course, that information over a very wide range of temperatures (or other parameters) can be obtained. In particular, this provides an accurate way of calculating the entropy, which has been a traditional problem for Monte Carlo simulations. For an Ising model on a

[#1] The relationship between the multiple-histogram equations and Bennett's work was pointed out by Park [25].

16×16 lattice, it has been shown that the entropy can be calculated over the entire temperature range from zero to infinity from just eight simulations. Using only a moderate amount of CPU time on a workstation, a four significant digits accuracy was achieved, which was far more than sufficient to see that the zero-temperature entropy does not vanish for this finite system, but takes on the value $(\ln 2)/256$ since the ground state is two-fold degenerate [21].

Two alternatives to eqs. (6) and (7) have been suggested. Alves, Berg and Villanova [26] have performed simulations using 64 replicas of their system at each temperature to obtain a direct estimate of the error in $N_i(E)$. This does indeed give the correct error for the individual entries (which is smaller than that given in eqs. (6) and (7)), but it totally neglects the correlations between entries in the histogram. Alves, Berg and Villanova also used a different method for combining data from histograms that does not produce full optimization and requires a separate, non-self-consistent determination of the f_i's at each value of the energy, E [26].

The second alternative was suggested by Huang, Moriarity, Myers and Potvin [27], who used the same g_i's as Ferrenberg and Swendsen [21] but used $N_i(E)$ to estimate the error instead of using the second part of eq. (6) to combine information from all simulations. They had to treat the case $N_i(E) = 0$ separately, since their equations would imply that the error was zero. The alternative equations are no simpler than the ones presented below, and I believe that the use of the full information on the density of states is more accurate than using only the information from a single histogram.

4. Applications

Since I do not have space to mention all studies involving histogram methods, I will just use a few to illustrate the power of the approach.

The high accuracy achievable with histograms is clearly illustrated by the recent calculations of Ferrenberg and Landau on the critical behavior of the three-dimensional Ising model [28]. Using histograms to evaluate a wide variety of quantities, especially Binder's fourth-order magnetization cumulant [15,16] and the logarithmic derivatives of powers of the magnetization, they were able to determine the critical temperature and the critical exponents to an accuracy matched only by the best series-expansion analyses and Monte Carlo renormalization-group calculations. The key feature was the high accuracy determination of the location of the peaks in the various quantities they observed. However, it should be noted that they also found that the limited range of validity of histogram extrapolations proved to be the most important

restriction on the u
quantities lay at ter
in the locations at t
by using the info
calculation using cl
significant digits ac

The efficiency of
three-dimensional
Swendsen [29,30].
simulation method
analyzed their resu
calculate ground-st
dimensions and co
comparison with e
dimensional model
transition to be ir
model.

One of the mo
methods has been
of polymer mixture
from the histogra
phase diagrams, ir
exponents and am

A new area of
properties of biolo
free-energy differe
free energy on a
application of mul
man and Rosenbe
with the puckering

5. First-order tra

One of the first
method for detern
cumulant [15,16]

$$U_L = 1 - \frac{\langle E^4 \rangle}{3\langle E^2 \rangle^2}$$

to investigate the

restriction on the ultimate accuracy of the calculation. The peaks of different quantities lay at temperatures that were sufficiently different to produce errors in the locations at the sixth significant digit. Ferrenberg and Landau noted that by using the information obtained on the location of the peaks, a new calculation using cluster simulation methods might be able to produce a seven significant digits accuracy.

The efficiency of the histogram approach was exploited in studies of two- and three-dimensional antiferromagnetic Potts models by Wang, Kotecky and Swendsen [29,30]. For this work they introduce a new cluster Monte Carlo simulation method specifically designed for antiferromagnetic Potts models and analyzed their results with the multiple-histogram method. They were able to calculate ground-state entropies to four significant digits in both two and three dimensions and confirmed the accuracy of their results in two dimensions by comparison with exact results. They also used the histograms for the three-dimensional model to do an accurate finite-size scaling study, which showed the transition to be in the same universality class as the three-dimensional XY model.

One of the most extensive applications of single- and multiple-histogram methods has been performed by Deutsch and Binder [31,32] in their analyses of polymer mixtures. Because of the increased amount of information obtained from the histogram methods, they were able to investigate a wide range of phase diagrams, including the locations of critical points, and the values of exponents and amplitudes within the context of a grand-canonical simulation.

A new area of application of histogram methods is in the calculation of properties of biological molecules. In this field it is often important to calculate free-energy differences and "potentials of mean force" (the dependence of the free energy on a particular coordinate of interest). As an example of the application of multiple histogram methods Kumar, Bouzida, Swendsen, Kollman and Rosenberg have calculated the potential of mean force associated with the puckering of the sugar ring in adenosine [33].

5. First-order transitions

One of the first applications of the high efficiency of the single histogram method for determining the location and height of peaks used Binder's reduced cumulant [15,16]

$$U_1 = 1 - \frac{\langle E^4 \rangle}{3\langle E^2 \rangle^2} \tag{12}$$

to investigate the first-order transition in the two-dimensional, eight-state Potts

model [13]. This cumulant takes on the value $\frac{2}{3}$ away from the transition, but goes to a non-trivial value exactly at the transition in the thermodynamic limit. For finite systems this produces minima on a very fine temperature scale. Because the peaks are so narrow (the full width of the peak for a 64×64 system is less than 0.2% of the transition temperature), they are well suited for accurate determination of the transition temperature. However, for the same reason, it is very difficult to find the extremum with a series of independent points. With the histogram analysis, one simulation proved sufficient to determine the location of the peak for each lattice size to better than four significant digits. Data from the same simulations were used to investigate the finite-size scaling of the specific heat and were easily able to show the exponentially small corrections to scaling.

Recently, Lee and Kosterlitz used histograms to make a significant advance in the analysis of first-order transitions [17,18]. They considered the double-peaked shape of the histogram itself at a transition and adjusted the temperature to make the two peaks of equal height. The ratio of the peak height to that of the minimum represents a free-energy barrier that scales (asymptotically) as L^{d-1} at the transition. They then noted that even when the asymptotic limit has not yet been reached, the qualitative behavior of the free-energy barrier is sufficient to determine the order of the transition. They provided a very convincing demonstration for the five-state Potts model in two dimensions, for which the correlation length is greater than 1000 lattice constants at the transition, which is well beyond the range of current simulations. Nevertheless, they were able to demonstrate clearly that the method was able to see that the transition was indeed first order.

This development leaves the question of how to determine the minimum in the histogram at a first-order transition. This has recently been solved by Berg and Neuhaus, who introduced a new simulation method, which they called a "multicanonical ensemble" [19].

Berg and Neuhaus noted that the low minimum in the histogram corresponds to the infrequent transitions between the two states represented by the peaks. By adding a term to an effective Hamiltonian that corresponds to the logarithm of the probability distribution between the peaks, they were able to compensate for the minimum; their simulations produce nearly flat histograms. The increased probability of being in the energy range between the peaks greatly reduces the transition time for the system to switch between the two phases. In their simulations, Berg and Neuhaus found speedups of one to two orders of magnitude. Finally, the enormously improved statistics in the intermediate region provides greatly improved estimates for the free-energy barrier. For the two-dimensional, ten-state Potts model, they found the interfacial free energy per unit area with an error of less than 1%.

6. Conclusions

Although the h
complicated than
they are straightfo
computer time to
analysis of the da
from the simulati
other information
dard practice for

Acknowledgement

I would like to
Steven Teitel for
during my visit
supported by the
Jerusalem, Israel.
Institute in Reho
would like to ack
No. DMR-900947

References

[1] Z.W. Salsburg, J
[2] D.A. Chesnut a
[3] I.R. McDonald
[4] J.P. Valleau and
[5] G. Torrie and J.
[6] M. Falcioni, E.
331.
[7] E. Marinari, Nu
[8] G. Bhanot, S. B
[9] G. Bhanot, K.M
[10] G. Bhanot, K.M
[11] G. Bhanot, R. S
[12] M. Karliner, R.
[13] A.M. Ferrenber
[14] A.M. Ferrenber
[15] K. Binder, Phys
[16] K. Binder, Z. P
[17] J. Lee and J.M.
[18] J. Lee and J.M.

om the transition, but
thermodynamic limit.
ne temperature scale.
he peak for a 64 × 64
hey are well suited for
lowever, for the same
series of independent
proved sufficient to
ze to better than four
used to investigate the
ily able to show the

e a significant advance
onsidered the double-
adjusted the tempera-
of the peak height to
r that scales (asymp-
l that even when the
tative behavior of the
of the transition. They
ate Potts model in two
ater than 1000 lattice
nge of current simula-
ly that the method was

rmine the minimum in
ly been solved by Berg
d, which they called a

histogram corresponds
resented by the peaks.
sponds to the logarithm
ey were able to com-
rly flat histograms. The
ween the peaks greatly
een the two phases. In
f one to two orders of
es in the intermediate
energy barrier. For the
interfacial free energy

6. Conclusions

Although the histogram methods discussed in this paper are somewhat more complicated than traditional analysis methods for Monte Carlo simulations, they are straightforward to code and require relatively insignificant amounts of computer time to use. Once they have been programmed, they speed up the analysis of the data and greatly increase the amount of information available from the simulation. Since this information is available without sacrificing any other information desired, the histogram methods are rapidly becoming standard practice for all Monte Carlo simulations.

Acknowledgements

I would like to thank Alan Ferrenberg, David Landau, Eytan Domany and Steven Teitel for interesting and valuable discussions. This paper was written during my visit to the Weizmann Institute in Rehovot, Israel, which was supported by the United States–Israel Binational Science Foundation (BSF), Jerusalem, Israel. I would also like to thank Prof. Domany and the Weizmann Institute in Rehovot, Israel for their kind hospitality during my visit. Finally, I would like to acknowledge support by the National Science Foundation Grant No. DMR-9009475.

References

[1] Z.W. Salsburg, J.D. Jackson, W. Fickett and W.W. Wood, J. Chem. Phys. 30 (1959) 65.
[2] D.A. Chesnut and Z.W. Salsburg, J. Chem. Phys. 38 (1963) 2861.
[3] I.R. McDonald and K. Singer, Discuss. Faraday Soc. 43 (1967) 40.
[4] J.P. Valleau and D.N. Card, J. Chem. Phys. 57 (1972) 5457.
[5] G. Torrie and J.P. Valleau, Chem. Phys. Lett. 28 (1974) 578; J. Comput. Phys. 23 (1977) 187.
[6] M. Falcioni, E. Marinari, M.L. Paciello, G. Parisi and B. Taglienti, Phys. Lett. B 108 (1982) 331.
[7] E. Marinari, Nucl. Phys. B 235 [FS11] (1984) 123.
[8] G. Bhanot, S. Black, P. Carter and R. Salvador, Phys. Lett. B 183 (1987) 331.
[9] G. Bhanot, K.M. Bitar, S. Black, P. Carter and R. Salvador, Phys. Lett. B 187 (1987) 381.
[10] G. Bhanot, K.M. Bitar and R. Salvador, Phys. Lett. B 188 (1987) 246.
[11] G. Bhanot, R. Salvador, S. Black and R. Toral, Phys. Rev. Lett. 59 (1987) 803.
[12] M. Karliner, R. Sharpe and Y.F. Chang, Nucl. Phys. B 302 (1988) 204.
[13] A.M. Ferrenberg and R.H. Swendsen, Phys. Rev. Lett. 61 (1988) 2635.
[14] A.M. Ferrenberg and R.H. Swendsen, Comput. Phys. (Sept/Oct 1989) 101.
[15] K. Binder, Phys. Rev. Lett. 47 (1981) 693.
[16] K. Binder, Z. Phys. B 43 (1981) 119.
[17] J. Lee and J.M. Kosterlitz, Phys. Rev. Lett. 65 (1990) 137.
[18] J. Lee and J.M. Kosterlitz, Phys. Rev. B 43 (1991) 3265.

[19] B.A. Berg and T. Neuhaus, Phys. Rev. Lett. 68 (1992) 9.
[20] J.M. Rickman and S.R. Phillpot, Phys. Rev. Lett. 66 (1991) 349.,
[21] A.M. Ferrenberg and R.H. Swendsen, Phys. Rev. Lett. 63 (1989) 1195.
[22] H. Müller-Krumbhaar and K. Binder, J. Stat. Phys. 8 (1973) 1.
[23] A.M. Ferrenberg, Ph.D. Thesis, Carnegie-Mellon University (1989).
[24] C.H. Bennett, J. Comput. Phys. 22 (1976) 245.
[25] H. Park, private communication.
[26] N.A. Alves, B.A. Berg and R. Villanova, Phys. Rev. B 41 (1990) 383.
[27] S. Huang, K.J.M. Moriarity, E. Myers and J. Potvin, Z. Phys. C 50 (1991) 221.
[28] A.M. Ferrenberg and D.P. Landau, Phys. Rev. B 44 (1991) 5081.
[29] J.S. Wang, R.H. Swendsen and R. Kotecky, Phys. Rev. Lett. 63 (1989) 109.
[30] J.S. Wang, R.H. Swendsen and R. Kotecky, Phys. Rev. B 42 (1990) 2465.
[31] H.-P. Deutsch, J. Stat. Phys. 67 (1992) 1039.
[32] H.-P. Deutsch and K. Binder, Macromolecules, in press.
[33] S. Kumar, D. Bouzida, R.H. Swendsen, P.A. Kollman and J.M. Rosenberg, J. Comput. Chem., in press.

PHYSICAL REVIEW

LETTERS

VOLUME 63 | 18 SEPTEMBER 1989 | NUMBER 12

Optimized Monte Carlo Data Analysis

Alan M. Ferrenberg[a] and Robert H. Swendsen

Department of Physics, Carnegie-Mellon University, Pittsburgh, Pennsylvania 15213

(Received 17 February 1989)

We present a new method for optimizing the analysis of data from multiple Monte Carlo computer simulations over wide ranges of parameter values. Explicit error estimates allow objective planning of the lengths of runs and the parameter values to be simulated. The method is applicable to simulations in lattice gauge theories, chemistry, and biology, as well as statistical mechanics.

PACS numbers: 05.50.+q, 64.60.Fr, 75.10.Hk

Recently, we showed that histograms can be used to greatly increase the amount of information obtained from a single computer simulation in the neighborhood of a critical point.[1] In particular, we demonstrated that the location and height of maxima and minima can be determined with higher accuracy and much less computer time than previously obtainable. We also noted that the region of validity of the single histogram method coincides with the finite-size scaling region, so that information about the critical region does not deteriorate with increasing system size.[2]

On the other hand, for more general problems it is often desired to investigate the behavior of the system over a wider range of parameter values. In this situation, it is necessary to perform simulations at more than one value of the parameters of interest.

This paper presents an optimized method for combining the data from an arbitrary number of simulations to obtain information over a wide range of parameter values in the form of continuous functions. The method goes beyond earlier methods[3] in that it provides an optimized combination of data from different sources, and can be applied to an arbitrary number of simulations. Errors can be calculated and provide a clear and simple guide to optimizing the length and location of additional simulations to provide maximum accuracy.

It is possible to extend the temperature range to generate the free energy and entropy from zero to infinite temperature.

Finally, our method can be used with any simulation method that provides data for a system in equilibrium, and it requires a negligible amount of additional computer time for its implementation. The method is therefore applicable to simulations in lattice gauge theories,[4] chemistry, and biology,[5,6] as well as statistical mechanics.

We will first describe the method and then demonstrate its efficiency by calculating the properties of an $L=16$, $d=2$ Ising model for the full range of temperatures.

Consider the general Hamiltonian

$$H(\sigma)=H_0(\sigma)+KS(\sigma),$$

where $S(\sigma)$ is an operator (energy, magnetization, etc.) defined on the spins $\{\sigma_i\}$, and factors of $-1/k_BT$ have been absorbed. We are interested in the behavior of the system as a function of K. For simplicity, we will consider only one parameter K, but the generalization to an arbitrary number of parameters is straightforward. The partition function is given by

$$Z(K)=\sum_{\sigma}\exp[H(\sigma)]=\sum_{S}W(S)[KS],$$

where $W(S)$ is the density of states.

Consider R Monte Carlo simulations. We perform the nth simulation at K_n and store the data as histograms, $\{N_n(S)\}$, with total numbers of values $\{n_n\}$. Errors are given by

$$\delta^2N_n(S)=g_n\overline{N_n(S)},$$

where we have used a bar over an expression to indicate the expectation value with respect to all Monte Carlo

12

(MC) simulations of length n_n. If successive MC configurations are independent, then $g_n = 1$, otherwise

$$g_n = 1 + 2\tau_n ,$$

where τ_n is the correlation time.

We can approximate the behavior of the partition function by

$$z_n(K) = \sum_S N_n(S) \exp[(K - K_n)S]$$

which is related to the true partition function by

$$\overline{z_n(K)} = n_n Z(K)/Z(K_n) .$$

The free energy is given by

$$F(K) - F(K_n) = \ln \overline{z_n(K)} - \ln n_n .$$

The density of states is related to the histogram by

$$W(S) = \overline{N_n(S)} n_n^{-1} \exp[f_n - K_n S] ,$$

where $f_n = F(K_n)$ is a parameter equal to the free energy at K_n, and will be evaluated self-consistently. If we perform simulations on a set of values $\{K_n \mid n = 1, R\}$, we can combine them to form a general expression which leads to an improved estimate for $W(S)$. This gives us

$$W(S) = \sum_{n=1}^{R} p_n(S) N_n(S) n_n^{-1} \exp[f_n - K_n S] \qquad (1)$$

with

$$\sum_{n=1}^{R} p_n(S) = 1 .$$

If we insert the actual histograms in (1) and minimize the error in the resultant estimate for $W(S)$,[7] we find

$$p_n(S) = \frac{n_n g_n^{-1} \exp[K_n S - f_n]}{\sum_{m=1}^{R} n_m g_m^{-1} \exp[K_m S - f_m]} . \qquad (2)$$

If we then define

$$P(S,K) = W(S) \exp[KS] ,$$

we obtain the essential multiple-histogram equations as

$$P(S,K) = \frac{\sum_{n=1}^{R} g_n^{-1} N_n(S) \exp[KS]}{\sum_{m=1}^{R} n_m g_m^{-1} \exp[K_m S - f_m]} , \qquad (3)$$

where

$$\exp[f_n] = \sum_S P(S,K_n) . \qquad (4)$$

The average value of any operator on S can then be evaluated as a function of K using

$$\langle A(S) \rangle (K) = \sum_S A(S) P(S,K)/z(K) ,$$

where

$$z(K) = \sum_S P(S,K) .$$

The values of f_n are found self-consistently by iterating (3) and (4). Efficient convergence is obtained by using the derivatives of the new values of f_n as functions of the old values in the iteration process. Note that from the form of Eqs. (3) and (4), an arbitrary constant can be added to each f_n without affecting the solution. This constant can be determined by evaluating the free energy at zero or infinite temperature, where it is known. Alternatively, it can be set to zero at some reference point, which is useful when applying the method to the calculation of interfacial free energies.

The statistical error in $P(S,K)$ is given by

$$\delta P(S,K) = \left[\sum_n g_n^{-1} N_n(S) \right]^{-1/2} P(S,K) \qquad (5)$$

from which it is clear that this method always reduces the statistical errors when additional MC simulations are added to the analysis. This expression also provides a clear guide for planning a series of simulations. The locations and heights of peaks in the relative error, plotted as a function of S, give direct quantitative indications of the optimum locations and lengths of additional MC simulations.

As a practical matter, it is useful to handle most of the calculations in terms of the logarithms of the various quantities in these equations. Also, if there is insufficient computer memory to save an entire two-dimensional histogram, the average of the magnetization and its square can be stored as functions of S to determine the magnetic susceptibility. It is sometimes preferable to store MC data in terms of lists rather than histograms, with obvious modifications of the equations. This will be preferable, for example, if $qn_n < r^q$, where q is the number of operators and r is the number of values each operator can assume.

If the method is restricted to two MC simulations, the calculated difference in the free energies between the simulated points

$$F(K_1) - F(K_2) = f_1 - f_2$$

is identical to that obtained by Bennett's method.[7,8]

We have tested this multiple-histogram method against the exact solution of the $d = 2$ Ising model with $L = 16$. The Hamiltonian is given by

$$H = K \sum_{(i,j)} \sigma_i \sigma_j ,$$

where the spins take on the values $+1$ and -1, and the sum is over all nearest-neighbors pairs.

For temperatures close to the critical temperature, a single simulation is sufficient for high accuracy. However, for temperatures more than about 20% away from the critical temperature, additional simulations are necessary. To improve the accuracy away from K_c, we added simulations at $K = 0.3$ and 0.64 to the first simulation at $K_c = 0.4406868$. Then, using plots of the relative error from (5) as a guide, we studied the results of adding

13

simulations at $K=0.0$, 0.1, 0.2, 0.375, and 0.525 for a total of eight. We simulated the model using the Swendsen-Wang algorithm,[9] taking 2×10^6 MC sweeps at each temperature, except for T_c, where we used 9.5×10^6 sweeps. The value of g_n at T_c was estimated to be 6, with smaller values at other temperatures.

The difference between the calculated and exact values of the specific heat over the full temperature range is less than 0.3%, and would differ from the exact values[10] by less than the width of the line if we were to include a plot with this paper.

The results for the entropy show dramatic improvement as data from additional MC simulations are included in the calculation. The entropy is given by the expression

$$\mathcal{S}=\ln z(K)-K_z^{-1}(K)\sum_S SP(S,K)$$

so that entropy differences can be calculated directly. As an example, consider the $d=2$ Ising model on an L by L lattice, for which the total differences in the entropy per site between zero and infinite temperature is $(1-L^{-1})\ln2$. For $L=16$, this has the value 0.69044. When the data are restricted to a simulation at the critical temperature, the calculated entropy difference is 0.591, which is off by 15%. Adding data from the simulations at $K=0.3$ and 0.64 gave a value of 0.6746, which already reduced the error to 2.3%. Finally, using data from all eight simulations, we obtained 0.69030 with a remaining error of only 0.02%. This is an order of magnitude better than would be necessary to see the L^{-2} term due to the twofold degeneracy at $T=0$.

Another application of the method, for which we already have preliminary results, is the calculation of the free energy of a seam of bonds with coupling $K'=\alpha K$. The point $\alpha=1$ corresponds to the usual periodic boundary conditions, $\alpha=0$ to free boundaries, and $\alpha=-1$ to antiperiodic boundary conditions. By taking two-dimensional histograms for the seam energy and the total energy of the system, we can calculate the surface free energy and the interface free energy as functions of the temperature and α. Taking MC data for just the two points $\alpha=1$ and -1, the results are quite good for the smallest lattices ($L=4$ and 6) as expected from the successes of Bennett's method.[7] However, even here the weakest part of the calculation is for seam energies near zero, and the addition of data from a third simulation at $\alpha=0$ provides improvement. For system sizes up to $L=8$, four simulations are sufficient to reproduce the interface free energy with an error of 0.05% (and agreement with the exact result). The same data give the surface free energy with an even smaller statistical error. By extrapolating the interface free energy as a function of L we can estimate the interface free energy per unit length in the thermodynamic limit as a function of temperature. We find that even restricting ourselves to data between $L=4$, 6, and 8, the interfacial free energy vanishes at a temperature within 0.2% of the exact value. This extension of the calculation to temperatures other than those at which the simulations were performed does not require additional parameters beyond the set $\{f_n\}$, which were already determined.

Since this method is able to combine Monte Carlo simulation data from different sources to increase the total accuracy of the results, it could even be used to combine data from different groups working on large problems, such as those encountered in lattice gauge theories. Because there are no limitations on the method of simulation, we also expect this approach to be useful for simulations in chemistry and biology.

We would like to thank H. Park for helpful discussions, and for calling our attention to the relationship between our results and those of Bennett. This work was supported by the National Science Foundation Grant No. DMR-8613218.

---

[a]Present address: Center for Simulational Physics, Department of Physics and Astronomy, University of Georgia, Athens, GA 30602.

[1]A. M. Ferrenberg and R. H. Swendsen, Phys. Rev. Lett. **61**, 2635 (1988).

[2]The maximum shift of the energy scales as the width of the energy distribution, which is related to the specific heat C. Since the specific heat also relates the change in energy to the corresponding change in temperature, the maximum temperature change can be easily seen to scale as $C^{-1/2}L^{-d/2}$, where L is the size, and d is the dimension of the system. At the critical point, C diverges as $L^{\alpha/\nu}$, so that the range of temperatures scales as $L^{-1/\nu}$, which coincides with the scaling region. Similar consideration of other variables shows that this scaling is a general result.

[3]Z. W. Salsburg, J. D. Jackson, W. Fickett, and W. W. Wood, J. Chem. Phys. **30**, 65 (1959); D. A. Chestnut and Z. W. Salsburg, J. Chem. Phys. **38**, 2861 (1963); I. R. McDonald and K. Singer, Discuss. Faraday Soc. **43**, 40 (1967); I. R. McDonald and K. Singer, J. Chem. Phys. **47**, 4766 (1967); J. P. Valleau and D. N. Card, J. Chem. Phys. **57**, 5457 (1972); G. Torrie and J. P. Valleau, Chem. Phys. Lett. **28**, 578 (1974); M. Falcioni, E. Marinari, M. L. Paciello, G. Parisi, and B. Taglienti, Phys. Lett. **108B**, 331 (1982); E. Marinari, Nucl. Phys. **B235 [FS11]**, 123 (1984); G. Bhanot, S. Black, P. Carter, and R. Salvador, Phys. Lett. B **183**, 331 (1987); G. Bhanot, K. M. Bitar, S. Black, P. Carter, and R. Salvador, Phys. Lett. B **187**, 381 (1987); G. Bhanot, K. M. Bitar, and R. Salvador, Phys. Lett. B **188**, 246 (1987); G. Bhanot, R. Salvador, S. Black, P. Carter, and R. Toral, Phys. Rev. Lett. **59**, 803 (1987).

[4]K. G. Wilson, Phys. Rev. D **10**, 2445 (1974); in *Recent Developments in Gauge Theories*, edited by G. 't Hooft *et al.* (Plenum, New York, 1980).

[5]J. A. McCammon and S. C. Harvey, *Dynamics of Proteins*

and Nucleic Acids (Cambridge Univ. Press, New York, 1987), and references therein.

[6]C. H. Brooks, III, M. Karplus, and B. M. Pettitt, in *Proteins: A Theoretical Perspective of Dynamics, Structure and Thermodynamics, Advances in Chemical Physics*, edited by I. Prigogine and S. Rice (Wiley, New York, 1988), Vol. 71.

[7]C. H. Bennett, J. Comput. Phys. **22**, 245 (1976).

[8]H. Park (private communication).

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

[10]A. E. Ferdinand and M. E. Fisher, Phys. Rev. **185**, 832 (1969).

Optimized Monte Carlo Data Analysis

Alan M. Ferrenberg and Robert H. Swendsen

The use of Monte Carlo (MC) methods to study physical systems was introduced by Metropolis *et al.*[1] over 35 years ago. Since that time, MC methods have been used extensively in the study of phase transitions,[2] lattice gauge theories,[3] and chemical and biological systems.[4–6] MC simulations yield estimates for the average values of thermodynamic quantities at particular values of external parameters such as the temperature, magnetic field, and chemical potential. Since we are usually interested in studying the behavior of the system over ranges of the external parameters, it is necessary, using standard MC methods, to perform many simulations for each desired value of the external parameters. The result for each average is a set of discrete points that is usually shown as a continuous line "drawn to guide the eye."

For systems whose behavior depends on more than one external parameter, it is necessary to perform enough simulations to cover a multidimensional region of parameter space. Using standard techniques, a multiparameter scan is extremely time consuming for two parameters and impractical for three or more.

The availability of averages at only particular values of the external parameters is particularly limiting for systems near first- and second-order phase transitions where the system exhibits narrow peaks in various thermodynamic functions. The positions and heights of the peaks provide important information about the nature of the transition but, because standard MC techniques provide only a set of discrete points, the position of these peaks can be determined only approximately.

An additional problem in the study of phase transitions is that the correlation time can become large if the system is near the transition. This problem was discussed by Harvey Gould and Jan Tobochnik in the July/August 1989 issue of *Computers in Physics*.

The approach we discuss here is the use of *histograms* to extract more information from a Monte Carlo simulation. The idea is to use our knowledge of the equilibrium probability distribution at one value of a parameter to determine the probability distribution at another value of the parameter. The application of this idea to MC simulations dates back to 1959. Salsburg *et al.*[7] were the first to discuss how a single histogram of an observable could be used to evaluate any function of that quantity at a neighboring value of the corresponding parameter. However, they used the histogram only at the temperature of the original simulation and did not obtain additional information. Chesnut and Salsburg[8] described the use of histograms to obtain information over a range of continuously varying parameters, but they also did not implement this idea.

To the best of our knowledge, McDonald and Singer[9] were the first workers to use a single histogram method to evaluate thermodynamic functions over a continuous range of temperatures. (They also introduced an alternative to the Metropolis importance sampling method. Their sampling method involves making random changes of the system configuration subject only to an upper limit on the total energy of the system.) They recognized that the range of temperatures for which a single histogram would give reliable results was limited by the width of the measured histogram, although they did not obtain this relationship. They also noticed that the range of temperatures decreased as the system size was increased. From these observations it was concluded, incorrectly as we shall see, that a single histogram was not useful for the study of phase transitions.

In order to understand the single histogram method, we consider a MC simulation of the Ising model. For simplicity, we assume that the external magnetic field is zero so that the only relevant external parameter is the temperature T. Suppose that we do a standard MC run at $T = T_1$ and measure the histogram $H_1(E)$, the number of configurations that have energy E during a run of N_1 MC steps per spin. The probability $P_E(\beta_1)$ that the system has energy E at $\beta_1 = 1/k_B T_1$ is given by

$$P_E(\beta_1) = H_1(E)/N_1 = W(E)[e^{-\beta_1 E}/Z(\beta_1)], \quad (1)$$

where $W(E)$ is the density of states at energy E. The partition function $Z(\beta_1)$ is given by

$$Z(\beta_1) = \sum_s e^{-\beta_1 E} = \sum_E W(E) e^{-\beta_1 E}. \quad (2)$$

Since the histogram $H_1(E)$ is proportional to P_E, a MC estimate for $W(E)$ is given by

$$W(E) = a_1 H_1(E) e^{\beta_1 E}, \quad (3)$$

where a_1 is a proportionality constant. Since $W(E)$ is independent of T, the probability that the system has energy E at $\beta = 1/k_B T$ takes the form

$$P_E(\beta) = H_1(E) e^{-(\beta - \beta_1)E} \times \left(\sum_E H_1(E) e^{-(\beta - \beta_1)E} \right)^{-1}. \quad (4)$$

Alan M. Ferrenberg completed the work described here as part of the requirements for the Ph.D. in physics at Carnegie-Mellon University. He is now a post-doctoral research associate at the Center for Simulational Physics, Department of Physics and Astronomy, University of Georgia, Athens, GA 30602. Robert H. Swendsen is a professor of physics at Carnegie-Mellon University, Pittsburgh, PA 15213.

Since β is a continuous variable, we can estimate the temperature dependence of the average value of any function of E, e.g.,

$$\langle A \rangle = \sum_E A(E) P_E(\beta). \tag{5}$$

The form of the histogram at $T = 2/\ln(1+\sqrt{2})$ for a 16×16 square lattice is shown in Fig. 1.

In Ref. 10 we showed that the single histogram method is much more efficient for studying phase transitions than had been previously believed. We were able to demonstrate that the range of validity of a single histogram taken at the critical point scales in the same way as the finite-size scaling region so that information about the transition does not deteriorate with increasing system size. Hence information about a phase transition can usually be obtained with a single simulation without the need for multistage sampling (described below). We also demonstrated that histogram techniques provide the most accurate method for determining the position and height of peaks associated with a phase transition.

An effort to overcome the limitations of the single histogram method and to use *multiple* histograms was made in 1972 by Valleau and Card.[11] They introduced the idea of multistage sampling in which supplemental or *bridging* distributions are used to provide information in the wings of the original distribution. They recognized that the proportionality constant a_1 in (3) cannot be obtained from a run at a single temperature, but a second histogram can be generated at a different temperature T_2. If T_1 and T_2 are not too different, the corresponding histograms $H_1(E)$ and $H_2(E)$ will overlap over a range of values of E and the ratio a_2/a_1 can be determined by integrating over the overlap region. This procedure can be repeated if additional runs are performed and determines the entire set of a_i values to within a multiplicative constant. If enough bridging distributions are generated, overlap with the histogram from an infinite temperature simulation can be attained. In this limit the total number of states available to the system can usually be found exactly and the proportionality constant a_0 can be determined. The determination of a_0 fixes all the other a_i values so that the partition function and the absolute free energy can be obtained.

An important contribution was made in 1976 by Bennett,[12] although the relevance of his work to the problem of linking MC simulations for use in multistage sampling was not immediately recognized. Bennett considered the problem of computing free-energy differences between two different temperatures and derived equations based on optimized contributions to the estimate of the density of states from each simulation *at each value of the energy*. The result has the form of expectation values of Fermi functions, with the free-energy difference playing the role of the chemical potential.

In 1977, Torrie and Valleau[13] introduced the method of "umbrella sampling" in order to generate probability distributions wider than the Boltzmann distribution. The problem is that importance sampling typically generates a narrow distribution centered about the average value. Hence, in order to use multistage sampling, it is necessary either to perform many runs, or to rely on the overlap of the tails of the distributions where the statistics are poor. Umbrella sampling was used[13] in a study of a Lennard-Jones system by simulating only the repulsive part of the interaction. Although the method led to an increased width, the simulation became much less efficient and the results using umbrella sampling showed little improvement over those obtained with simple multistage sampling.

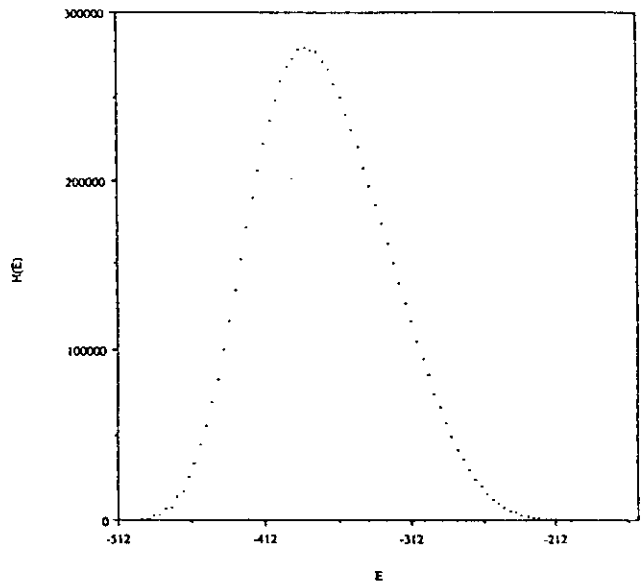


FIG. 1. The energy dependence of the histogram $H(E)$ for the Ising model on a 16×16 square lattice at $T = 2/\ln(1+\sqrt{2})$. A total of 8×10^6 Monte Carlo steps per spin was used.

An important application of histograms for the study of phase transitions was made in 1982 by Falcioni and co-workers[14] and in 1984 by Marinari.[15] These workers used the single histogram method, but extended it to complex temperatures. In this way they were able to compute the zeros of the partition function in the complex temperature plane, and hence obtain additional information about the critical behavior at phase transitions.[16]

Bhanot and co-workers[17–20] wrote a series of papers in 1987 on the application of a combination of MC methods. They used a multistage sampling approach,[11] including the method of matching overlap regions, and a simple sampling method[9] with an upper bound on the total energy. They also introduced a lower bound so that each simulation covered only a narrow range of energies. Karliner *et al.*[21] developed a modification of this approach and maintained a narrow energy range, but reintroduced importance sampling within this range. The use of importance sampling resulted in increased efficiency and lower statistical errors.

Recently, we[22] introduced a new method of optimized multiple histogram data analysis that builds on the

multistage sampling method of Valleau and Card.[11] Our approach is similar to that adopted by Bennett[12] for calculating free-energy differences. In our multiple histogram method, the data from each simulation are combined to form an estimate for the density of states which is optimized for each value of the energy. For simplicity, we will consider only a one-parameter Hamiltonian with T the relevant external parameter. The main result of the method is summarized in (13) and (14).

Suppose that we do R MC simulations. The ith simulation, with N_i MC updates, is performed at $T = T_i$ and yields a histogram $H_i(E)$. The histogram provides an estimate for the equilibrium probability distribution, which we write in the form

$$P(E) = H_i(E)/N_i = W(E)e^{-\beta_i E + f_i}, \tag{6}$$

where f_i is a parameter related to the free energy at T_i by $f_i = \beta_i F(\beta_i)$. (The parameter f_i is related to the unknown values of a_i in the multistage sampling method.[11]) Equation (6) can be inverted to obtain an estimate for the density of states:

$$W(E) = [H_i(E)/N_i]e^{\beta_i E - f_i}. \tag{7}$$

Of course, due to statistical errors, the estimate (7) from one run will be reliable only over some range of E values. Since each of the R simulations will yield a different estimate for $W(E)$, an improved estimate for W can be determined as a weighted sum over each individual estimate for the density of states:

$$W(E) = \sum_{i=1}^{R} p_i(E)H_i(E)N_i^{-1}e^{\beta_i E - f_i}. \tag{8}$$

This estimate for $W(E)$ can be optimized for *each* value of E, by choosing $p_i(E)$ so as to minimize the error in the estimate for W. The uncertainty in the histogram values is given by

$$\delta^2 H_i(E) = g_i \overline{H_i(E)}, \tag{9}$$

where the bar indicates the expectation value with respect to all MC simulations of duration N_i. If the successive MC configurations are independent, then $g_i = 1$; otherwise, we have[23]

$$g_i = 1 + 2\tau_i, \tag{10}$$

where τ_i is the correlation time.

If we minimize the error in the resultant estimate for $W(E)$, we obtain

$$p_i(E) = N_i g_i^{-1} e^{-\beta_i E + f_i}\left(\sum_{i=1}^{R} N_i g_i^{-1} e^{-\beta_i E + f_i}\right)^{-1}. \tag{11}$$

We define

$$P(E,\beta) = W(E)e^{-\beta E} \tag{12}$$

and write the essential multiple histogram equations as

$$P(E,\beta) = \left[\left(\sum_{i=1}^{R} g_i^{-1}H_i(E)\right)e^{-\beta E}\right] \times \left(\sum_{i=1}^{R} N_i g_i^{-1} e^{-\beta_i E + f_i}\right)^{-1}, \tag{13}$$

where

$$e^{-f_i} = \sum_E P(E,\beta_i). \tag{14}$$

Equations (13) and (14) can be iterated to determine the values of f_i self-consistently. The convergence can be accelerated by making use of derivatives of f values on one iteration with respect to those of the previous iteration. If we extend the histograms to sufficiently low or high T where the free energy can be determined exactly, the absolute free energies can be computed.

(As with other Monte Carlo techniques for calculating free energies, (13) and (14) determine the free energy to within an additive constant. For convenience, we can set $f_1 = 0$ and then determine the other f values with respect to f_1.)

The statistical error in $P(E,\beta)$ is given by

$$\delta P(E,\beta) = \left(\sum_i g_i^{-1}H_i(E)\right)^{-1/2} P(E,\beta). \tag{15}$$

From (15) we see that the method always reduces the statistical errors when additional MC simulations are added to the analysis. This expression also provides a clear guide for planning a series of simulations. The positions and heights of peaks in the relative error, plotted as a function of E, give a direct quantitative indication of the optimum locations and durations of additional MC simulations.

Once the values of f_i are determined, (13) can be used to calculate the average value of any function of E as a function of β,

$$\langle A \rangle = \sum_E A(E)P(E,\beta)\left(\sum_E P(E,\beta)\right)^{-1}. \tag{16}$$

In particular, the specific heat C is given by

$$VC(T) = (1/k_B T^2)(\langle E^2 \rangle - \langle E \rangle^2), \tag{17}$$

where V is the volume of the system. The multiple histogram method has already been applied to several MC studies of phase transitions. For the $d = 2$ Ising model the results obtained[22] compared favorably to the exact solution for finite lattices.[24] New results for the three-state antiferromagnetic Potts model, including zero-temperature entropies, were recently obtained by Wang *et al.*[25] using the multiple histogram method. Other recent applications include an SU(2) lattice gauge calculation and calculations for several lattice dimer models.

Since the multiple histogram method is able to combine MC simulation data from different sources to increase the total accuracy of the results, it could be used to combine data from different groups working on large problems, such as those encountered in lattice gauge theories. Because there are no limitations on the method

of simulation, we also expect this method to be useful for simulations in chemistry and biology.

Suggestions for Further Study

1. Use the single histogram method to show that the free-energy difference can be expressed as

$$\beta_2 F(\beta_2) - \beta_1 F(\beta_1) = -\ln \sum_E P(E) e^{-(\beta_2 - \beta_1)E}, \tag{18}$$

where $P(E) = H(E)/N_1$. (The unknown constant a_1 does not appear in the free-energy difference.)

2. Write a program to simulate the $d = 2$ Ising model on a $L \times L$ square lattice at a temperature T in a zero magnetic field. As a check on your program, compare your results for $L = 2$ to the results obtained by an exact enumeration of the 16 possible states. Then choose $L = 4$ and compute $H(E)$ at the critical temperature of the infinite system, $T_c = 2/\ln(1 + \sqrt{2}) \approx 2.269$. Use the single histogram method to estimate P_E at various T up to $T = 3.13$. How do your predicted results for P_E compare to the results for P_E when measured directly? What is the approximate range of applicability of the single histogram method? Use the single histogram at $T = T_c$ to estimate C in the critical region. How do your results for $L = 4$ compare with the exact solution?[24] Is the temperature at which C is a maximum above or below T_c? Repeat the above measurements of $H(E)$ at $T = T_c$ for $L = 8$ and $L = 16$ and determine the L dependence of the maximum of C and applicability of the single histogram method.

3. Consider the Gaussian probability distribution

$$P_E = (1/\sigma)(1/\sqrt{2\pi}) e^{-E - \langle E \rangle)^2/2\sigma^2}, \tag{19}$$

where $\sigma^2 = \langle E^2 \rangle - \langle E \rangle^2$. Assume that the histogram has the Gaussian form (19) at $T = T_1$ and show that this form implies that $C(T) = T_1^2/T^2 C(T_1)$.

4. Choose $L = 16$ and compute $H(E)$ at $T = T_c$. Would it be preferable to use a cluster-flip or single-flip algorithm to obtain new configurations? How well can this histogram be fitted by a Gaussian function? Why would a Gaussian yield a better fit away from T_c?

5. Use the multiple histogram equations (13) and (14) to combine the results from runs at different temperatures. The correlation times can be computed as discussed in the July/August issue of *Computers in Physics*.

6. A system of particles interacting via the interparticle potential $V(r) = \epsilon(\sigma/r)^{12}$ can be characterized by a single dimensionless parameter $\Gamma = \beta V(r = a)$, where $4\pi n a^3/3 = 1$ and n is the particle density. Since the energy of the system is a continuous variable, how can the histogram be computed? Use the multiple histogram method to compute the mean energy and other thermodynamic quantities for Γ in the range 0.1–300. Compare your results with the MC results of Ref. 26. Can the multiple histogram method be used to determine the fluid–solid boundary? (Although computer time can be saved by considering only 32 particles, this problem is computer intensive.)

7. Use the multistage sampling method[11] to combine the histograms of two simulations. Compare the results with those obtained using the multiple histogram method. Which method produces better results?

The success of this column depends on reader input. Please send us your results, comments, and suggestions for future columns. Regular columnists Gould and Tobochnik will be back next issue. Messages can be sent via email to hgould@clarku or tobochnik%heyl.dnet@gw.wmich.edu.

Acknowledgments

The authors would like to thank H. Park, K. Bassler, J.-S. Wang, and M. de Meo for helpful discussions. We also thank Harvey Gould and Jan Tobochnik for inviting us to write this column and for their clarifying suggestions. ■

References

1. N. Metropolis, A. W. Rosenbluth, A. H. Teller, and E. Teller, J. Chem. Phys. 21, 1087 (1953).
2. K. Binder, J. Comput. Phys. 59, 1 (1985).
3. K. G. Wilson, Phys. Rev. D 10, 2445 (1974); *Recent Developments in Gauge Theories*, edited by G. 't Hooft *et al.* (Plenum, New York, 1980).
4. See, for example, the collection of papers in *Simulation of Fluids*, edited by G. Ciccotti, D. Frenkel, and I. McDonald (North-Holland, Amsterdam, 1987).
5. J. A. McCammon and S. C. Harvey, *Dynamics of Proteins and Nucleic Acids* (Cambridge U. P., New York, 1987) and references therein.
6. C. H. Brooks III, M. Karplus, and B. M. Pettitt, "Proteins: A Theoretical Perspective of Dynamics, Structure and Thermodynamics," in *Advances in Chemical Physics*, edited by I. Prigogine and S. Rice (Wiley, New York, 1988), Vol. LXXI.
7. Z. W. Salsburg, J. D. Jackson, W. Fickett, and W. W. Wood, J Chem. Phys. 30, 65 (1959).
8. D. A. Chesnut and Z. W. Salsburg, J. Chem. Phys. 38, 2861 (1963).
9. I. R. McDonald and K. Singer, Discuss. Faraday Soc. 43, 40 (1967).
10. A. M. Ferrenberg and R. H. Swendsen, Phys. Rev. Lett. 61, 2635 (1988).
11. J. P. Valleau and D. N. Card, J. Chem. Phys. 57, 5457 (1972).
12. C. H. Bennett, J. Comput. Phys. 22, 245 (1976).
13. G. Torrie and J. P. Valleau, Chem. Phys. Lett. 28, 578 (1974); J. Comput. Phys. 23, 187 (1977).
14. M. Falcioni, E. Marinari, M. L. Paciello, G. Parisi, and B. Taglienti, Phys. Lett. B 108, 331 (1982).
15. E. Marinari, Nucl. Phys. B 235 [FS11], 123 (1984).
16. C. N. Yang and T. D. Lee, Phys. Rev. 87, 404 (1952).
17. G. Bhanot, S. Black, P. Carter, and R. Salvador, Phys. Lett. B 183, 331 (1987).
18. G. Bhanot, K. M. Bitar, S. Black, P. Carter, and R. Salvador, Phys. Lett. B 187, 381 (1987).
19. G. Bhanot, K. M. Bitar, and R. Salvador, Phys. Lett. B 188, 246 (1987).
20. G. Bhanot, R. Salvador, S. Black, and R. Toral, Phys. Rev. Lett. 59, 803 (1987).
21. M. Karliner, R. Sharpe, and Y. F. Chang, Nucl. Phys. B 302, 204 (1988).
22. A. M. Ferrenberg and R. H. Swendsen, Carnegie–Mellon preprint.
23. H. Müller-Krumbhaar and K. Binder, J. Stat. Phys. 8, 1 (1973).
24. A. E. Ferdinand and M. E. Fisher, Phys. Rev. 185, 832 (1969).
25. J. S. Wang, R. H. Swendsen, and R. Kotecký, Phys. Rev. Lett. 63, 109 (1989).
26. W. G. Hoover, M. Ross, K. W. Johnson, D. Henderson, J. A. Barker, and B. C. Brown, J. Chem. Phys. 52, 4931 (1970).

-5-

Three-state antiferromagnetic Potts models: A Monte Carlo study

Jian-Sheng Wang*
HLRZ, Forschungszentrum Jülich, D-5170 Jülich 1, Federal Republic of Germany

Robert H. Swendsen
Department of Physics, Carnegie Mellon University, Pittsburgh, Pennsylvania 15213

Roman Kotecký
Department of Mathematical Physics, Charles University, V Holešovičkách 2, 180 00 Praha 8, Czechoslovakia
(Received 28 March 1990)

We present a study of the three-state antiferromagnetic Potts model in two and three dimensions, using a cluster-flip Monte Carlo simulation algorithm. The new approach enables us to perform simulations with greatly improved efficiency. We have obtained results for the ground-state entropy and critical exponents in two and three dimensions. The low-temperature phase in three dimensions is shown to have long-range order with a finite-size dependence of the magnetization and susceptibility similar to that of the XY model.

I. INTRODUCTION

In this paper we give a detailed account of our study of the three-state antiferromagnetic Potts model in two and three dimensions, using a cluster-flip algorithm.[1]

Antiferromagnetic Potts models have been shown to possess interesting and unusual properties. The ground-state entropy is nonzero whenever the number of spin states is $q > 2$. The $q = 3$ model on a square lattice has a critical point only at zero temperature.[2-6] In three dimensions, the evidence indicates the existence of phase transitions for $q = 3$, 4, and 5, although the nature of these transitions has been uncertain.[7-12] Notice that the ground-state restrictions are too weak for large q to create an order at low temperatures, and the thermodynamic disorder prevails up to the vanishing temperature. Indeed, using the Dobrushin uniqueness theorem, one can prove[13] that there is no phase transition if $q > 3 \times 2^d$, where d denotes the dimension of the lattice. Other studies have shown that the addition of second-neighbor interaction,[14-17] mixed anisotropic interactions,[18,19] or an external magnetic field[20] can produce new types of ordering and new phase transitions.

The highly degenerate ground states in the antiferromagnetic Potts models could lead to interesting consequences. Berker and Kadanoff[21] suggested from a one-parameter renormalization-group consideration that, similarly to the XY model in two dimensions, a critical low-temperature phase may appear, with an algebraic decay of correlations. However, this conclusion was criticized, suggesting that it is an artifact of the one-parameter renormalization-group treatment.[20] Banavar, Grest, and Jasnow[7] made the first study of the three-dimensional Potts model for $q = 3$ and 4. From a field-theoretic calculation, they conclude that the critical behavior of the three-state model belongs to the universality class of the XY model in three dimensions and the four-state model belongs to the universality class of Heisenberg model, if the transitions are continuous. Their Monte Carlo simulation indicates a continuous transition. They found that there is a nonzero magnetization, also confirmed by Hoppe and Hirst,[8] and recently by Ueno, Sun, and Ono,[12] unlike the behavior suggested by Berker and Kadanoff. On the other hand, Ono[10] suggested that there is no spontaneous magnetization at low temperatures and the low-temperature phase is of the Kosterlitz-Thouless type.[22]

Our Monte Carlo simulation results in two dimensions are consistent with a zero-temperature transition. In three dimensions we find that the critical exponents and low-temperature phase are similar to that of XY model, as proposed in Ref. 7. Our data are much more accurate than previous work[6-8,10,12] due to a new algorithm.[1] In the next section we give a description of our algorithm. Results of simulations for the two- and three-dimensional Potts models are presented in the subsequent sections. We summarize our results in the last section. In the Appendix a careful consideration of order parameters is given.

II. SIMULATION ALGORITHM

The difficulty encountered in the single-spin-flip algorithm[23,24] is the phenomenon of critical slowing down at a second-order phase transition. The correlation time, which is roughly the time needed to generate a statistically independent configuration, measured in Monte Carlo steps, goes as $\tau \propto L^z$, where L is the linear size of the system and z is dynamic critical exponent ($z \approx 2$ for order-parameter nonconserving dynamics[25-27]). This drastically reduces the accuracy of Monte Carlo data, since the statistical error is inversely proportional to the square root of the number of *independent* configurations. The recently developed cluster-flip-type algorithms[28-33] have

been shown to reduce z considerably.[34–36] Thus much better accuracy with less computational efforts can be achieved.

We use[1] a generalization of the algorithm of Swendsen and Wang[28] (SW) to the antiferromagnetic Potts models. This generalization is closely related to Wolff's embedding of Ising reflection variables in $O(n)$ models.[32,37,38] To present the algorithm, let us consider the Potts model defined by the Hamiltonian[39]

$$H=-K\sum_{\langle i,j\rangle}\delta_{\sigma_i,\sigma_j}\ , \tag{1}$$

where the Potts variable σ_i takes the value $1,2,\ldots,q$, and the summation is over nearest-neighbor pairs of sites on a square or cubic lattice. $K=J/k_BT$ is a dimensionless coupling constant, and $J<0$ for antiferromagnetic interactions (we set $J/k_B=-1$). The partition function is given by

$$Z=\sum_{\sigma}e^{-H}\ . \tag{2}$$

Following Edwards and Sokal,[30] our algorithm for the Potts antiferromagnetic model may be introduced as a contraction of a joint probability distribution on the spin configurations $\{\sigma_i\}$, $\sigma_i=1,2,3$, and bond configurations $\{n_{ij}\}$, $n_{ij}=0,1,2,3$, where $\langle i,j\rangle$ runs over nearest-neighbor pairs of the lattice sites. Namely, the joint probability

$$\mu_{\text{joint}}=Z^{-1}\prod_{\langle i,j\rangle}\left[(1-p)\delta_{n_{ij},0}+\sum_{\alpha=1}^{3}p\delta_{n_{ij},\alpha}(1-\delta_{\sigma_i,\sigma_j})(1-\delta_{\sigma_i,\alpha})(1-\delta_{\sigma_j,\alpha})\right], \tag{3}$$

where $p=1-e^{-|K|}$. Such a joint-probability distribution yields the following marginal distributions: For configurations σ it is just the distribution of the original Potts antiferromagnet; for configurations n it is the distribution with the weights

$$\prod_{n_{ij}=0}(1-p)\prod_{n_{ij}=1,2,3}p2^{N_c}$$

(here N_c is the number of α clusters, defined by the constancy of $n_{ij}=\alpha\neq 0$); different α yield different clusters, and for configurations $\bar{n}$ defined by $\bar{n}_{ij}=0$ if $n_{ij}=0$ and $\bar{n}_{ij}=1$ otherwise, the marginal distribution is defined by the weights

$$\prod_{\bar{n}_{ij}=0}(1-p)\prod_{\bar{n}_{ij}=1}p3^{N_c'}$$

(here the clusters are those just mentioned glued together).

Our algorithm then is actually an alternate application of the following two conditional probabilities:

$$P(n|\sigma)=\begin{cases}\text{surely } n_{ij}=0, \text{ whenever } \sigma_i=\sigma_j\\ n_{ij}=\alpha \text{ with probability } p \text{ and } n_{ij}=0 \text{ with probability } 1-p,\\ \quad\text{whenever } \sigma_i\neq\sigma_j,\ \sigma_i\neq\alpha, \text{ and } \sigma_j\neq\alpha,\end{cases}$$

or, more accurately, the probability $P(n_{A_\alpha}|\sigma,n_{A_\alpha^c})$ under the condition that together with σ is fixed a set A_α of bonds on which either $n_{ij}=0$ or α with a fixed configuration on the complement $A_\alpha^c(\langle i,j\rangle\in A_\alpha^c\Rightarrow n_{ij}\neq 0,\alpha)$,

$$P(n|\sigma)=\begin{cases}\text{surely } n_{ij}=0, \text{ whenever } \sigma_i=\sigma_j\\ n_{ij}=\alpha \text{ with probability } p \text{ and } n_{ij}=0 \text{ with probability } 1-p, \text{ whenever } \sigma_i\neq\sigma_j,\ \sigma_i\neq\alpha, \text{ and } \sigma_j\neq\alpha,\end{cases}$$

for every $\langle i,j\rangle\in A_\alpha^c$, and $P(\sigma|n)$ defined as a random distribution that assigns that distribution to each α cluster with equal probability to those two configurations for which $\sigma_i\neq\sigma_j$, $\sigma_i\neq\alpha$, and $\sigma_j\neq\alpha$.

For actual implementation the algorithm consists of the following steps.

(1) One chooses a pair of Potts states among the q different states at random.

(2) Bonds are formed between nearest-neighbor sites occupied by those chosen states if the two sites are in different Potts states, and a uniformly distributed random number in the interval 0 and 1 is less than $p=1-e^{-|K|}$.

(3) Clusters are identified. A cluster can be either a single site or a set of sites connected through bonds. The sites not in the chosen Potts states are not counted as clusters; their Potts variables do not change.

(4) For each cluster, with equal probability, we either keep its original Potts states, or interchange the chosen Potts state on the sites in the cluster. We then go back to step (1).

For $q=2$, this algorithm reduces to the original SW algorithm for antiferromagnetic Ising models, using the concept of antibonds between spins of opposite signs. The algorithm for $q>2$ updates a subset of the lattice

sites, where the reduced Hamiltonian in one-step updating is that of a dilute antiferromagnet. Of course the reduced (effective) Hamiltonian changes in the next Monte Carlo step, determined by the state of the system. If $q \geq 4$, two or more pairs of states can be updated simultaneously. The relation to Wolff's algorithm[32] is clear, since a q-state Potts model can be thought of as an $O(n)$ model with $n = q - 1$ and with unit vectors taking only a discrete sets of values. The reflection of a unit vector with respect to a certain plane is equivalent to the exchange of Potts states.

A key step in implementing this algorithm is a cluster-labeling scheme, so that sites in the same cluster receive the same labeling number, while sites that belong to different clusters have distinct labels. This is done using an auxiliary list of labels. The labeling algorithm is similar to the Hoshen-Kopelman algorithm[40] used in the cluster-counting problem in percolation. An integer array $A(I)$ is initialized to $A(I)=I$, meaning site I has label I. If $A(I)<I$, the label of site I is the same as the label of the site $I'=A(I)$. A proper label at any given moment has the property $A(I)=I$. As one goes through each pair of nearest neighbors, if the condition for having a bond in step (2) is fulfilled, the current labels of the two sites are found iteratively: Site I has the label $A(I)$; if the label is equal to I, then I is the current label; otherwise site I has the same label as site $A(I)$. One traces back this list until $A(J)=J$; then J is the current label of site I. The list is updated so that the labels of the two sites reset to the current, smaller label. A final check is needed to go through the list to ensure that each site has the proper label. The operation needed is proportional to the total number of sites. Thus the computational speed does not slow down as system size increases. The memory requirement is twice the number of sites.

III. DEFINITION OF ORDER PARAMETER AND SUSCEPTIBILITY

Since the interactions are antiferromagnetic, we expect some kind of staggered order that breaks the sublattice symmetry.[14] We define

$$m_\mu = \frac{2}{L^d}\left[\sum_{i\in a}\delta_{\sigma_i,\mu} - \sum_{i\in b}\delta_{\sigma_i,\mu}\right], \tag{4}$$

where μ takes values $1,2,\ldots,q$; L is the linear size of the system; a and b are the two sublattices such that the sites on the sublattice a have their nearest neighbors on the sublattice b and the other way around. For the order parameter we may take

$$\langle m \rangle = \frac{1}{q}\sum_{\mu=1}^{q}\langle |m_\mu| \rangle . \tag{5}$$

Our definition is the same as that introduced in Ref. 14 except a constant factor ($\frac{2}{3}$).

The susceptibility is in the disordered phase given by

$$\chi = \frac{L^d}{q}\sum_{\mu=1}^{q}\langle m_\mu^2 \rangle . \tag{6}$$

By noting that each cluster has freedom to choose, independently of other clusters, two possible new states (step 4), we can explicitly perform an average over all possible assignments of the new configurations. Namely, if N_α^a and N_α^b denote the number of sites in the cluster α on the a and b sublattices, respectively, and $N_\alpha = N_\alpha^a + N_\alpha^b$ is the total number of sites in the cluster, assuming that we are interchanging, say, the states 1 and 2, we get

$$m_1 L^d/2 = \sum_\alpha n_\alpha N_\alpha^a - (1-n_\alpha)N_\alpha^b , \tag{7}$$

where $n_\alpha = 1$ if state 1 is on sublattice a, and 0 otherwise. Averaging over the random variables n_α for which $\langle n_\alpha \rangle = \frac{1}{2}$ and $\langle n_\alpha n_{\alpha'} \rangle = \frac{1}{4} + \frac{1}{4}\delta_{\alpha,\alpha'}$, we obtain

$$\chi = \frac{L^{-d}}{q}\left[\left\langle \sum_\alpha N_\alpha^2 \right\rangle + \left\langle \left| \sum_\alpha (N_\alpha^a - N_\alpha^b) \right|^2 \right\rangle\right], \quad q=3 . \tag{8}$$

This formula has the advantage over (6) that the variance of χ is reduced, since it takes into account many configurations.

Another choice of order parameters is given by considering the three Potts states as a unit vector taking three directions 120° apart. Then we have a two-component order parameter. It has been used by Nightingale and Schick,[3] and by Ono.[10] We show in the Appendix that this latter definition is actually more appropriate. We use this definition for the low-temperature susceptibility.

IV. TWO-DIMENSIONAL RESULTS

A. Zero-temperature simulation

The efficiency of an algorithm is characterized by a correlation time, determined from the equilibrium time-dependent correlation function, defined by

$$f(t) = \frac{\langle m(t+t_0)m(t_0)\rangle - \langle m(t_0)\rangle^2}{\langle m(t_0)^2\rangle - \langle m(t_0)\rangle^2} , \tag{9}$$

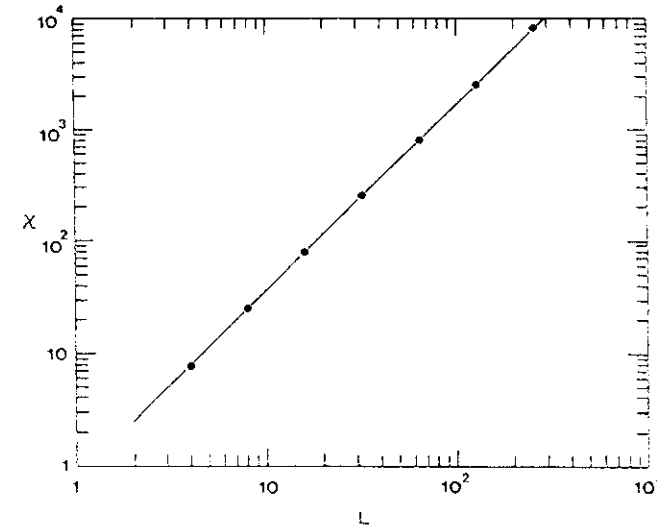


FIG. 1. Log-log plot of the two-dimensional susceptibility χ vs linear size L at $T=0$. The straight line has a slope $\gamma/\nu = \frac{5}{3}$.

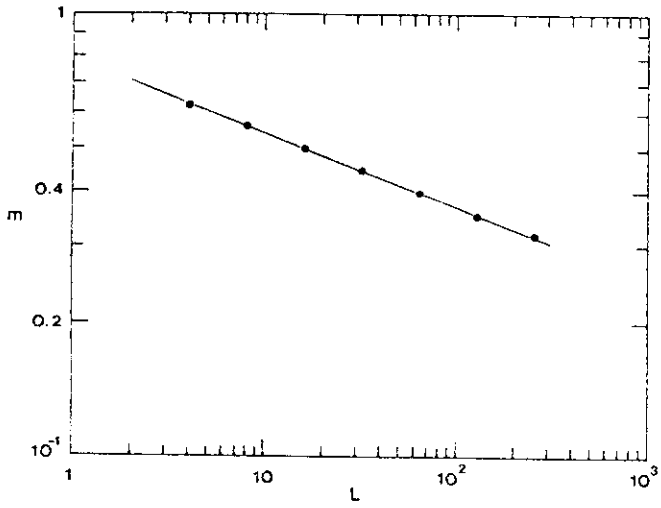


FIG. 2. Log-log plot of the two-dimensional magnetization m vs linear size L at $T=0$. The straight line has a slope $\beta/\nu=\frac{1}{6}$.

where m is the total instantaneous magnetization, defined in Eq. (5), and the angular brackets denote an average over a sequence of configurations generated in a Monte Carlo simulation. The time dependence is, to a very good approximation, exponential, $f(t)\propto e^{-t/\tau}$. The coefficient τ here is the characteristic correlation time.

For the two-dimensional three-state antiferromagnetic Potts model at zero temperature, the standard Monte Carlo gives a correlation time $\tau_{\rm std}\sim 0.32L^z$, with a dynamic critical exponent $z\approx 2.0$. Our algorithm gives a correlation time $\tau\approx 7$ for $L=4-64$. It means that critical slowing down essentially disappears and much more accurate results are obtained.

Figure 1 is a log-log plot of the susceptibility versus size L, for L up to 256. A nearly straight line yields a very accurate estimate of the exponents ratio $\gamma/\nu=1.666\pm 0.002$. Assuming scaling, this corresponds to $\eta=2-\gamma/\nu=0.334\pm 0.002$, which characterizes the decay of pair-correlation function, $g(r)\sim r^{-d+2-\eta}$ $(d=2)$. Of course, the exponent γ or ν separately is not uniquely defined due to a zero-temperature transition. Park and Widom[41] have recently found an exact value $\gamma/\nu=\frac{5}{3}$, confirming our numerical result.

Figure 2 is a log-log plot for the magnetization. It decreases with size as expected, with the exponents ratio $\beta/\nu=0.170\pm 0.006$ (exact value $\beta/\nu=\frac{1}{6}$). The hyperscaling relation $\gamma/\nu+2\beta/\nu=d$ is satisfied within statistical errors.

B. Finite-temperature simulation

To calculate the entropy and other thermodynamic quantities, we used the multiple-histogram method.[42] The central idea of the histogram method[43] is to collect distribution of quantities of interest at one temperature; the value at nearby temperatures is generated according to Gibbs formula. The multiple-histogram method[42] combines simulation results at different temperatures, and regenerates data as a function of parameter in the model (typically temperature) in a continuous smooth fashion. As a by-product, the free energy can also be obtained.

In Figure 3 we plotted the magnetization as a function of temperature for different sizes $L=4$, 8, 16, and 32. At high temperatures $m\propto L^{-d/2}$ is observed. At very low temperatures the magnetization decreases with size as $L^{-\beta/\nu}$ as already discussed in Sec. IV A. Our Monte Carlo data are consistent with no spontaneous magnetization at all temperatures in the infinite-size limit. Earlier results (Ref. 14, Fig. 1) might indicate a nonzero magnetization or nonzero T_c. This could just be a finite-size effect.

Figure 4 is the reduced free energy per site, defined by $f=L^{-d}\ln Z$, plotted as a function of temperature for sizes $L=4$, 8, and 16. In the high-temperature limit f approaches ln3, while the zero-temperature limit is the ground-state entropy. Finite-size effect shows up at low temperatures.

In two dimensions the ground-state entropy is known to be $s_\infty=\frac{3}{2}\ln\frac{4}{3}$ for an infinite system due to mapping onto an ice model.[44] Finite-size correction is given by[41]

$$s(L)\simeq s_\infty+L^{-2}\ln 2.935\,779\,65\ . \tag{10}$$

We obtained $s(4)=0.5000$, $s(8)=0.4484$, and $s(16)$

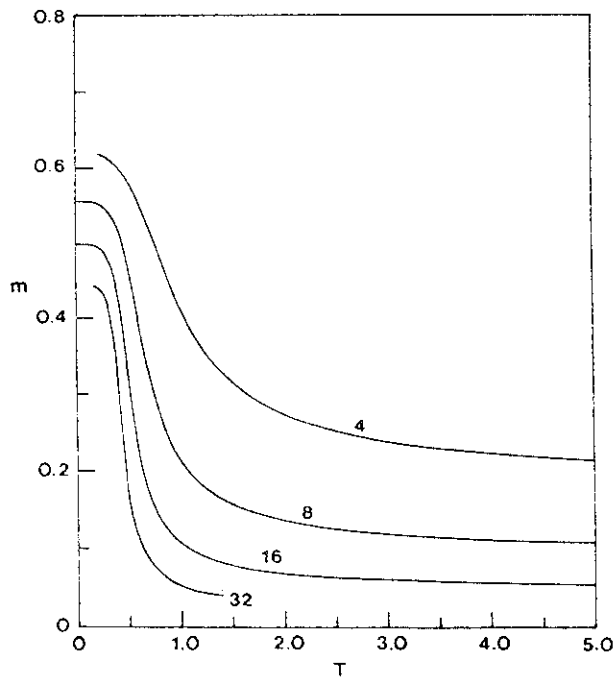


FIG. 3. Magnetization as a function of temperature in two dimensions for size $L=4$, 8, 16, and 32. The smooth curves are obtained using the multiple-histogram method (Ref. 43). This and following plots in two dimensions combine data from five simulations each for $L=4$ and 8, and nine simulations for $L=16$.

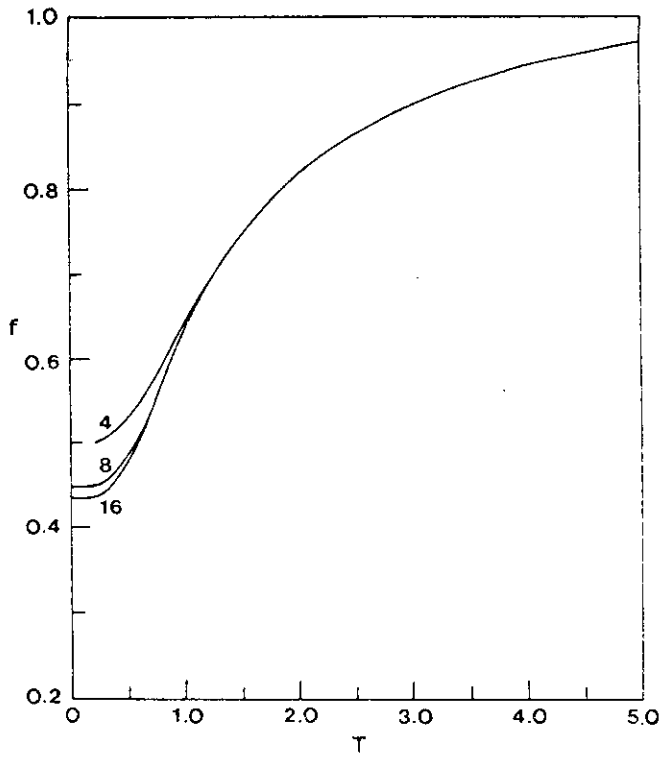


FIG. 4. Reduced free energy $f=L^{-d}\ln Z$ as a function of temperature in two dimensions for size $L=4$, 8, and 16. The value f approaches the ground-state entropy as temperature T approaches zero.

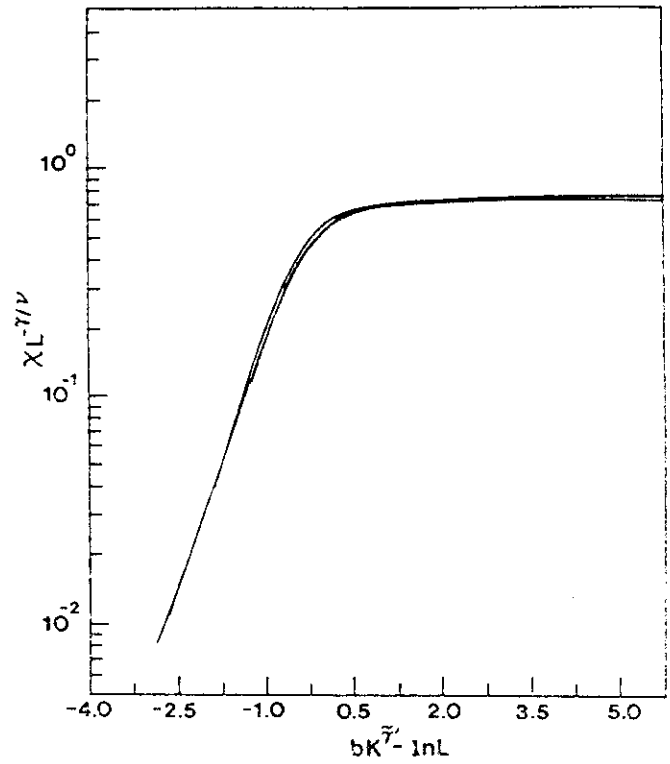


FIG. 6. Scaling plot $L^{-\gamma/\nu}\chi$ vs ξ/L in logarithmic scale, where $\xi=e^{0.90K^{\tilde{\nu}}}$, with $\gamma/\nu=1.666$, $\tilde{\nu}=1.3$, using the same set of data as in Fig. 5.

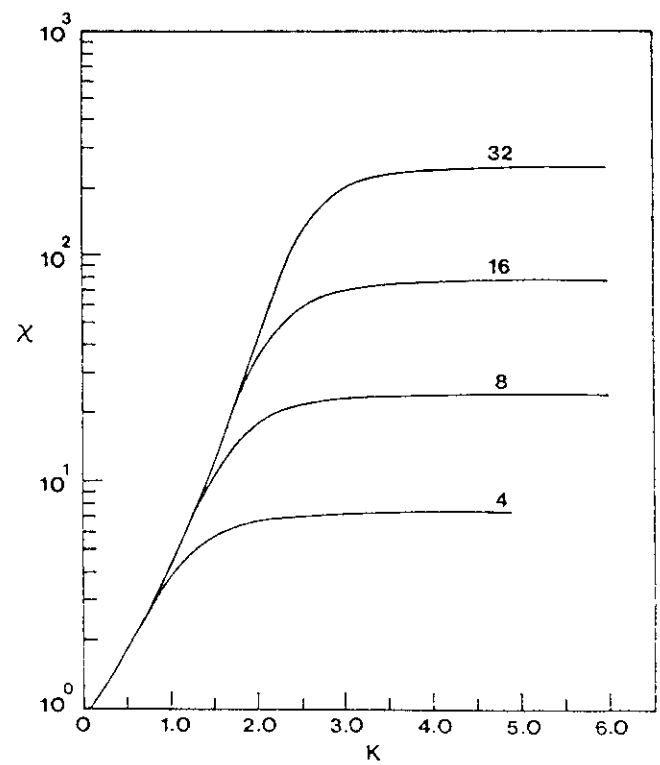


FIG. 5. Susceptibility as a function of coupling strength $K=1/T$ on a semilogarithmic plot for size $L=4$, 8, 16, and 32 in two dimensions.

$=0.4359$ from simulation. These values are in good agreement with exact result in Eq. (10).

In Fig. 5 the susceptibility is plotted against $K=1/T$ on a semilogarithmic scale. A clear curvature is observed, indicating that the susceptibility is growing with K (in the large size limit) faster than a simple exponential. Previous data[6] were analyzed in the form of $\chi\propto|T-T_c|^{-\gamma}$, and it was found $T_c\approx 0$ with $\gamma\approx 5$. Our susceptibility data are not compatible with this functional form.

If we assume a finite-size scaling structure for the susceptibility

$$\chi\simeq L^{\gamma/\nu}\tilde{\chi}(\xi/L)\ , \tag{11}$$

we have to assume that the correlation length takes a form $\xi\simeq e^{bK^{\tilde{\nu}}}$. In the thermodynamic limit the susceptibility then has an essential singularity, $\chi\propto e^{aK^{\tilde{\nu}}}$. Figure 6 is a scaling plot with $\gamma/\nu=1.666$ and $\tilde{\nu}=1.30$. Deviation from scaling is small even for $L=4$. The exponent $\tilde{\nu}$ is in agreement with a phenomenological renormalization-group calculation.[3]

V. THREE-DIMENSIONAL RESULTS

A. Correlation time

Correlation time is, for our algorithm, also reduced at the critical temperature in three dimensions. We found that the correlation time is approximately given by

$\tau \approx 2L^{0.48\pm0.04}$ ($L \leq 32$), while for the standard simulation it is found to be $\tau_{\text{std}} \approx \frac{1}{6}L^{2.00\pm0.01}$ ($L \leq 16$). Thus the cluster algorithm becomes advantageous for size $L > 5$.

However, below the critical temperature, where the correlation time is usually size independent, even our algorithm shows a strong size dependence. At a temperature $T/T_c = 0.68$, correlation-time data from system sizes $L = 4$, 8, and 16 are consistent with a dynamic critical exponent $z \approx 2$; it is the same as single-spin-flip dynamics at criticality. This peculiar behavior can be interpreted as that the low-temperature phase exhibits critical fluctuation as suggested by Ono.[10] Our susceptibility data support this interpretation.

B. Low-temperature behavior

Figure 7 is a plot of magnetization as a function of temperature in three dimensions for system sizes $L = 4$, 8, 16, and 32 (for earlier data, see Ref. 7). Unlike the ferromagnetic Ising models, we found a strong dependence of the order parameter on the size of the system. The magnetization approaches a nonzero value as $1/L$ below T_c. Figure 8 is a plot of magnetization as a function of inverse system size, $1/L$, at a temperature $T/T_c = 0.68$. The data linearly extrapolate to a nonzero finite-size limit $m(L = \infty) = 0.491$. Our result is in contrast to Ono's conclusion[10] that magnetization is zero with a massless phase below T_c, but is in agreement with that of Banavar *et al.*,[7] Hoppe and Hirst,[8] and Ueno *et al.*[12] As has been observed by Banavar *et al.*,[7] the magnetization at zero temperature, $m(T = 0) = 0.62$, is close to, but less than, $\frac{2}{3}$, the value yielded by maximal possible order. The maximal order (breaking sublattice symmetry) is obtained if one of the spin states is on sublattice a and the other two states are distributed randomly on sublattice b.

If the (truncated) correlation function decayed algebraically, then the susceptibility, defined by the fluctuation of the magnitude of a two-component vector order parameter,

$$\chi = L^{d}[\langle \xi_1^2 + \xi_2^2 \rangle - \langle (\xi_1^2 + \xi_2^2)^{1/2} \rangle^2]$$

(see the Appendix for the definition of ξ_1 and ξ_2), would diverge with system size. Figure 9 is a log-log plot of the

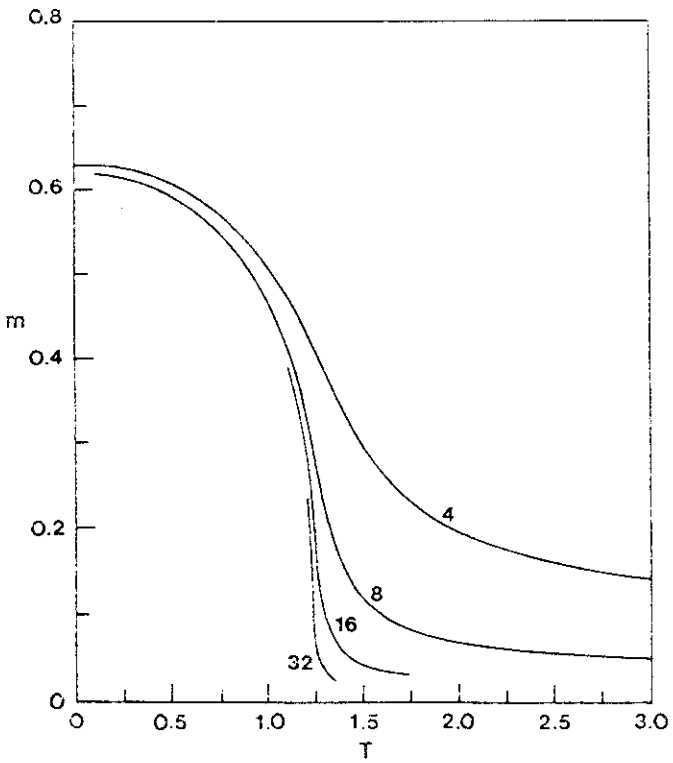


FIG. 7. Magnetization as a function of temperature in three dimensions for system sizes $L = 4$, 8, 16, and 32.

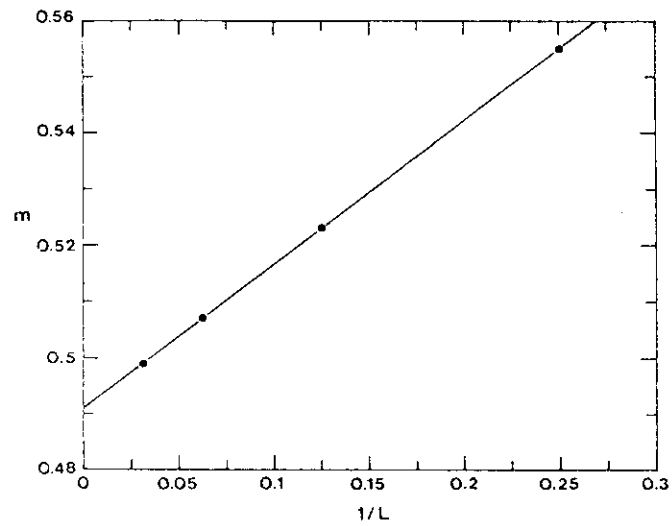


FIG. 8. Magnetization at $T/T_c = 0.68$ vs $1/L$ in three dimensions. The straight line is given by $m = 0.491 + 0.256/L$.

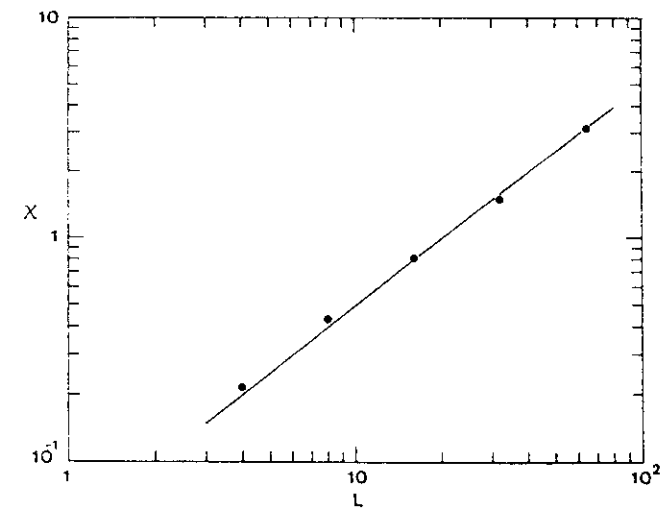


FIG. 9. Susceptibility in three dimensions at $T/T_c = 0.68$ as a function of system size L plotted in log-log scale. The slope of the straight line is 1.

susceptibility at $T/T_c=0.68$ against system size L. The susceptibility linearly depends on system size. This behavior is fully consistent with that of the XY model in three dimensions at low temperature. The $1/L$ dependence of the finite-size magnetization is also analogous to the XY model. A spin-wave calculation of the low-temperature magnetization and susceptibility of the XY model would give such a finite-size dependence.[45]

The distribution of the two-component order parameter yields some information about features of the low-temperature phase. At very low temperatures ($T=0.1$, for example), the order parameters take values close to full order. We see sixfold-symmetry peaks for $L=4$ and 8. As system size increases, the peaks become sharper. For $L \geq 16$, only three peaks remain; they are related by exchanging Potts states. The symmetry associated with exchange of sublattices is appearantly broken. As the temperature raises, the peaks becomes less sharp. For temperature not very far from T_c, the distribution looks more rotationally symmetric, with a sixfold anisotropy (or threefold for large system size). Very close to T_c, the distribution is nearly rotationally symmetric. It appears that the effect in Ono's simulation (Ref. 10, Figs. 5 and 6) is due to a finite observation time. If one simulates time long enough, or the dynamics are fast enough, one would see an approximate circle instead of an arc in the locus of the instantaneous order parameters.

The ground-state entropy is also calculated in three dimensions. Although we have only the values of the residual entropy for $L=4$ and 8 [$s(4)=0.3953$ (10 histogram) and $s(8)=0.3708$ (16 histograms)], we note that they agree with the equation $s(L) \approx s_\infty + L^{-d} \ln 6$, which would predict that $s_\infty = 0.3673$. (Hoppe and Hirst[8] obtained a higher value: 0.376.) This equation is also in reasonable agreement with the exact number of states for an $L=2$ lattice (113.3 versus 126). Borgs and Imbrie[46] rigorously proved that for a class of models with well-defined energetic barriers between phases, the similar coefficient is the logarithm of the number of phases. Even though their theory does not apply in our case, it makes plausible the speculation that ln6 should mean the existence of six phases at low temperatures.

C. Critical behavior

We used a single histogram method[43] to calculate the specific heat, the susceptibility, and the fourth-order cumulant, simulating at $T=1.22549$. The length of the runs were more than 10^6 Monte Carlo steps except the largest size, $L=64$, which was 1.8×10^5 Monte Carlo steps.

The peak of the specific heat grows as the system size increases. Figure 10 is a plot of the peak of specific heat versus size L in a log-log scale. The approach to the asymptotic behavior seems to be rather slow. Assuming that $C_{\text{peak}} \propto L^{\alpha/\nu}$, the effective exponent α/ν decreases from 0.3 to 0.16, while if one assumes a cusp, $C_{\text{peak}} = C_\infty + aL^{\alpha/\nu}$, the α/ν value is negative in the range $-0.2 - -0.1$. The data may also be compatible with a logarithmic divergence for $L>8$. A precise estimate for α/ν is not possible. This difficulty is also

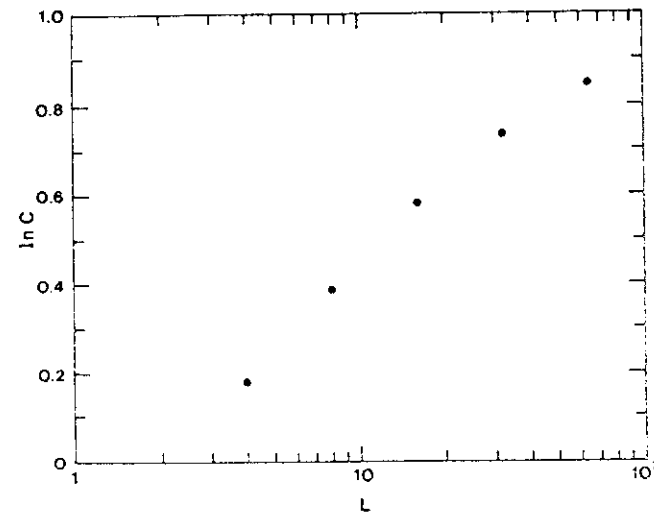


FIG. 10. The peak of specific heat in three dimensions vs size L in log-log scale.

reflected in the location of the peak, shown in Fig. 11. We found that the peak first moves toward lower temperatures as L increases, and then rises back slightly for larger systems. This makes it difficult to apply the standard finite-size scaling for the peak position,

$$T(L) \simeq T_c + bL^{-1/\nu} . \tag{12}$$

A more accurate result is obtained for the critical temperature from the fourth-order cumulant[47]

$$g(T,L) = \frac{1}{2}\left[3 - \frac{\langle m^4 \rangle}{\langle m^2 \rangle^2}\right] \simeq \bar{g}(|T-T_c|L^{1/\nu}) . \tag{13}$$

In the scaling region close to T_c, a different choice of L should have a unique intersection point at $T=T_c$. Figure 12 is $g(T,L)$ for $L=16$, 32, and 64. From this

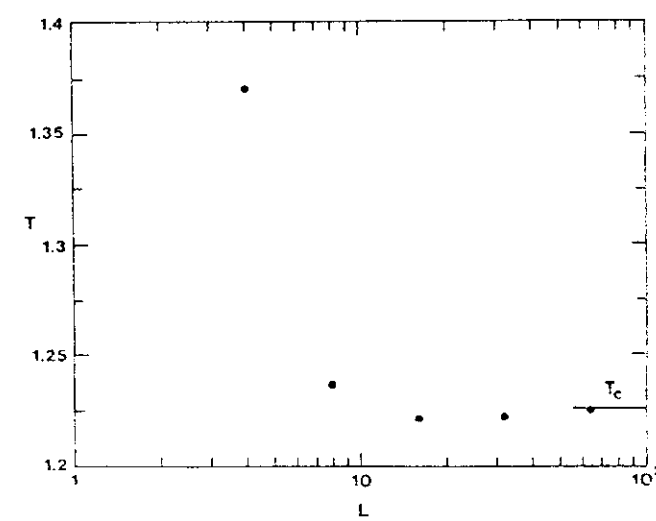


FIG. 11. The location of peak of specific heat in three dimensions as a function of size L.

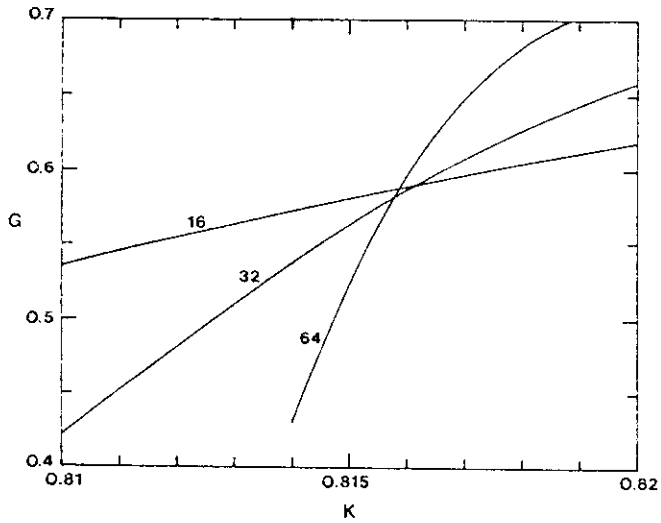


FIG. 12. Fourth-order cumulant g vs inverse temperature K for system size $L=16$, 32, and 64 in three dimensions.

analysis we get an estimate for $T_c=1.2259\pm0.0007$. This should be compared with the results of Ono[10] (1.25), Hoppe and Hirst[8] (1.28±0.04), and Ueno *et al.*[12] (1.235).

The critical exponents γ and ν are obtained from a finite-size scaling plot for the susceptibility, shown in Fig. 13. We find $\gamma/\nu=1.99\pm0.03$ and $\nu=0.66\pm0.03$, using $T_c=1.2259$. These exponents agree with those of the XY model[48–50] within errors. Our results are in disagreement with $\nu=0.58\pm0.01$, $\gamma=1.10\pm0.02$ obtained by Ueno *et al.*, using an interface approach.[12] The discrepancy may be due to the smaller system sizes ($L\leq24$) they have used and the location of the critical temperature.

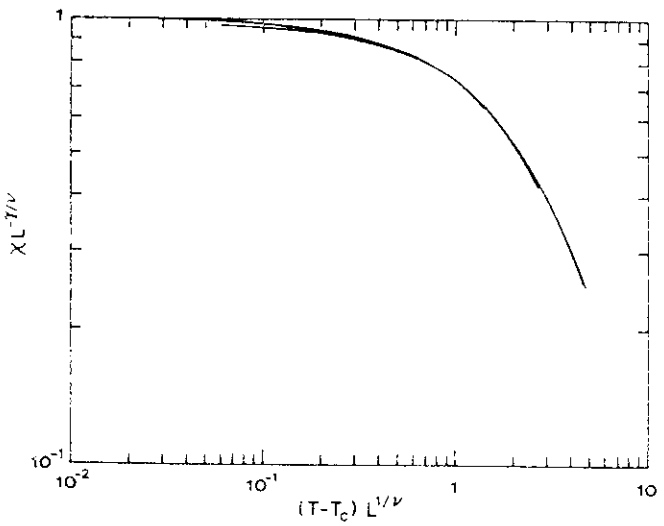


FIG. 13. Scaling plot for the three-dimensional three-state antiferromagnetic Potts model of the susceptibility times $L^{-\gamma/\nu}$ as a function of $(T-T_c)L^{1/\nu}$, using $T_c=1.2259$, $\gamma/\nu=1.99$, and $\nu=0.66$, for $L=8$, 16, 32, and 64.

VI. CONCLUSION

We have made a high-precision Monte Carlo simulation of the three-state antiferromagnetic Potts model in both two and three dimensions. This is only possible due to a fast algorithm for equilibration. In two dimensions, our results are consistent with previous conclusion that the transition is at zero temperature. The numerical estimates of the exponent η and the ground-state entropy are in good agreement with the exact results. The susceptibility is consistent with a form of an essential singularity. In three dimensions, we give estimates of the critical temperature and exponents. The exponents agree with XY model value within statistical errors. At low temperatures, the correlation time, magnetization, and susceptibility data indicate that the correlation length is infinite; but the system has a long-range order (nonzero order parameter). The question remains open whether the anisotropy at low temperatures is relevant to the critical behavior. While Ueno *et al.* claimed a new universality, our results are in favor of XY universality class for the model. Clearly more theoretical understanding is needed to clarify the issue.

ACKNOWLEDGMENTS

We would like to thank J.L. Lebowitz and the Department of Mathematics of Rutgers University for their generous hospitality while part of this work was being carried out. J.S.W. would like to thank B. Dünweg for a suggestion to consider order parameters in a systematic way that lead to the Appendix. He also thanks K. Binder, H. Müller-Krumbhaar, and W. Selke for stimulating discussions. J.S.W. and R.K. were supported in part by the Rutgers Supercomputer program and by National Science Foundation Grant No. DMR-8612369. R.H.S. was supported by the Rutgers Supercomputing program and by National Science Foundation Grant No. DMR-8613218. Some of the computations were performed at the John von Neumann National Supercomputer Center and at HLRZ, Jülich.

APPENDIX: ORDER PARAMETERS

Consider three atomic species 1, 2, and 3 (or Potts variables) to be placed on sublattice a and b. We shall study order parameters assuming that macroscopic states are completely characterized by particle concentrations c_1^a,c_2^a,c_3^a and c_1^b,c_2^b,c_3^b. They are defined by

$$c_\mu^l=\frac{2}{L^d}\left\langle\sum_{i\in l}\delta_{\sigma_i,\mu}\right\rangle , \tag{A1}$$

where δ is the Kronecker symbol, σ_i, $\mu=1,2,3$, and $l=a,b$. We use angular brackets to denote a thermodynamic average.

Since each site must be occupied by either 1, 2, or 3, we have a conservation law for the concentration:

$$c_1^a+c_2^a+c_3^a=1 , \tag{A2a}$$

$$c_1^b+c_2^b+c_3^b=1 . \tag{A2b}$$

Thus only four of the six parameters given are independent. Choosing c_1^a, c_2^a, c_1^b, and c_2^b as the independent ones, we may characterize the ferromagnetic order by (suitably rescaled) overall concentrations.

$$\Phi_1=\tfrac{3}{4}(c_1^a+c_1^b)-\tfrac{1}{2}\ , \tag{A3a}$$

$$\Phi_2=\tfrac{3}{4}(c_2^a+c_2^b)-\tfrac{1}{2}\ , \tag{A3b}$$

and the antiferromagnetic order by the difference of sublattice concentrations

$$\Psi_1=c_1^a-c_1^b\ , \tag{A4a}$$

$$\Psi_2=c_2^a-c_2^b\ . \tag{A4b}$$

Now, for a paramagnetic phase we have $\Phi_1=\Phi_2=\Psi_1=\Psi_2=0$, while a complete ferromagnetic order, say $\sigma_i=1$, has $\Phi_1=1$, $\Phi_2=-\frac{1}{2}$, $\Psi_1=0$, $\Psi_2=0$, and a complete antiferromagnetic order, say with 1 and 2 on sublattice a and b, respectively, has $\Psi_1=1$, $\Psi_2=-1$, $\Phi_1=\frac{1}{4}$, $\Phi_2=\frac{1}{4}$.

The dependent variables Φ_3, Ψ_3 can be defined in a similar fashion with the conservation law Eq. (A2) rewritten as

$$\Phi_1+\Phi_2+\Phi_3=0\ , \tag{A5a}$$

$$\Psi_1+\Psi_2+\Psi_3=0\ . \tag{A5b}$$

Let us consider invariants of the symmetries of the model (needed, e.g., when constructing a Landau Hamiltonian). Taking into account the symmetries with respect to arbitrary permutations of atomic species 1,2,3 and with respect to the interchange of the sublattices a and b, we infer that invariants are symmetric functions in the variables Φ_1, Φ_2, Φ_3 as well as Ψ_1, Ψ_2, Ψ_3, and are symmetric with respect to the transformation $\Psi\rightarrow-\Psi$. The second-order invariants are

$$\Phi_1^2+\Phi_2^2+\Phi_3^2\ , \tag{A6a}$$

$$\Psi_1^2+\Psi_2^2+\Psi_3^2\ . \tag{A6b}$$

The third-order term

$$\Phi_1^3+\Phi_2^3+\Phi_3^3 \tag{A7}$$

contains only ferromagnetic order parameters and, in the Landau Hamiltonian, may be viewed as that one responsible for a first-order phase transition in the ferromagnetic Potts model in three and higher dimensions. The fourth-order terms are

$$(\Phi_1^2+\Phi_2^2+\Phi_3^2)^2\ , \tag{A8a}$$

$$(\Psi_1^2+\Psi_2^2+\Psi_3^2)^2\ . \tag{A8b}$$

These are the only possible invariants up to the fourth order.

Let us choose independent parameters (cf. Ref. 18) that diagonalize the quadratic invariant, say $\Phi_1-\Phi_2$, Φ_3, $\Psi_1-\Psi_2$, and Ψ_3. Rescaling (with respect to shifting) them slightly we get the Landau Hamiltonian

$$H_\xi=a+b(\xi_1^2+\xi_2^2)+c(\xi_1^2+\xi_2^2)^2\ , \tag{A9}$$

$$H_\eta=d+e(\eta_1^2+\eta_2^2)+f\eta_1\left[\frac{\eta_1^2}{3}-\eta_2^2\right]+g(\eta_1^2+\eta_2^2)^2\ , \tag{A10}$$

with order parameters

$$\xi_1=-\frac{\sqrt{3}}{2}(c_3^a-c_3^b)\ , \tag{A11a}$$

$$\xi_2=\tfrac{1}{2}[(c_1^a-c_1^b)-(c_2^a-c_2^b)]\ , \tag{A11b}$$

and

$$\eta_1=-\tfrac{3}{4}(c_3^a+c_3^b)+\tfrac{1}{2}\ , \tag{A12a}$$

$$\eta_2=\frac{\sqrt{3}}{4}[(c_1^a+c_1^b)-(c_2^a+c_2^b)]\ . \tag{A12b}$$

These are the order parameters used by Ono[10] and have an illustrative interpretation when Potts spins σ_i are viewed as unit vectors in three directions forming the angles $2\pi/3$. Taking into account that the concentrations c_μ^l belong to the interval [0,1], the vector $\xi=(\xi_1,\xi_2)$ is restricted to fall within a uniform hexagon of diameter 1, while $\eta=(\eta_1,\eta_2)$ falls into a uniform triangle (cf. Ref. 10). In term of order parameters ξ and η, the three possible states of perfect ferromagnetic order are described by unit vectors η in the vertices of the triangle, while $\xi=0$. The states of perfect antiferromagnetic order are represented by unit vectors ξ in the vertices of the hexagon with vector η (of the length $\frac{1}{2}$) in the centers of the sides of the triangle. For the three-state Potts antiferromagnet in three dimensions a broken-sublattice (BS) order is expected to occur at low temperatures. Namely, the order with one sublattice is occupied by one of the three states, while the other sublattice is occupied by the remaining two states at random. The BS states are represented by vectors ξ of the length $\sqrt{3}/2$ in the centers of the sides of the hexagon with vector η of the length $\frac{1}{4}$ in the vertices of a suitably rescaled triangle.

The state constructed under periodic boundary conditions has all the symmetries of the model. In particular, all the concentrations c_μ^l are clearly equal to $\frac{1}{3}$ and the parameters Φ and Ψ are, strictly speaking, vanishing. In Sec. III we are using (following Ref. 7) the order parameter $\langle m\rangle$ defined by formulas (4) and (5). This parameter may be thought of as a sum of the parameter $|\Psi_1|+|\Psi_2|+|\Psi_3|$ over all ordered states (existing at given temperatures). Notice, however, that it cannot discern antiferromagnetic and BS states.

---

*Present address: Institut für Physik, Johannes-Gutenberg Universität, Staudinger Weg 7, D-6500 Mainz, Federal Republic of Germany.

[1]J.-S. Wang, R. H. Swendsen, and R. Kotecký, Phys. Rev. Lett. 63, 109 (1989).
[2]R. Baxter, Proc. R. Soc. London, Ser. A 383, 43 (1982).
[3]M. P. Nightingale and M. Schick, J. Phys. A 15, L39 (1982).
[4]C. Jayaprakash and J. Tobochnik, Phys. Rev. B 25, 4890

(1982).
[5]T. Temesvari, J. Phys. A **15**, L625 (1982).
[6]F. Fucito, J. Phys. A **16**, L541 (1983).
[7]J. R. Banavar, G. S. Grest, and D. Jasnow, Phys. Rev. Lett. **45**, 1424 (1980); Phys. Rev. B **25**, 4639 (1982).
[8]B. Hoppe and L. L. Hirst, J. Phys. A **18**, 3375 (1985); Phys. Rev. B **34**, 6589 (1986).
[9]R. Kotecký, Phys. Rev. B **31**, 3088 (1985).
[10]I. Ono, Prog. Theor. Phys. Suppl. **87**, 102 (1986).
[11]M. C. Marques, J. Phys. A **21**, 1061 (1988); **21**, 1297 (1988).
[12]Y. Ueno, G. Sun, and I. Ono, J. Phys. Soc. Jpn. **58**, 1162 (1989).
[13]H.-O. Georgii, *Gibbs Measures and Phase Transitions* (de Gruyter, Berlin, 1988), p. 148.
[14]G. S. Grest and J. R. Banavar, Phys. Rev. Lett. **46**, 1458 (1981).
[15]M. P. M. den Nijs, M. P. Nightingale, and M. Schick, Phys. Rev. B **26**, 2490 (1982).
[16]J. R. Banavar and F. Y. Wu, Phys. Rev. B **25**, 1511 (1984).
[17]I. Ono, J. Phys. Soc. Jpn. **53**, 4102 (1984).
[18]W. Kinzel, W. Selke, and F. Y. Wu, J. Phys. A **14**, L399 (1981).
[19]K. Yasumura, J. Phys. A **20**, 4975 (1987).
[20]Z. Racz and T. Vicsek, Phys. Rev. B **27**, 2992 (1983).
[21]A. N. Berker and L. Kadanoff, J. Phys. A **13**, L259 (1980).
[22]J. M. Kosterlitz and K. J. Thouless, J. Phys. C **6**, 1181 (1973).
[23]N. Metropolis, A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller, and E. Teller, J. Chem. Phys. **21**, 1087 (1953).
[24]K. Binder, *Monte Carlo Methods in Statistical Physics*, edited by K. Binder (Springer-Verlag, Berlin, 1979); *Applications of the Monte Carlo Method in Statistical Physics*, edited by K. Binder (Springer-Verlag, Berlin, 1984).
[25]P. C. Hohenberg and B. I. Halperin, Rev. Mod. Phys. **49**, 435 (1977).
[26]S. K. Ma, *Modern Theory of Critical Phenomena* (Benjamin, Reading, MA, 1976), Chaps. 11–14.
[27]S. Wansleben and D. P. Landau, J. Appl. Phys. **61**, 3968 (1987).
[28]R. H. Swendsen and J.-S. Wang, Phys. Rev. Lett. **58**, 86 (1987).
[29]D. Kandel, E. Domany, D. Ron, A. Brandt, and E. Loh, Jr., Phys. Rev. Lett. **60**, 1591 (1988); D. Kandel, E. Domany, and A. Brandt, Phys. Rev. B **40**, 330 (1988).
[30]R. G. Edwards and A. D. Sokal, Phys. Rev. D **38**, 2009 (1988).
[31]F. Niedermayer, Phys. Rev. Lett. **61**, 2026 (1988).
[32]U. Wolff, Phys. Rev. Lett. **62**, 361 (1989).
[33]R. C. Brower and P. Tamayo, Phys. Rev. Lett. **62**, 1087 (1989).
[34]D. W. Heermann and A. N. Burkitt, Physica A **162**, 210 (1990); J.-S. Wang, *ibid.* **164**, 240 (1990).
[35]P. Tamayo, R. C. Brower, and W. Klein, J. Stat. Phys. **58**, 1083 (1990).
[36]U. Wolff, Phys. Lett. B **228**, 379 (1989).
[37]R. G. Edwards and A. D. Sokal, Phys. Rev. D **40**, 1374 (1989).
[38]M. Hasenbusch, Nucl. Phys. **B333**, 581 (1990).
[39]R. B. Potts, Proc. Cambridge Philos. Soc. **48**, 106 (1962).
[40]J. Hoshen and R. Kopelman, Phys. Rev. B **14**, 3438 (1976).
[41]H. Park and M. Widom, Phys. Rev. Lett. **63**, 1193 (1989).
[42]A. M. Ferrenberg and R. H. Swendsen, Phys. Rev. Lett. **63**, 1195 (1989); Computers in Physics **3(5)**, 101 (1989); see also P. B. Bowen *et al.*, Phys. Rev. B **40**, 7439 (1989).
[43]A. M. Ferrenberg and R. H. Swendsen, Phys. Rev. Lett. **61**, 2635 (1988).
[44]E. Lieb, Phys. Rev. **162**, 162 (1967).
[45]K. Kawasaki and H. Mori, Prog. Theor. Phys. (Kyoto) **27**, 529 (1962); **28**, 690 (1962).
[46]C. Borgs and J. Imbrie, Commun. Math. Phys. **123**, 305 (1989).
[47]K. Binder, Z. Phys. B **43**, 119 (1981).
[48]M. Ferer, M. A. Moore, and M. Wortis, Phys. Rev. B **8**, 5205 (1973).
[49]J. C. Le Guillon and J. Zinn-Justin, Phys. Rev. Lett. **39**, 95 (1977).
[50]D. Z. Albert, Phys. Rev. B **25**, 4810 (1982).

Efficient Monte Carlo methods for the computer simulation of biological molecules

Djamal Bouzida*
Department of Physics, Carnegie Mellon University, Pittsburgh, Pennsylvania 15213

Shankar Kumar
Department of Biological Sciences, University of Pittsburgh, Pittsburgh, Pennsylvania 15260

Robert H. Swendsen
Department of Physics, Carnegie Mellon University, Pittsburgh, Pennsylvania 15213
(Received 29 August 1991)

We present an alternative approach to efficient Monte Carlo simulations of biological molecules. By relaxing the usual restriction to Markov processes, we are able to optimize performance while dealing directly with the inhomogeneity and anisotropy inherent in these systems. This approach allows us to sample configurational space more efficiently than with either standard Monte Carlo or molecular-dynamics methods.

PACS number(s): 87.15.−v, 02.50.+s, 31.15.+q

I. INTRODUCTION

In recent years, considerable effort and computational resources have been devoted to simulating biological molecules. These molecules, which include polymers, proteins, and nucleic acids, are of great importance in physics, chemistry, biology, and medicine [1,2]. The goal of these computer simulations is to provide insights into the structure-function relationships in biomolecular interactions and energetics [3] by supplying detailed information about the conformations and internal motions of biologically important molecules.

Due to the size and complexity of the task, these simulations require enormous amounts of supercomputer time. This has led us to investigate alternative methods for improving efficiency with the goal of reducing the burden on supercomputing resources, broadening the scope of applications, and increasing the reliability of the results.

Currently, most computer simulations of thermodynamic systems use either molecular-dynamics (MD) or Monte Carlo (MC) methods. Both methods involve the generation of molecular conformations to represent thermal fluctuations. Equilibrium properties are found by computing appropriate averages over the resulting set of conformations. The interactions between atoms in the molecule are represented by an effective Hamiltonian that has been constructed empirically from a variety of experimental information [4–8].

Molecular dynamics is a deterministic method that computes classical trajectories by iterating a discretized representation of Newton's equations. The total energy of the system is conserved so that the generated configurations trace out a microcanonical ensemble (if the system is ergodic). An advantage of the MD method is the possibility of following an explicit classical trajectory of the system.

Since it is often more convenient to analyze data obtained at constant temperature (and/or pressure), various modifications of the MD method have been developed. One possibility is to introduce terms representing noise and dissipation into the equations, which leads to a canonical ensemble (constant temperature) [9–11]. Another possibility is the rescaling of atomic velocities to impose a fixed average temperature [11,12]. Nosé has introduced particularly interesting modified equations of motion that couple the system to a fictitious external degree of freedom [13,14].

The main adjustable parameter that enters a molecular-dynamics simulation is the size of the time step. Increasing the time step moves the system through phase space more rapidly, but it can introduce errors and can even affect the stability of the algorithm. For either Gear [15] or Verlet [16] algorithms, the maximum time step for reasonable accuracy must be less than about $\frac{1}{15}$ of the shortest period of vibration. In the case of most biological molecules, this corresponds to roughly 1 fs, which is indeed commonly used in such work. It might also be noted in passing that since the highest frequencies are associated with the smallest masses, one way to increase the efficiency of the MD method without distorting the equilibrium properties would be to set all masses equal [17].

A Monte Carlo simulation is a stochastic Markov process that generates a sequence of configurations representing a canonical ensemble. Trial moves are generated from a random distribution and are either accepted or rejected with a probability given by the Boltzmann factor. The Markov property of the MC process means that the probability of transition to a new state depends only on the present state. A fundamental theorem of Markov processes states that if the transition probabilities satisfy detailed balance, and if any configuration can be reached from any other configuration in a finite number of steps with nonzero probability (ergodicity), then the simulation can correctly reproduce the equilibrium behavior [18].

There are also mixtures of MC and MD methods known as hybrid MC [19] and hybrid MD methods [20]. In these methods, larger time steps are used with a global acceptance step to ensure thermal equilibrium.

In developing alternative methods, we must keep in mind the nature of the system and the types of questions that are asked. Some of these questions concern the dynamics of short-time behavior (i.e., a few picoseconds), and the only current option is the MD method. However, most questions concern equilibrium properties. The equilibrium configuration is of major importance in determining biological function, and free-energy calculations are needed to predict and understand biochemical reactions. Many calculations of relaxation times actually require quasiequilibrium determination of free-energy barriers.

Early work by Northrup and McCammon [21] indicated that the MD method was more efficient than the MC method, even for equilibrium properties. This conclusion was based on their observation that the MC method required much more computer time to carry out the updating process. However, the two methods actually require about the same amount of computer time per sweep. When this is corrected for, their data, based on the rms deviations per sweep, indicate that the standard MC method is more than the MD method. We have chosen to base our approach on the MC method partly for this reason, but primarily because of the flexibility for introducing additional new moves in MC simulations to improve the efficiency.

Both Monte Carlo and molecular-dynamics methods were originally developed to simulate fluids. In setting up a simulation, both methods require preliminary calculations to equilibrate the system and to determine the appropriate simulation parameters (time step in the MD method or maximum jump size in the MC method) and to ensure stability in the case of the MD method. When simulating fluids, optimization of parameters is done on a global basis, which is appropriate because these systems are homogeneous and isotropic. However, the local structure of molecules is inhomogeneous and highly anisotropic [1,2,21]. Thus it is necessary to optimize the parameters locally in order to increase the efficiency of the simulation. Furthermore, both the inhomogeneity and anisotropy change with time.

We have developed alternative MC methods to address these problems. The essential feature of our approach is the optimization of the Monte Carlo parameters from data collected during the simulation. Because this feature allows us to optimize local moves, we are able to deal explicitly with the inhomogeneity and the local anisotropy of biological molecules. Our approach also allows for the optimization of a wide variety of global moves without long preliminary studies. This turns out to be extremely important in designing efficient algorithms.

Since our methods use information gathered during the simulation, they are no longer strictly Markovian, although they are *almost* or *piecewise* Markov processes [22]. Detailed balance is satisfied and we have been able to demonstrate that our algorithms reproduce the correct equilibrium behavior while substantially improving speed and efficiency. In this paper, we present a description of our approach and illustrate its advantages by applying it to a small molecule.

II. ALMOST MARKOV SIMULATION METHODS

In this section, we derive two simulation methods from an analysis of a simple harmonic oscillator (SHO). The acceptance-ratio method (ARM) is an optimization technique that treats the inhomogeneity of biomolecules efficiently. The second method, the dynamically optimized Monte Carlo (DOMC) method, treats both the inhomogeneity and the anisotropy. These methods are then shown to retain their efficiency for the anharmonic potentials found in biological molecules.

The potential energy of a $d=1$ SHO is given by $V(x)=\frac{1}{2}kx^2$. The efficiency of a MC simulation is determined by the choice of maximum step size δ from which the MC trial moves are generated. If an optimum δ is known for some given k and β, then the optimum δ for any other values of k and β is also known through the scaling relation

$$\tfrac{1}{2}k\beta\delta^2=F^2 , \tag{1}$$

where $\beta=1/k_BT$. Thus, if we can determine the optimal scale factor F for any SHO, we have solved the problem for all such models.

An important quantity for understanding the efficiency of MC simulations is the average acceptance ratio $\langle P\rangle$, which is defined as the ratio of accepted moves to trial moves during a simulation. As shown in Fig. 1 (which

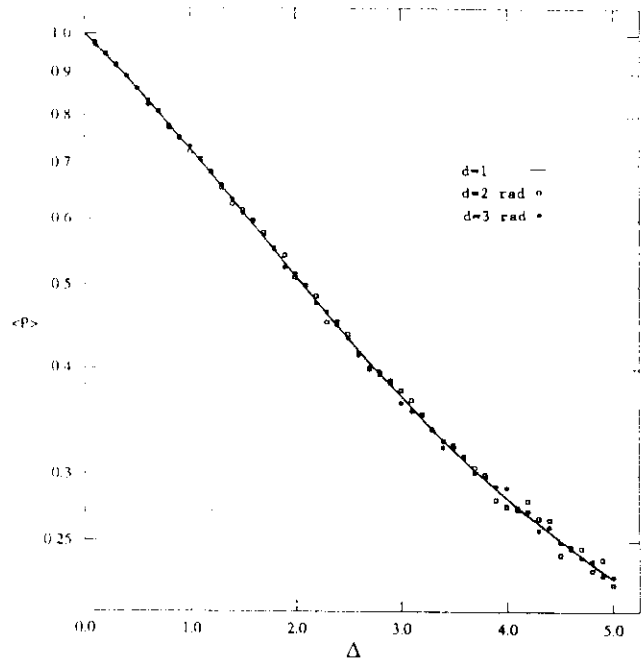


FIG. 1. Semilog plot for the average acceptance ratio $\langle P\rangle$ as a function of the maximum step size for $d=1$, 2, and 3 SHO with $\beta=1$. The one-dimensional curve is exact, while in higher dimensions, each point is obtained by averaging over 10 000 MC steps from a simulation where the trial jumps are generated uniformly in radius.

also shows data for higher dimensions discussed below), the acceptance ratio decreases monotonically as a function of step size δ. This is expected because small trial step sizes corresponding to small energy changes will produce high acceptance ratios, while large moves have a high probability of being rejected due to large energy differences. The acceptance ratio decreases approximately exponentially as a function of δ for the range of values shown in Fig. 1. For larger δ, the acceptance ratio varies inversely as δ.

To determine the optimal value of F, we first performed a series of MC studies of a $d=1$ SHO, $V(x)=x^2$ with $\beta=1$. In addition to the acceptance ratio, we monitored the autocorrelation time τ and two measures of the displacements per MC step, $\langle(\Delta x)^2\rangle^{1/2}$ and $\langle|\Delta x|\rangle$, where Δx represents the displacement during a MC move and the angular brackets indicate the usual thermal average.

The simplest measures of efficiency are the rms [21] and average absolute displacements. These quantities should go to zero for small acceptance ratio $\langle P\rangle$ since most moves are rejected, and for $\langle P\rangle$ near 1.0 since each trial move is small. Therefore, we expect a maximum in each curve, as seen in Fig. 2. The maximum rms displacement $\langle(\Delta x)^2\rangle^{1/2}$ occurs at $\langle P\rangle=0.42$ ($\delta=F=2.62$), while $\langle|\Delta x|\rangle$ has a maximum at the larger value of $\langle P\rangle=0.56$ ($\delta=F=1.76$). The optimal value of F clearly depends on what is being calculated. However, even if $\langle P\rangle$ differs from the optimal value by as much as ±0.15, the displacements are only reduced by 10%. This leaves a fairly large region around $\langle P\rangle=0.5$ ($\delta=F=2$) for which both quantities are nearly optimized.

The integrated correlation time τ was determined from the normalized time-dependent energy-energy correlation function

$$f(t)=\frac{\langle E(t_0)E(t_0+t)\rangle-\langle E\rangle^2}{\langle E^2\rangle-\langle E\rangle^2} \tag{2}$$

by the usual expression

$$\tau=\sum_{t=1} f(t_i)\ , \tag{3}$$

where the sum is cut off when the fluctuations drive the correlation function negative [23]. The statistical error is proportional to $\sqrt{1+2\tau}$.

Figure 3 shows a plot of the correlation time τ as a function of the acceptance ratio $\langle P\rangle$. This plot shows a minimum in the correlation time with $\tau_{\min}\simeq1.4$ MC steps corresponding to a 50% acceptance ratio. This provides a justification for the common practice of tuning the step size to accept about one-half of the trial moves. In fact, using $\langle P\rangle=0.5$ ($\delta=F=2$) to minimize τ also gives rms and average absolute displacements within 2.5% and 1.3% of their respective maxima.

In higher dimensions, the parameters that enter the optimization scheme are found to depend on how the jumps are generated. Two possibilities are to generate them uniformly in either the radius or the volume of a sphere (uniformly in radius or area of a circle for $d=2$). Interestingly, when the jumps are generated uniformly in radius, the acceptance ratio, the rms displacement, and the absolute displacement are nearly independent of the dimension as

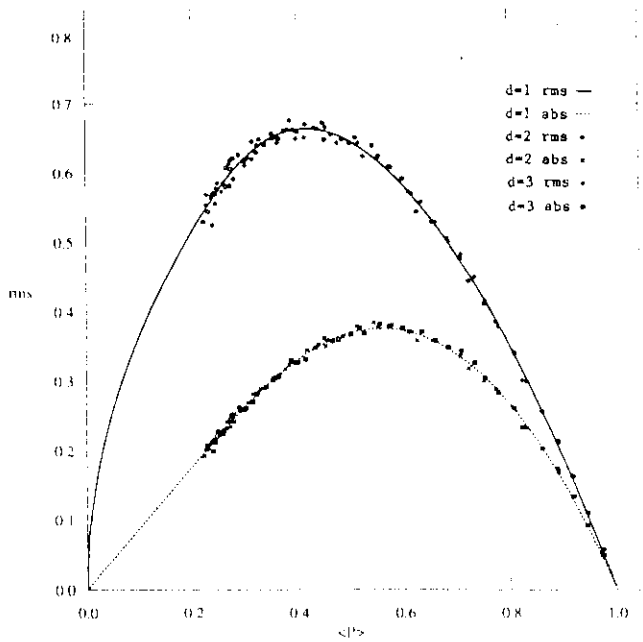


FIG. 2. The rms and absolute displacements vs acceptance ratio for $d=1$, 2, and 3 SHO with $\beta=1$. The one-dimensional curve is exact, while in higher dimensions, each point is obtained by averaging over 100 000 MC steps from a simulation where the trial jumps are generated uniformly in radius.

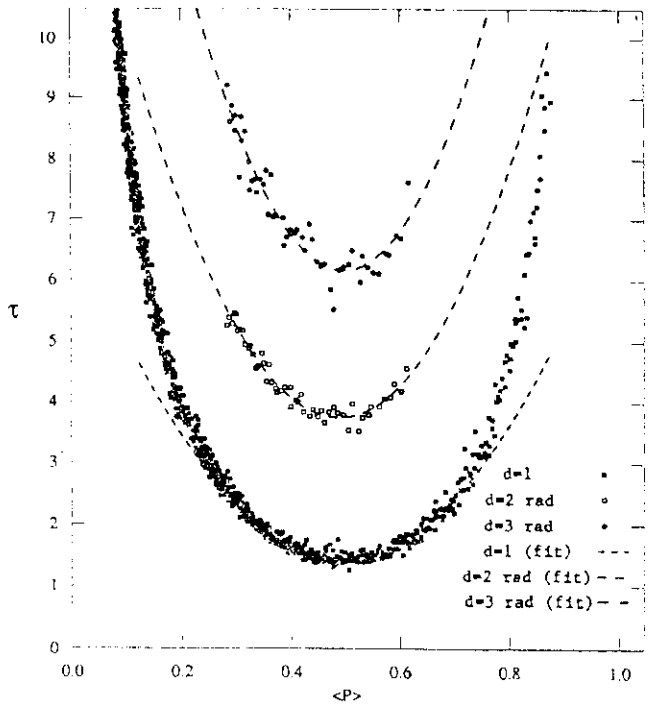


FIG. 3. Plot of the energy-energy autocorrelation time τ as a function of acceptance ratio $\langle P\rangle$ for $d=1$, 2, and 3 SHO with $\beta=1$. The trial jumps are generated uniformly in radius, and the dashed lines represent parabolic fits to the curves.

own in Figs. 1 and 2. The correlation time still has a inimum at $\langle P \rangle = 0.5$, but its value increases for higher imensions as shown in Fig. 3. The minimum correlation me is 3.8 MC steps in two dimensions, and 6.1 MC steps three dimensions. Thus the optimal value of F when ie jumps are generated uniformly in radius is still $F = 2$ rresponding to a $\langle P \rangle = 0.5$.

When the trial jumps are generated uniformly in olume, the behavior is somewhat different. As shown in ig. 4, the acceptance ratio decreases more rapidly as a nction of δ than in the previous case, but is still a near- exponential function. The maximum rms and absolute splacements are higher than in the previous case but cur at lower acceptance ratios as shown in Fig. 5. The inimum correlation time occurs at lower acceptance ra- os for higher dimensions as shown in Fig. 6. The inimum correlation time is 2.8 MC steps corresponding $\langle P \rangle = 0.42$ in two dimensions, and it is 4.4 MC steps curring at $\langle P \rangle = 0.39$ in three dimensions. The value F corresponding to the optimal correlation time de- eases only slightly in higher dimensions, being about 9 for $d = 2$, and about 1.8 for $d = 3$ SHO.

Although most workers fix the step size for the dura- on of the simulation, some efforts have been made to odate it as new information is generated. Allen and idesley suggested raising or lowering the global step ze by 5% depending on whether the measured accep- nce ratio is above or below 50% [24]. Corana *et al.* in- oduced variations in maximum step sizes to maintain e acceptance ratio at 50% in simulated annealing runs r minimizing functions of continuous variables [25]. nce they were not concerned with equilibrium proper- es, they did not discuss the non-Markovian nature of eir procedure.

Our first optimization procedure is an acceptance-ratio

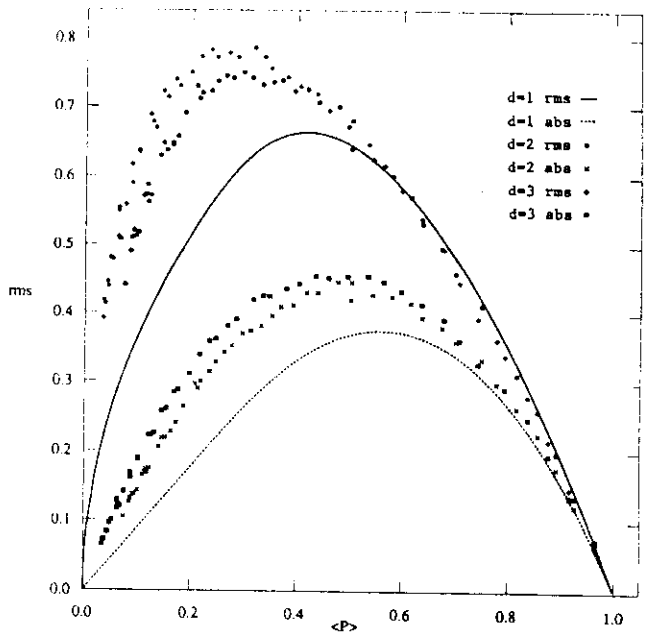


FIG. 5. Same as in Fig. 2, except the trial jumps are generated uniformly in volume (uniformly in area for $d = 2$).

method to carry out equilibrium simulations with different dynamically determined step sizes for each particle. We have used the approximately exponential dependence of $\langle P \rangle$ on δ

$$\langle P \rangle = \exp(-\delta/\delta_0) , \tag{4}$$

where δ_0 is some constant. Let $\langle P_i \rangle$ be the ideal or the

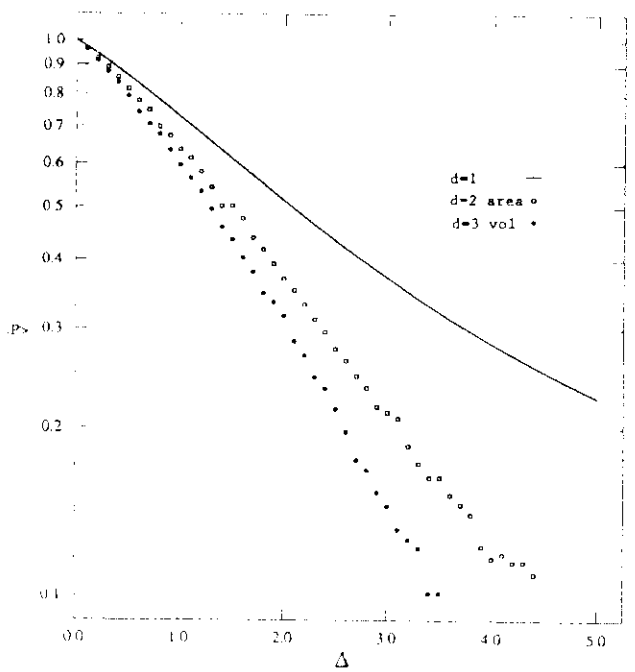


FIG. 4. Same as in Fig. 1, except the trial jumps are generated uniformly in volume (uniformly in area for $d = 2$).

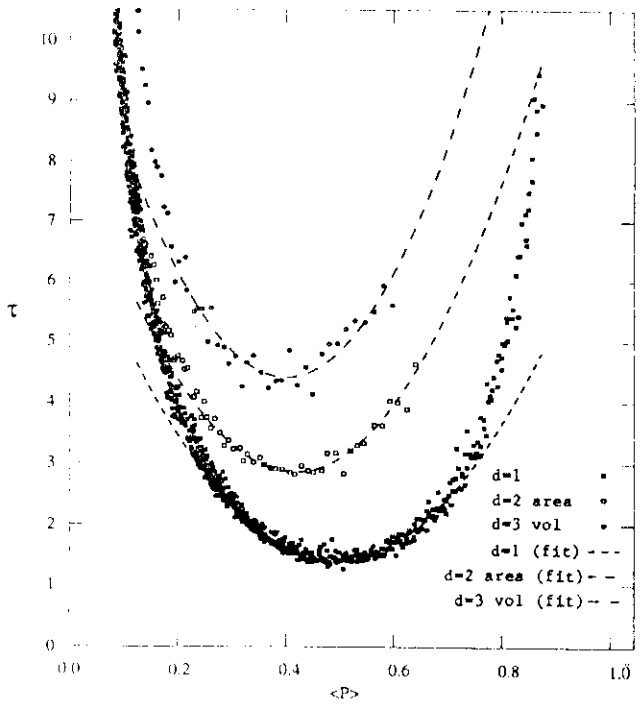


FIG. 6. Same as in Fig. 3, except the trial jumps are generated uniformly in volume (uniformly in area for $d = 2$).

desired acceptance probability corresponding to an ideal maximum step size δ_i. Then, clearly

$$\delta_i = \delta \frac{\ln\langle P_i \rangle}{\ln\langle P \rangle} . \tag{5}$$

A simulation is set up as a sequence of cycles. During a given cycle characterized by a maximum step size δ_{old}, the acceptance ratio $\langle P_{\text{old}} \rangle$ for each particle is computed. An iteration procedure using Eq. (5) is set to update new values of δ. However, this equation must be protected against overflow problems that occur whenever P is either 0 or 1. Therefore, we have modified it to read

$$\delta_{\text{new}} = \delta_{\text{old}} \frac{\ln(a\langle P_i \rangle + b)}{\ln(a\langle P \rangle + b)} , \tag{6}$$

where a and b are real parameters chosen such that δ_{old} is either multiplied or divided by a convenient scale factor (about 5 or 10) whenever P is 0 or 1 [25].

The ARM is a robust optimization technique and is especially useful at high temperatures for simulated annealing experiments. An advantage of the method is that each atom is treated separately, thus dealing with the inhomogeneity of the system efficiently. A weakness is that the accuracy of the optimization is limited by the discrete estimates of the acceptance ratio from the finite length of each simulation cycle. Another weakness is that the ARM does not deal with the local anisotropy of macromolecules, although it can be applied effectively to rotations of part of the molecule or other global moves. However, these problems can be dealt with using a different method we call the dynamically optimized Monte Carlo method.

We first discuss the one-dimensional DOMC equations for a $d=1$ SHO. Square brackets denote a direct average over all attempted moves regardless of whether they are accepted or not. It is easy to show that

$$[\Delta E] = \tfrac{1}{2} k [(\Delta x)^2] , \tag{7}$$

where ΔE represents the energy change corresponding to the jump size Δx. Combining this with Eq. (1) and eliminating k, we obtain the DOMC estimate for the optimum value of δ from the simulation data

$$\delta = F \left[\frac{[(\Delta x)^2]}{\beta[\Delta E]} \right]^{1/2} , \tag{8}$$

or, more generally

$$\delta = F \left[\frac{[(\Delta x)^{2n+2}]}{\beta[\Delta E (\Delta x)^{2n}]} \right]^{1/2} , \quad n = 0, 1, 2, \ldots . \tag{9}$$

This is the fundamental DOMC equation for $d=1$ systems. As discussed above, the scale factor F for the SHO is about 2 for optimum efficiency. The extension of the DOMC method to anisotropic systems in two and three dimensions is described in the Appendix. Trial moves are made within an ellipsoid (an ellipse for $d=2$) that reflects the local anisotropy. The advantage of choosing moves from an ellipsoid was suggested by Northrup and McCammon [21] in 1980, although they had not developed a method for implementing it.

During each simulation cycle, the averages $[(\Delta x$ and $[\Delta E]$ are computed locally for each variable, and new value of the maximum step size is obtained for t next cycle [26]. This procedure is done only once eve cycle, and takes a negligible amount of computer time- about 3% for the adenosine simulations discussed belo and even less for larger molecules.

Because both the ARM and the DOMC method use i formation from past configurations in determining t transition probabilities, they are not strictly Markovia which raises the possibility that systematic errors mig be generated. To test for systematic errors, we have us the DOMC method to calculate the energy of a SHO v ing very short cycles. In the extreme case of only 1 M step/cycle, we do find a large systematic error of 42% However, with even 2 MC steps/cycle the error drops 6%, and for 3 MC steps/cycle it is about 1%. No sy tematic error was measurable for 4 or more M steps/cycle. To obtain small statistical errors in the es mates of δ, at least 10 MC steps/cycle are needed f one-dimensional moves, and 50 to 100 MC steps/cyc for three-dimensional moves. Consequently, we conclu that the systematic errors are negligible for practical a plications.

We found DOMC to be extremely effective for a wi range of anharmonic systems including both symmetr and asymmetric double-well potentials. The results we qualitatively similar to those for the SHO, although t optimal values of F tended to be higher—about 3 or 4. was also interesting to note that the average displac ments showed a broader maximum when plotted again F than when plotted against δ, which implies that t precise value of F is even less critical for strongly anha monic systems. Comparisons between the results DOMC and direct numerical integrations aga confirmed the absence of any measurable systematic err for a cycle length of more than 4 MC steps. Furthe more, DOMC easily achieves optimal efficiency for tw and three-dimensional models, with anisotropies of 10 in the ratio of the coupling constants. Tests on two- an three-dimensional potentials with anharmonicity an even double minima have demonstrated that the DOM equations remain remarkably efficient.

In practice, it is possible for the straight averages use in Eq. (9) to take either very large or negative value This may arise, for example, when a trial move puts a atom very close to another. Generally, this is rather rar at normal or low temperatures. However, to account fo such cases, the stability of the program is protected b providing a branch to an alternate updating of the ste size based on the ARM for the particular cycle that ha run into a problem. The program reverts to the DOM equations on the following cycle.

Although the ARM and the DOMC method have bee derived and discussed in terms of single-particle moves this is not a real restriction. In fact, collective moves tha are important in biomolecules may be easily optimize using this approach. A particularly important class o such moves involves global rotations of a group of atom with respect to the rest of the molecule [27–29]. Thes

oves are described by angles of rotation, and the only odification needed is to restrict the maximum trial ove to the range $[-\pi,\pi]$. The DOMC method can be pecially useful when the rotations are highly correlated. ficient MC moves involving groups of two or three ro- ions can be optimized with the equations derived in e Appendix.

III. APPLICATIONS

As a simple example, we have applied these simulation ethods to adenosine, one of the building blocks of NA, using the united-atom force field developed by ollman and co-workers [8]. The potential energy is ven by

$$E=\sum_{\text{bonds}} K_b(b-b_{\text{eq}})^2+\sum_{\text{angles}} K_\theta(\theta-\theta_{\text{eq}})^2 + \sum_{\text{dihedral}} \frac{V_\varphi}{2}[1+\cos(n\varphi-\gamma)] + \sum_{\text{nonbonds}} \left[\frac{A_{ij}}{R_{ij}^{12}}-\frac{B_{ij}}{R_{ij}^{6}}\right]+\sum_{\text{H bonds}} \left[\frac{C_{ij}}{R_{ij}^{12}}-\frac{D_{ij}}{R_{ij}^{10}}\right] + \sum_{\substack{i,j\\ i<j}} \frac{q_i q_j}{\epsilon R_{ij}} . \quad (10)$$

ne first two terms represent the bond-length and bond- ngle strain energies. The dihedral energy is represented / a cosine function, where φ is the dihedral angle, n enotes the symmetry of the torsional barrier, V_φ is the rce constant, and γ is the phase angle. The last three rms describe the "nonbonded" interactions between airs of atoms that do not belong to the same bond or an- e. A hydrogen bond is represented by a 10-12 potential, nd a Lennard-Jones potential is used for the other non- onded pair interactions. The last term is intended to epresent both the direct electrostatic interaction and the creening effects of solvent. This term is not well deter- ined for typical separations of a few angstroms, but we ave followed the common practice of taking the dielec- ic "constant" to be $\epsilon(R_{ij})=R_{ij}$ [30].

In performing the simulations, trial MC moves includ- d nine global rotations along with single-particle jumps. he possibility of including such global moves to ac- elerate the simulation is a great advantage of the Monte arlo approach.

One measure of the DOMC efficiency is the energy- nergy correlation time, which is 8.6 sweeps, indicating a apid sampling of phase space. However, this quantity is ot easily compared with the results of other methods for easons that will become clear below. Therefore, we have sed the time dependence of the approach of the average ms displacement $\langle(\Delta\mathbf{r})^2\rangle^{1/2}$ to its equilibrium value as a easure of efficiency [21]. This quantity is defined as

$$\langle(\Delta\mathbf{r})^2\rangle^{1/2}=\left\langle \frac{1}{N}\sum_{i=1}^{N}\{\mathbf{r}_i-\langle\mathbf{r}_i\rangle\}^2\right\rangle^{1/2}, \quad (11)$$

here N is the number of atoms and $\mathbf{r}_i$ and $\langle\mathbf{r}_i\rangle$ are the osition and the average position of atom i, respectively. It was computed by superposing all the stored structures to the initial equilibrated structure. The superposition was done following the procedure introduced by Kabsch [31] with a modification that assures that the matching of structures is done through a rotation and not an inversion [32].

We have performed conventional MC, MD, and DOMC simulations at room temperature (298 K) using the same amount of computer time to account for the fact that the full DOMC simulation took a factor of 3 more computer time per sweep due to the global rotations. The DOMC simulation consisted of 1800 cycles with 100 sweeps each. Every single-particle move and rotation was performed once each sweep. Snapshots of the molecule are taken every 10 sweeps. The computed average position of each atom $\langle\mathbf{r}_i\rangle$ after superposition is updated after each snapshot. The rms is found to converge to about 1 Å.

In the MD simulation, a Verlet leap-frog algorithm was used with a time step of 10^{-15} s, starting from the same equilibrated initial configuration as the DOMC simulation. We used the average potential energy from the DOMC simulation to initialize the velocities of the atoms. Velocities were initially generated from a Maxwellian distribution and a global adjustment was made to set the total linear and angular momenta to zero. The velocities were then rescaled to make the initial kinetic energy equal to the difference between the previously calculated total energy and the configuration's poten-

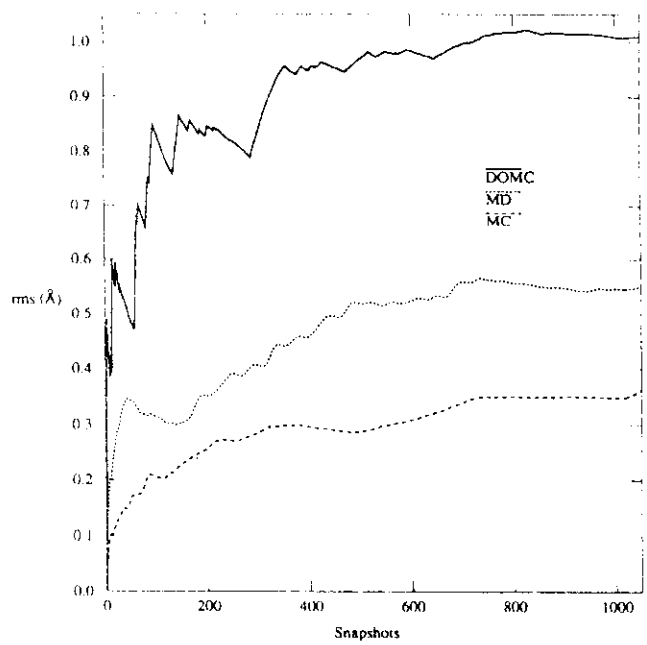


FIG. 7. The rms displacements for simulations of adenosine obtained from MC, MD, and DOMC simulations corresponding to the same amount of computer time. On a DEC 3100 workstation, the CPU times are 0.12 s/sweep for the DOMC simulation, and 0.04 s/sweep for the MC and MD simulations. For the DOMC simulation, snapshots are taken every 10 sweeps, while snapshots are taken every 30 sweeps for the MC and MD simulations. In this figure, the rms obtained during the first 1000 snapshots is shown.

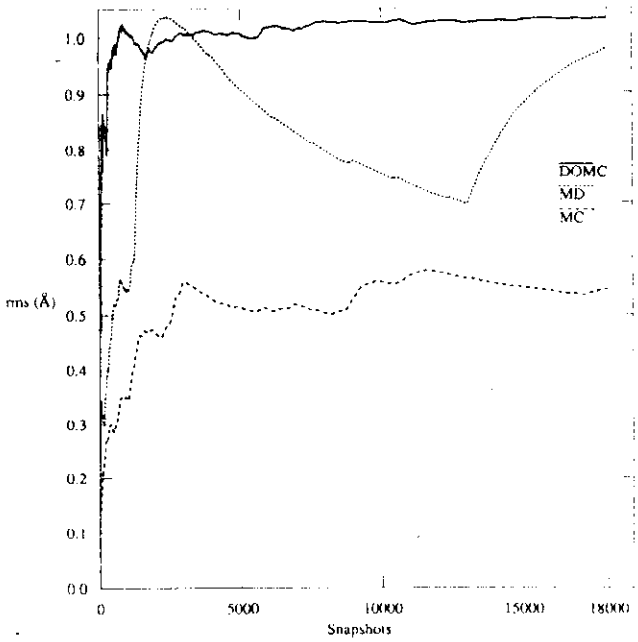


FIG. 8. Same as in Fig. 7, except that the length of the runs are longer. The DOMC simulation is 180 000 sweeps, while the MC and MD simulations are 540 000 sweeps long.

tial energy.

The standard MC simulation was done with a global optimizaton. Single atom moves were generated uniformly in volume from spherical neighborhoods with a radius of 0.08 Å, corresponding to the maximum rms displacement for this method. The overall global acceptance ratio was 31%.

Plots of rms displacements obtained for each of the three methods starting from the same well-equilibrated structure corresponding to the same amount of computer time are shown in Fig. 7 and 8. Figure 7 shows the "short-time" behavior of the rms fluctuations, while a lengthier run is shown in Fig. 8. From the short run, the rms fluctuations obtained from the DOMC simulation exceed the MD fluctuations, and approach the asymptotic value rapidly.

The plot of the DOMC rms fluctuations also shows some kinks (or zigzags). These represent energy-barrier crossings when the molecule changes conformation states. These kinks are also seen in the MD data in Fig. 8; however, they occur on much longer (and unpredictable) time scales.

The rms fluctuations obtained from MD depend strongly on the initial conditions—both on the initial configuration and the initial random Maxwellian velocities. They do not usually rise as fast as shown in Fig. 7. Our work indicates that for any initial configuration, the DOMC simulation converges much faster than either standard MC or MD simulation.

IV. CONCLUSION

In this paper, we have introduced two methods for carrying out optimized Monte Carlo simulations of thermodynamic systems with strong inhomogeneity and local anisotropy. This approach is particularly intended for simulations of macromolecules, although we expect it to be useful in other situations. These methods make essential use of information gathered during the course of the simulation, which requires a slight relaxation of the usual restriction of Monte Carlo simulations to Markov processes. Our calculations have shown that under normal conditions, all systematic errors introduced by the non-Markovian nature of the simulation are negligible.

An important advantage of the current approach is the automatic optimization of any kind of MC move that would be useful in accelerating the convergence of the simulations. Large-scale collective motions can be emphasized, simulations can be carried out in either internal coordinate space or Cartesian space, or a mixture of both. This has far-reaching implications, especially in the calculation of free-energy differences by free-energy perturbation, multistage sampling, or umbrella sampling techniques, where lack of proper convergence can make the simulations very long and time consuming [33].

By providing for each incorporation of NOE and/or crystallographic constraints, the ARM and the DOMC method can be used for efficient structure refinement using simulated annealing techniques in either NMR or crystallographic studies. Such applications of the ARM and the DOMC method are already in progress.

We have demonstrated the efficiency of these methods by applying them to a small molecule that exhibits many of the characteristics that make simulations of larger molecules difficult. However, the structure of proteins presents certain problems that require specialized techniques beyond the scope of the present paper. Simulations of larger molecules, including progress on simulated annealing to find protein conformations, are planned to be discussed elsewhere.

ACKNOWLEDGMENTS

We would like to acknowledge support by the National Science Foundation Grant No. DMR-9009475 (D.B. and R.H.S.) and NIH Grant No. GM25671 and the National Science Foundation Grant No. ASC-9015310 (S.K.). We also thank Charles Brooks, Stephen Garoff, John Nagle, and John Rosenberg for helpful comments and valuable discussions.

APPENDIX: $d=2$ AND $d=3$ DOMC METHOD

To extend DOMC equations to higher dimensions, we first consider an effective anisotropic simple harmonic oscillator of the form

$$\mathcal{H}=\tfrac{1}{2}\sum_{i,j} k_{ij}x_i x_j \ ,$$

where k_{ij} represents the spring constants. We use a transformation matrix D to generate moves $\{\eta_i\}$ in an ellipsoid (or an ellipse for $d=2$) given by

$$\eta_i=\sum_{j=1}^{3} D_{ij}\xi_j \ .$$

$\{\xi_j\}$ is a random vector chosen from a unit sphere (unit circle for $d=2$). The optimum jump size scales as the contours of constant energy and is determined by a dimensionless parameter F given by

$$F^2=\tfrac{1}{2}\beta\mathbf{D}^t\cdot\mathbf{k}\cdot\mathbf{D} ,$$

where $\mathbf{D}^t$ is the transpose of the transformation matrix $\mathbf{D}$ and $\beta=1/k_BT$. F is a scale factor chosen to optimize the efficiency of the simulation. The matrix $\mathbf{k}$ is determined from the simulation using

$$[\Delta E\eta_l\eta_m]=\tfrac{1}{2}\sum_{i,j}k_{ij}[\eta_i\eta_j\eta_l\eta_m] .$$

In this linear system of equations, ΔE denotes the change of energy for an attempted move $\{\eta_i\}$ and the square brackets indicate an average over all attempted moves, whether or not they were accepted [26]. The matrix $\mathbf{D}$ is then obtained from

$$D_{in}=F\left[\frac{2}{\beta\lambda_n}\right]^{1/2}V_{in} ,$$

where λ_n is an eigenvalue of $\mathbf{k}$ and $\mathbf{V}$ is the corresponding normalized eigenvector. $\mathbf{D}$ is then updated every cycle to adapt to the changing local environment of each atom.

---

*Present address: Department of Biomedical Engineering, Boston University, Boston, MA 02215.

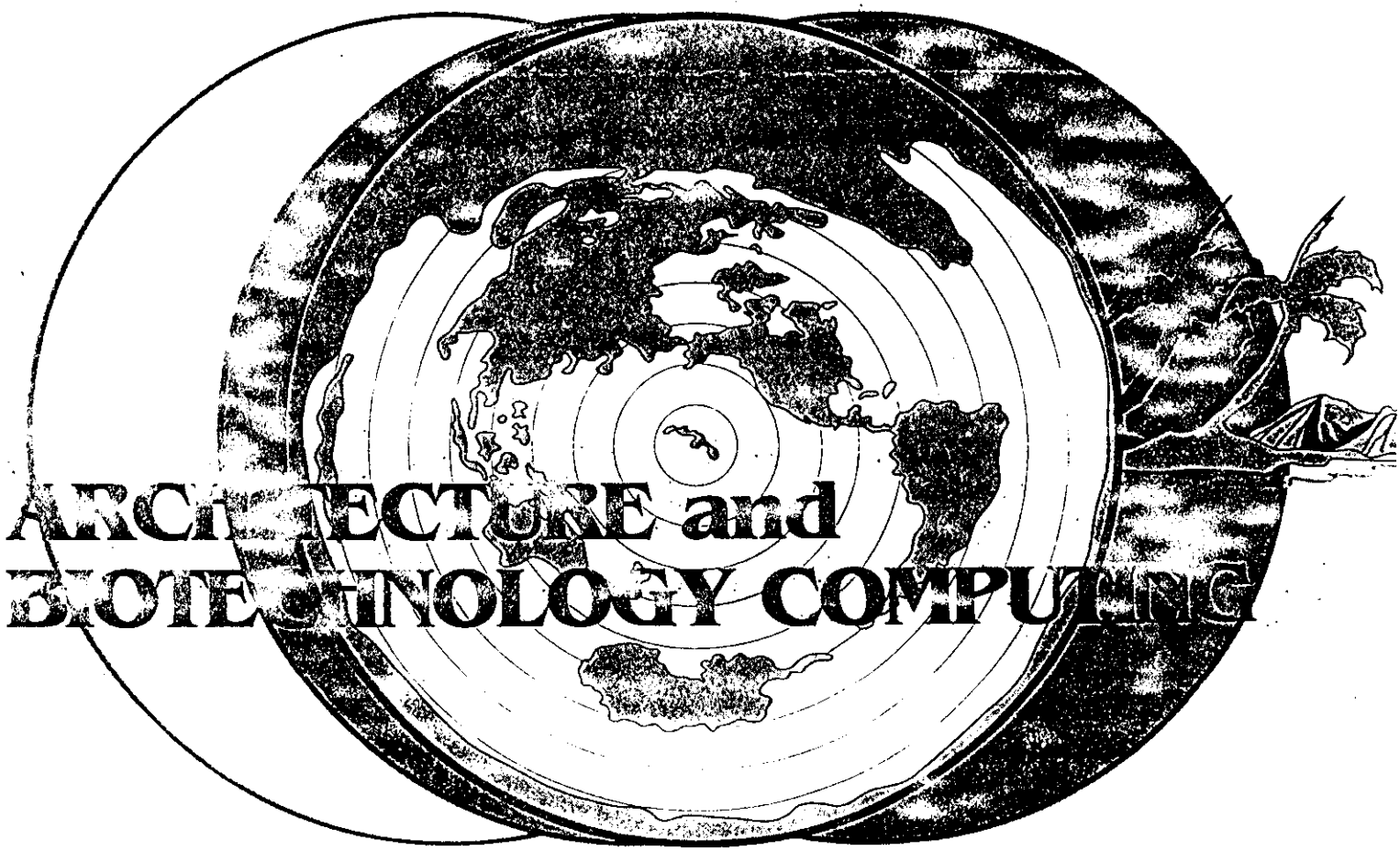
[1] C. L. Brooks III, M. Karplus, and B. M. Pettitt, *Proteins: A Theoretical Perspective of Dynamics, Structure, and Thermodynamics*, Advances in Chemical Physics Vol. LXXI (Wiley, New York, 1988).
[2] J. A. McCammon and S. C. Harvey, *Dynamics of Proteins and Nucleic Acids* (Cambridge University Press, New York, 1987).
[3] See, for example, M. Karplus and J. A. McCammon, Annu. Rev. Biochem. **52**, 263 (1983), and references therein.
[4] U. Burkert and N. L. Allinger, *Molecular Mechanics* (American Chemical Society, Washington, DC 1982).
[5] K. D. Gibson and H. A. Scheraga, Proc. Natl. Acad. Sci. U.S.A. **58**, 420 (1967).
[6] M. Levitt, J. Mol. Biol. **168**, 595 (1983).
[7] B. R. Brooks, R. Bruccoleri, B. Olafson, D. States, S. Swaminathan, and M. Karplus, J. Comput. Chem. **4**, 187 (1983).
[8] S. J. Weiner, P. A. Kollman, D. A. Case, U. C. Singh, C. Ghio, G. Alagona, S. Profeta, and P. Weiner, J. Am. Chem. Soc. **106**, 765 (1984).
[9] H. C. Andersen, J. Chem. Phys. **72**, 2384 (1980).
[10] H. J. C. Berendsen, J. P. M. Postma, W. F. van Gunsteren, A. DiNola, and J. R. Haak, J. Chem. Phys. **81**, 3684 (1984).
[11] J. M. Haile and S. Gupta, J. Chem. Phys. **79**, 3067 (1983).
[12] L. V. Woodcock, Chem. Phys. Lett. **10**, 257 (1971).
[13] S. Nosé, Mol. Phys. **52**, 255 (1984).
[14] S. Nosé, J. Chem. Phys. **81**, 511 (1984).
[15] C. W. Gear, *Numerical Initial Value Problems in Ordinary Differential Equations* (Prentice-Hall, Englewood Cliffs, NJ, 1971).
[16] L. Verlet, Phys. Rev. **159**, 98 (1967).
[17] C. H. Bennett, J. Comput. Phys. **19**, 267 (1975).
[18] A. A. Markov, Izv. Adad. Nauk SPB **VI**, 61 (1907).
[19] S. Duane, A. D. Kennedy, B. J. Pendleto, and D. Roweth, Phys. Lett. B **195**, 216 (1987).
[20] D. W. Heermann, P. Nielaba, and M. Rovere, Comput. Phys. Commun. **60**, 311 (1990).
[21] S. H. Northrup and J. A. McCammon, Biopolymers **19**, 1001 (1980).
[22] S. L. Shumway and J. P. Sethna [Bull. Am. Phys. Soc. **35**, 500 (1990)] present a novel technique for optimizing MC moves using almost Markov simulations. Their specific technique is quite different from ours, but there are similarities between their approach and ours. See also S. L. Shumway and J. P. Sethna, Phys. Rev. Lett. **67**, 995 (1991).
[23] H. Müller-Krumbhaar and K. Binder, J. Stat. Phys. **8**, 1 (1973).
[24] M. P. Allen and D. J. Tildesley, *Computer Simulation of Liquids* (Clarendon, Oxford, 1987), p. 122.
[25] A. Corana, M. Marchesi, C. Martini, and S. Ridella, Assoc. Comput. Mach. Trans. Math. Soft. **13**, 262 (1987).
[26] Charles Brooks has pointed out that the straight averages of $[(\Delta x)^2]$ and $[\eta_i\eta_j\eta_l\eta_m]$ (see Appendix) can be calculated exactly by integrating over the distribution function from which the MC steps are sampled (private communication).
[27] D. C. Rapaport and H. A. Scheraga, Macromolecules **14**, 1238 (1981).
[28] T. Noguti and N. Go, Biopolymers **24**, 527 (1985).
[29] A. Kolinski, J. Skolnick, and R. Yaris, Biopolymers **26**, 937 (1987).
[30] A. Warshel and M. Levitt, J. Mol. Biol. **103**, 227 (1976).
[31] W. Kabsch, Acta Crystallogr. Sect. A **32**, 922 (1976).
[32] A. D. Machlan, J. Mol. Biol. **128**, 49 (1979). See Appendix.
[33] A. M. Ferrenberg and R. H. Swendsen, Phys. Rev. Lett. **63**, 1195 (1989), and references therein.

Proceedings of the Twenty-Sixth Annual Hawaii International Conference on System Sciences

Edited by
TREVOR N. MUDGE
VELJKO MILUTINOVIC
LAWRENCE HUNTER

1993

Vol I



ARCHITECTURE and BIOTECHNOLOGY COMPUTING

IEEE Computer Society Press

The Institute of Electrical and Electronics Engineers, Inc.

Protein Structure Prediction

Minitrack Coordinator: R.H. Lathrop

A Simulated Annealing Approach for Probing Biomolecular Structures

D. Bouzida†, S. Kumar‡, and R. H. Swendsen†
† Department of Physics, Carnegie-Mellon University, Pittsburgh, PA 15213
‡ Department of Biological Sciences, University of Pittsburgh, Pittsburgh, PA 15260

Abstract We present new Monte Carlo methods for finding global energy minima for biological molecules using simulated annealing. The new methods are designed to compensate for the inhomogeneity and local anisotropy of biological molecules. A wide variety of Monte Carlo moves have been implemented, and full optimization of every move individually maintained. Using these new simulated annealing methods, we have investigated the low-energy configurations of a commonly used model of small biological molecules.

1 Introduction

The prediction of the three-dimensional structure of a protein from the sequence of its residues is one of the most important open problems in molecular biology. According to the "thermodynamic hypothesis", proteins in an aqueous environment and under normal conditions generally fold into conformations with the least free energy [1, 2]. Hence, in principle, we need only minimize the free energy to find the biologically active structure. Unfortunately, finding the lowest-energy state turns out to be an extremely difficult computational problem, so that the thermodynamic hypothesis has not really been tested by computational methods.

The problem of finding a global minimum of a function having many local minima belongs to the class of large-scale combinatorial optimization problems which can be solved only approximately on present state-of-the-art computers. Many approaches for solving problems of this sort are available. One can do an exhaustive search of the entire energy surface, but this is feasible only for small peptides comprising a small number of residues. The search will undoubtedly yield a globally optimal solution, but requires a prohibitive amount of computer time for large system sizes. A second approach is to follow the gradients of the potential function to the energy minimum. This energy minimization approach is fundamentally correct but suffers from not sampling far enough from a local minimum. A more feasible approach is to use an approximation algorithm yielding approximate nearly optimal solutions in an acceptable amount of computer time.

The most popular general approach is simulated annealing [3]. Simulated annealing is a stochastic computational approach for finding near-optimal solutions to large non-linear optimization problems [3, 4]. As the temperature of a system is lowered, the particles will rearrange themselves into lower energy states. To obtain the lowest-energy state, the annealing has to be carried out very slowly, so that the system reaches equilibrium at a sequence of decreasing temperatures.

The solutions obtained by simulated annealing do not, in principle, depend on the initial configuration, and the algorithm has been proven to have a polynomial upper bound for its implementation [5, 6]. Furthermore, it is relatively simple to implement and is inherently massively parallel [7].

The main problem with simulated annealing is that it requires highly efficient simulation methods to be successful. There are two broad classes of computational methods that are potentially in use when studying thermodynamic systems: deterministic methods such as Molecular Dynamics (MD) and stochastic methods such as Monte Carlo (MC). Both classes of methods can be used to carry the system from one state to the next in phase space, and thus generate molecular conformations representing the system's thermal fluctuations. The main difference between the two methods resides in the manner in which the conformations are generated.

In the deterministic methods such as molecular dynamice, the time dynamic behavior of the propagation of the system is determined by integrating the classical Newtonian equations of motion. The momenta of the different particles will carry the system from one state in phase space to another. Thus, given the initial state s_0 of the system, a trajectory of the system in phase space is traced. The set of all the produced conformations along the given trajectory form a microcanonical ensemble (constant total energy). In the stochastic methods such as Monte Carlo, the conformations are generated from a Markov process, and are either accepted or rejected with a probability given by the Boltzmann distribution. This idea of importance sampling for drawing sample conformations from "important" regions in phase space was first introduced by Metropolis et al. [8].

In contrast to the original applications of simulated annealing [3], which made use of the Metropolis algorithm, modified molecular dynamics (at constant temperature) has been used to simulate the structures — the main reason being the belief that the direct application of the Metropolis algorithm to macromolecules is inefficient [9]. It is interesting to note, however, that this inefficiency is due mainly to the inhomogeneous

0-8186-1060-3425/93 $03.00

and anisotropic properties inherent in these systems, and the fact that only single atom moves (and thus local moves) have been implemented.

In a previous paper, we have discussed new MC methods for efficient equilibrium simulations [10] that allow us to individually optimize any MC move that might be effective in establishing equilibrium. These methods are not restricted to single atom moves. In fact, collective moves involving many atoms may be easily optimized using this approach. Global rotations of a group of atoms with respect to the rest of the molecule are especially important. These moves are described by angles of rotation, and the only modification needed is to restrict the maximum trial move to the range $[-\pi, \pi]$. The new methods can be especially useful when the rotations are highly correlated. In our simulations, we used both global rotations around chosen flexible bonds and single atom moves.

In this paper, we will describe an extension of these methods to non-equilibrium simulated annealing with new procedures for driving the system to lower minima. We present applications, and discuss the results and implications for future work.

2 Methods

For a fully successful simulated annealing calculation on a physical system, we need an accurate model energy function (force field), a concise description of the state of the system, an efficient method for generating system configurations, and a good annealing schedule.

Force Field The energy function is usually a mechanical molecular model [11] with parameters determined empirically from comparison with experimental data, and from ab initio quantum mechanical calculations. In this work, the parameters are obtained from the AMBER force field of Kollman and coworkers [12]. As often done in previous molecular simulations, a distance-dependent dielectric "constant" is used to simulate screening due to solvent. This choice is not without problems, as discussed below.

States of the System The states of a many-particle system are the set of coordinates and momenta

$$s = \{x_1, x_2, \cdots x_n; p_1, p_2, \cdots p_n\} \qquad (1)$$

where n is the number of degrees of freedom of the system. The set of all the states constitutes the phase space volume. It should be noted that the velocities are automatically integrated out when doing a Monte Carlo simulation, and only the position coordinates are used.

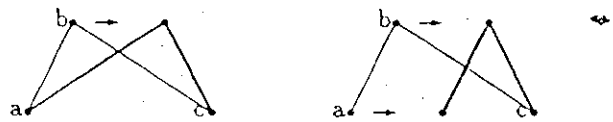
Simulation Methods We have developed two new MC methods to simulate the thermal fluctuations of the system at a given temperature during the annealing [10]. The essence of the new methods stems from the fact that the maximum step size for each atom is not fixed for the duration of the simulation. Instead, it is updated from cycle to cycle according to the information gathered from recent configurations.

The Acceptance Ratio Method (ARM), which can be particularly useful at high temperatures, is based on the acceptance ratio of the different MC moves. The almost exponential dependence of the acceptance ratio on the maximum step size is used to optimize the local jumps of each atom inside a sphere [10]. The ARM method treats each atom separately, and thus deals with the inhomogeneity of the system efficiently.

The Dynamically Optimized Monte Carlo (DOMC) method was developed to treat both the anisotropy and the inhomogeneity of biological molecules by making use of correlations computed during the simulation [10]. The resulting MC moves are generated from ellipsoids instead of spheres to take the anisotropy of the local environment into account.

To efficiently sample the vast phase phase space available to biological molecules, collective moves such as global rotations and translations involving whole groups of atoms, in addition to the single atom jumps which are needed to insure ergodicity, are also optimized using these methods.

At high temperatures, the peptide bonds assume random conformations. A peptide rotation corresponds to a global rotation around the peptide bond of all the atoms on one side of it with respect to the other atoms. As the annealing proceeds, the peptide rotations will drive the peptide bonds to the single energy minimum available to them. As we will see below, this low energy state corresponds to the trans conformation. In addition to the peptide rotations, soft rotations around the Ramachandran angles [13], that we denote (ϕ,ψ) rotations, and side-chain rotations are also implemented and optimized in our annealing calculations.



Single atom move Translational move

Figure 1: Usefulness of translational moves

The usefulness of the translational moves can be seen by considering a simple example of three bonded atoms {a,b,c}, as shown in Figure 1. Let's assume that the bond lengths are not equal to their equilibrium values, and we want to devise moves that will drive the triplet of atoms into equilibrium. A standard MC move of atom b would involve the distortion of both bonds ab and bc. This corresponds to the picture on the left where bond ab is elongated while bond bc is compressed, thus driving the system into another less probable conformation. On the other hand, if we move atom b by dragging atom a with it, the system will equilibrate faster. We denote this second kind of moves a translational move, which is especially useful at the beginning of the annealing run when the ini-

tial configuration of the protein is set up as randomly placed atoms along a long straight chain.

Annealing Schedule We begin the simulated annealing procedure at a high temperature and then decrease the temperature exponentially (i.e. by a constant ratio) [14]

$$T_{i+1} = \gamma T_i \quad \text{where} \quad 0 < \gamma < 1, \tag{2}$$

such that the algorithm runs at each temperature long enough to allow all chosen moves to take place some specified number of times. The factor γ is easily determined from the initial temperature T_i and the final temperature T_f as

$$\gamma = (\frac{T_f}{T_i})^{\frac{1}{Nc-1}}, \tag{3}$$

where Nc is the number of different temperatures (or cycles) at which annealing is performed.

Initially, the simulated annealing explores random configurations at a high temperature. Each lowering of the temperature then restrains state exploration further until a low-energy state is reached. This low-energy state is not guaranteed to be the ground state of the system, as a finite probability of reaching a local minimum exists. This is not a serious problem for a single degree of freedom, since the probability for finding the correct state can be quite high. However, to fold a molecule correctly, many degrees of freedom must have the correct values and the probability of them all being right drops exponentially with the number of variables.

We have solved one aspect of the problem by introducing biased moves at the end of each cycle. To push the system in the direction of the global minimum, we keep track of the energy change ΔE associated with every move during each temperature cycle of the annealing run. At the end of the cycle, a move to the value of each variable corresponding to the lowest relative energy for that variable is attempted, and accepted only if the total energy is lowered. This method ensures that if a variable has been in a low-energy well at any time during the entire simulation, it will end up in that low-energy well.

To illustrate the method, let's consider an angular potential of the form

$$V(\varphi) = A \sin \varphi + \cos 2\varphi, \tag{4}$$

where A is a positive constant and φ can take any value between $[0, 360°]$. This represents a double-well potential with minima at 90° and 270° as shown in Figure 2. The value of A controls the relative depths of the two wells. When A is zero, the wells have equal depths, and the ground state of the system is degenerate. A simulated annealing run will have a 50% chance of ending in any of the two wells. However, as A is increased, the well on the left moves upwards, while the other well shifts downwards. The degeneracy is then lifted, and the ground state of the system is at $\varphi = 270°$. The probability of ending up in either well is given by the Boltzmann distribution. Thus, if N simulated annealing runs are carried out, a large fraction of them will end up in the low-energy well, while a smaller fraction will end up in the other well.

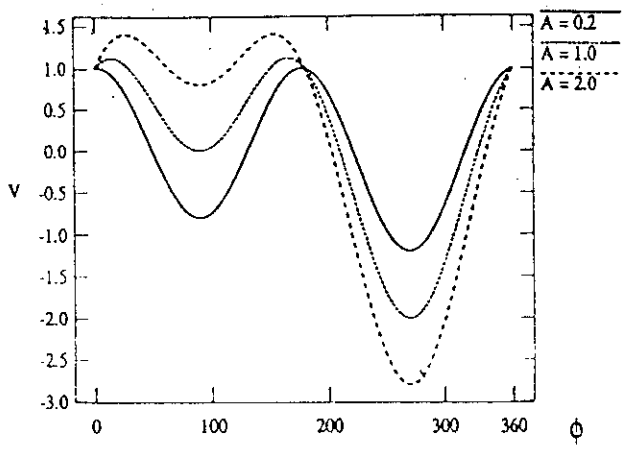


Figure 2: Double-well potentials.

Our objective is to anneal to the global minimum. To push the system in that direction, we keep track of each move $\Delta\varphi$ and energy change ΔE during each temperature cycle of the annealing run. The angle φ_{min} corresponding to the lowest energy E_{min} during the cycle is found and stored. At the end of the cycle,

```
    ⋮
;Given Annealing Cycle at T
    Δφ_min = 0;
    Δφ_sum = 0;
    ;Repeat until equilibrium
        φ' ← φ + Δφ;
        ΔE ← E(φ') − E(φ);
        Metrop_Accept();
        If φ' accepted
          then update(φ, Δφ_min, Δφ_sum);

;Preferential Bias Move
    φ' ← φ + (Δφ_min − Δφ_sum);
    ΔE ← E(φ') − E(φ);
    Accept and update φ only if ΔE < 0;
    ⋮
```

Figure 3: Pseudo-code for a simulated annealing cycle at T including preferential bias moves.

a move to φ_{min} is accepted only if the energy difference is negative. A pseudo-code for this procedure at a given temperature T and inital state φ is shown in Figure 3, where Metrop_Accept is the metropolis acceptance test for which moves are accepted with probability $\min(1, \exp(-\frac{\Delta E}{k_B T}))$.

40

3 Application

In carrying out annealing experiments on a protein using the AMBER force field [12], we also had to deal explicitly with problems concerning the chirality of the different residues and the conformation of the peptide bonds.

We handled the chirality problem by modifying the appropriate term in the Hamiltonian in order to raise the energy of the D-chirality states. For arrangements with the proper chirality, there was no change in the potential.

For the peptide bonds, we wanted to maintain the usual *trans* planar conformation. We modified the value of V_1 in the corresponding dihedral energy term

$$E(\varphi) = \frac{V_1}{2}\,[1 + \cos\varphi] + \frac{V_2}{2}\,[1 - \cos(2\varphi)]. \quad (5)$$

to strongly penalize the *cis* conformations by changing the cis local minimum into a maximum ($V_1 = 4V_2 = 20$ kcal/mole).

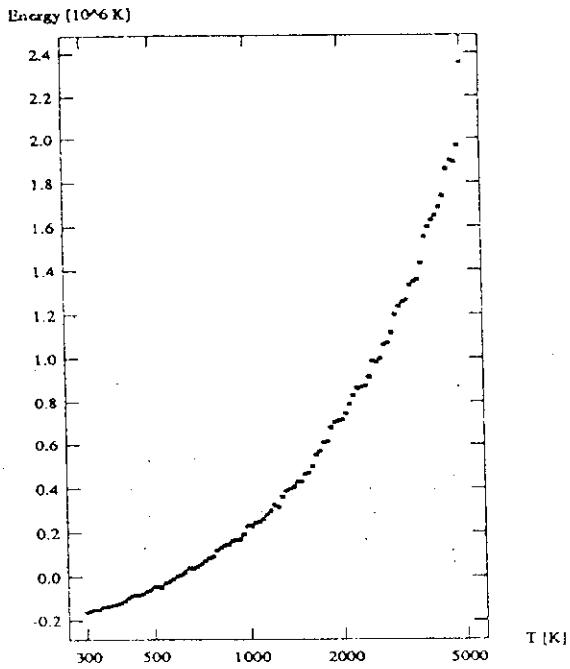


Figure 4: Average energy at each annealing temperature during the second stage $[5000K, 298K]$ of the simulated annealing run on glucagon.

With these changes, we carried out simulated annealing experiments using the ARM method on the glucagon pancreatic hormone. We have chosen this 29-residue protein for its simplicity and the fact that its crystal structure is known fairly well. The experiments were set up in three stages differing in temperature range and in the kinds of moves that were included. The first stage consisted of an annealing at high temperatures $[10^6K \leftrightarrow 5000K]$ starting with random atomic positions along a linear chain of the glucagon sequence residues. These initial conditions

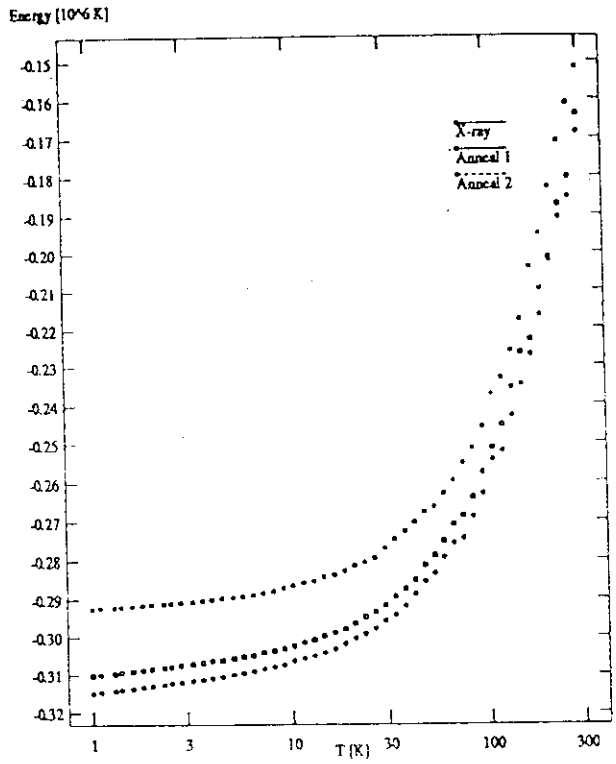


Figure 5: Average energies from three simulated annealing runs, one starting from the equilibrated X-ray structure at $298K$ and the other two from the annealed configurations.

were used because we wanted to establish that the method could efficiently assemble the residues and find the correct chirality. For future simulations, it would be more efficient to use our knowledge of the basic structure of the peptide chain.

In addition to the single atom moves, and the rotations around flexible bonds within the side-chains, we also used translational moves and peptide rotations as described above. During the second stage, the temperature range was between $[5000K \leftrightarrow 298K]$, and the trial moves used in this stage included the single atom moves, rotations around the Ramachandran ϕ and ψ angles, the peptide rotations, and the side-chain rotations. The final stage consisted of lowering the temperature from $298K$ to $1K$ with the same trial moves as the ones used during the second stage.

The average energy at each temperature $[5000K \leftrightarrow 298K]$ of the second stage of the simulated annealing experiment is shown in Figure 4. Each point on the graph corresponds to an equilibrium average for one annealing temperature. The horizontal axis is a logarithmic scale of temperatures. This plot shows a fast relaxation of the system to equilibrium. The average annealing energy obtained at $298K$ is found to be $\sim$ 20kcal/mole smaller than the corresponding

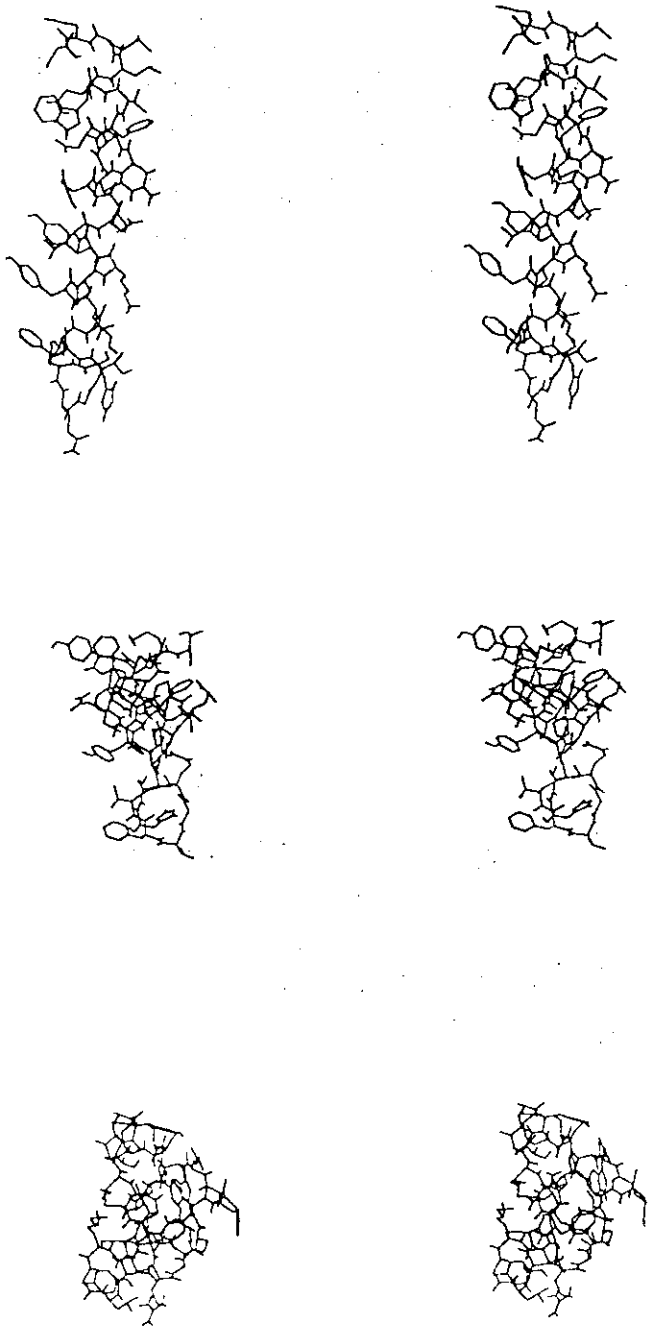


Figure 6: Stereo views of the three final annealed structures of glucagon at 1K. (top) X-ray structure. (middle) Final structure from Annealing I. (bottom) Final structure from Annealing II.

average energy of the equilibrated X-ray structure.

In the third stage [$298K \leftrightarrow 1K$] of the annealing experiment, we had available two initial configurations: the final configuration from stage 2 and the DOMC equilibrated X-ray structure at $298K$. As in previous stages, the ARM method is used with an ideal acceptance ratio set equal to 35% and all (local and global) trial moves are generated uniformly in volume of spherical neighborhoods in their respective spaces. Since glucagon is composed of 29 residues, the global moves include 29 groups of (ϕ, ψ) rotations, 28 peptide rotations, and 128 chosen rotations in the side-chains. During the simulation, the step size ranges and step angle ranges are changed using ARM according to the acceptance ratio obtained.

The Helmholtz free energy F of the protein is given by

$$F = E - TS, \tag{6}$$

where E is the average internal energy and S is the entropy available to it. At very low temperatures, the free energy is equal to the total energy since the entropic effects are negligible. The average energy as a function of the annealing temperatures is plotted in Figure 5. The annealing from the equilibrated X-ray structure at $298K$ actually reaches a state of higher free energy, which is not consistent with the thermodynamic hypothesis. To prove that the result is reproducible, we have carried out another three-stage simulated annealing experiment starting from a different initial random configuration. For comparison, the average energy obtained during the third stage of this run is also shown in Figure 5. Both annealing runs starting from annealed configurations yield a lower free-energy state than the run whose initial configuration is the equilibrated X-ray structure, although they did not find the same low energy state.

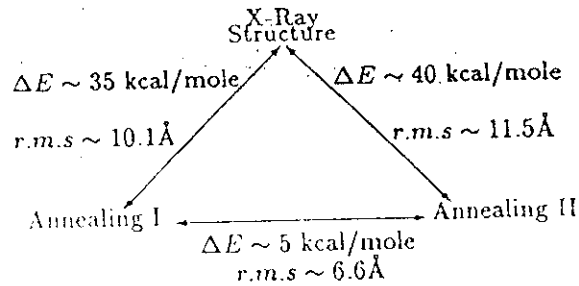


Figure 7: $r.m.s$ and energy differences between the annealed structures.

In Figure 6, we show stereo views of the final structures of glucagon obtained at 1K. The X-ray annealed structure is shown on top, while the two other structures are shown below. The energy differences between the three different annealed structures are schematically shown in Figure 7. As much as 35 and 40 kcal/mole energy differences are found between the annealed structures and the X-ray structure at 1K. In addition to the energy differences, the $r.m.s$ differences are also shown.

4 Interpretation of the Simulated Annealing Results

The annealed configurations are quite distinct from the X-ray structure; they are rather globular instead of helical. The most striking feature is that many of the polar (hydrophilic) side-chains are found in close contact on the inside of the molecule, instead of pointing outward as in the X-ray structure. This behavior directly contradicts some of our most fundamental ideas about folding, since the hydrophilic residues should be attracted to the outside of the molecule where they could be close to water. Something is clearly wrong.

The source of the problem is easy to find. The charged and polar residues experience a strong mutual attraction from the electrostatic interaction in the model we used, even with the effective dielectric "constant" taken to be $\varepsilon(R_{ij}) = R_{ij}$. The strong interaction makes it natural (within the model!) for opposite charges to attract and pack closely inside the molecule. However, this simply means that the form of the "effective" electrostatic interaction is qualitatively incorrect, since it fails to incorporate the distinction between hydrophilic and hydrophobic residues!

It should be noted that this qualitative failure of the Hamiltonian is in addition to the well-known quantitative errors in the strength of the effective electrostatic interaction; the low-temperature configurations of the model will not unfold until the temperature is raised by a few thousand degrees, while the actual molecules denature when the temperature is raised a few tens of degrees.

The conclusion is that the Hamiltonian we used is not a suitable representation of the true interactions of a biological molecule in water. The difficulties we found would not be nearly as obvious in an "equilibrium" simulation starting from a "known" configuration, since the standard representation of the electrostatic interaction is so strong that the model's low-energy states would never be found. However, such "equilibrium" simulations might still suffer from large systematic errors.

For equilibrium simulations, these problems could presumably be avoided by performing the simulations with explicit water molecules. However, that would greatly increase the necessary computer time for annealing calculations because of the surpression of large rotational moves. We therefore conclude that further progress in the energy minimization problem with this approach will depend on finding a qualitatively and quantitatively satisfactory representation of the screened electrostatic interaction at short distances.

5 Summary

In this paper, we have introduced new computer simulation methods for the annealing of models of biological molecules. These new methods are very efficient in finding low-energy configurations. By the objective criterion of how much computer time is necessary to lower the energy, the new methods are much faster than either molecular dynamics or standard MC. In fact, contrary to expectations, we found energy states that were lower than the relaxed crystallographic state, while MD runs with the same amount of CPU time were not able to come within 480 kcal/mole (240,000K) above the energy of the crystallographic state. A corresponding standard MC simulation with single atom moves gave even higher energies. The flexibility of our approach, and the ability to optimize any kind of MC move are great advantages for future work.

On the other hand, application of our methods to the study of glucagon has shown that the most standard Hamiltonians ("force fields") are deficient in their representation of the effect of water on the electrostatic interaction at short distances. The common form of the potential that we used is not only much too strong, but our calculations show that the lack of a proper distinction between hydrophilic and hydrophobic interactions between residues leads to low-energy configurations that do not represent the true behavior of the molecules.

From the point of view that the task of a simulation method is to expose the properties of the model being investigated, we have been quite successful. Even though the Hamiltonian contained many very deep local minima, we were able to find very low-energy configurations efficiently. The prospects for successful simulations of an improved model with a better representation of the electrostatic interaction are very bright.

6 Aknowledgements

We would like to acknowledge support by the National Science Foundation Grant No. DMR-9009475 (DB and RHS) and NIH Grant No. GM25671 (SK). We also thank Charles Brooks, Stephen Garoff, John Nagle, and John Rosenberg for helpful comments and valuable discussion.

References

[1] C. B. Anfinsen. Principles that govern the folding of protein chains. *Science*, 181:223–230, 1973.

[2] C. B. Anfinsen and H. A. Scheraga. Experimental and theoretical aspects of protein folding. *Advances in Protein Chemistry*, 29:205–300, 1975.

[3] S. Kirkpatrick, C. D. Gelatt Jr., and M. P. Vecchi. Optimization by Simulated Annealing. *Science*, 220:671–680, 1983.

[4] S. Kirkpatrick and R. H. Swendsen. Statistical mechanics and disordered systems. *Communications of the ACM*, 28(4):363–373, 1985.

[5] M. Lundy and A. Mees. Convergence of an Annealing Algorithm. *Mathematical Programming*, 34:111–124, 1986.

[6] P. J. M. van Laarhoven and E. H. L. Aarts. *Simulated annealing: Theory and applications*. D. Reidel Publishing Company, Dordrecht, Holland, 1987. Chapter III.

[7] D. R. Greening. Parallel simulated annealing technique. Research Report IBMC 15268, IBM, 1989.

[8] N. Metropolis, A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller, and E. Teller. Equation of state calculations by fast computing machines. *Journal of Chemical Physics*, 21(6):1087–1092, 1953.

[9] S. H. Northrup and J. A. McCammon. Simulation methods for protein structure fluctuations. *Biopolymers*, 19:1001–1016, 1980.

[10] D. Bouzida, S. Kumar, and R. H. Swendsen. Efficient Monte Carlo methods for the computer simulation of biological molecules. *Physical Review A.*, 45(12):8894–8901,1992.

[11] U. Burkert and N. L. Allinger. *Molecular mechanics*. American Chemical Society, Washington, D. C., 1982.

[12] S. J. Weiner, P. A. Kollman, D. A. Case, U. C. Singh, C. Ghio, G. Alagona, S. Profeta, and P. Weiner. A new force field for molecular mechanical simulation of nucleic acids and proteins. *Journal of the American Chemical Society*, 106:765–784, 1984.

[13] G. N. Ramachandran, C. Ramakrishnan, and V. Sasisekharan. Stereochemistry of polypeptide chain configurations. *Journal of Molecular Biology*, 7:95–99, 1963.

[14] M. Widom, K. J. Strandburg, and R. H. Swendsen. Quasicrystal equilibrium state. *Physical Review Letters*, 58(7):706–709, 1987.

The Weighted Histogram Analysis Method for Free-Energy Calculations on Biomolecules. I. The Method

Shankar Kumar,[1] Djamal Bouzida,[2] Robert H. Swendsen,[2] Peter A. Kollman,[3] and John M. Rosenberg[1]*

[1]*Department of Biological Sciences, University of Pittsburgh, Pittsburgh, Pennsylvania 15260*
[2]*Department of Physics, Carnegie Mellon University, Pittsburgh, Pennsylvania 15213*
[3]*Department of Pharmaceutical Chemistry, University of California-San Francisco, San Francisco, California 94143*

Received 13 February 1992; accepted 28 April 1992

The Weighted Histogram Analysis Method (WHAM), an extension of Ferrenberg and Swendsen's Multiple Histogram Technique, has been applied for the first time on complex biomolecular Hamiltonians. The method is presented here as an extension of the Umbrella Sampling method for free-energy and Potential of Mean Force calculations. This algorithm possesses the following advantages over methods that are currently employed: (1) It provides a built-in estimate of sampling errors thereby yielding objective estimates of the optimal location and length of additional simulations needed to achieve a desired level of precision; (2) it yields the "best" value of free energies by taking into account all the simulations so as to minimize the statistical errors; (3) in addition to optimizing the links between simulations, it also allows multiple overlaps of probability distributions for obtaining better estimates of the free-energy differences. By recasting the Ferrenberg-Swendsen Multiple Histogram equations in a form suitable for molecular mechanics type Hamiltonians, we have demonstrated the feasibility and robustness of this method by applying it to a test problem of the generation of the Potential of Mean Force profile of the pseudorotation phase angle of the sugar ring in deoxyadenosine.

INTRODUCTION

Several methods have been used to calculate the changes in the free energies between interacting molecules and to investigate relative stabilities of the different conformational states of a given molecule with respect to a conformation coordinate of interest. Such calculations are especially important in providing valuable insight into the role of structure–function relationships in biomolecular interactions and in providing a rational basis for the design and modeling of new drugs. However, free-energy calculations for large molecules are computationally demanding, because the entropy that depends on the extent of the phase space of the molecular system cannot generally be extracted from a simple ensemble average of some property of the given system. Hence, new methods for fast, efficient, and accurate determination of free-energy differences are needed. An increase in efficiency can be achieved in two ways: (1) by improving the efficiency of the simulational method itself and (2) by maximizing the amount of information obtained from either Monte Carlo (MC) or Molecular Dynamics (MD) simulations. This article deals with (1) the Single Histogram (SH) method and (2) the Extended Ferrenberg-Swendsen (WHAM) algorithm, which belong to the latter category; the WHAM equations developed here are extensions of the Multiple Histogram equations developed by Ferrenberg and Swendsen.[1-3] The SH and WHAM methods are applicable for both (constant temperature) MD and MC simulations. Methods for increasing the efficiency of the simulational protocol have been discussed elsewhere.[4-6]

We will first describe the nature of the problems that can be treated by these methods. This will be followed by a brief description of the SH and WHAM equations that can be used for biomolecular systems. An outline of the derivation of the WHAM equations will be given in the Appendix. Finally, we will apply these methods to generate the Potential of Mean Force (PMF) profile of the pseudorotation phase angle of the sugar ring in deoxyadenosine with the objective being to demonstrate the feasibility and robustness of the WHAM algorithm when applied to biomolecular systems.

*Author to whom all correspondence should be addressed.

Journal of Computational Chemistry, Vol. 13, No. 8, 1011–1021 (1992)

CCC 0192-8651/92/081011-11$04.00

BASIC STRUCTURE OF THE PROBLEM

The problem of calculating free energies can be broadly divided into two classes for computational purposes: (1) those involving the generation of a PMF profile along a coordinate and (2) those involving the computation of free-energy differences as a given molecular system is modified from a standard initial state to a final state. The latter are special cases of the former class. The approaches that have been commonly used so far in the solution of these problems are Free Energy Perturbation (FEP), and Umbrella Sampling methods.[7-15]* In both the Umbrella Sampling and the FEP methods the Hamiltonian $\hat{H}_0(x)$ is replaced by a modified potential, $\hat{H}_{\{\lambda\}}$, of the form

$$\hat{H}_{\{\lambda\}}(x) = \hat{H}_0(x) + \sum_{i=1}^{L} \lambda_i \hat{V}_i(x) = \sum_{i=0}^{L} \lambda_i \hat{V}_i(x) \quad (1)$$

with $\lambda_0 = 1$ and $\hat{V}_0(x)$ defined as being identical to $\hat{H}_0$. Circumflexes over the symbols denote functions.† Here the coordinates of the atoms of the molecule are represented by x; the L functions, $\hat{V}_1(x)$, $\hat{V}_2(x)$, . . . , $\hat{V}_L(x)$, are restraining potentials. The restraining potentials are functions of the molecular coordinates x. The λ_i are coupling parameters. The symbol in braces, $\{\lambda\}$, denotes the set of values $\lambda_1, \lambda_2, \lambda_3, \ldots, \lambda_L$. Thus $\{0\}$ indicates that all the λ_i ($i = 1, 2, \ldots, L$) have been set to zero; unless stated otherwise λ_0 always takes on the value of unity. The restraining potentials are chosen in such a manner that the sampling distribution is shifted along a coordinate of interest such as a reaction coordinate. Multiple restraining potentials are useful for sampling "long" reaction pathways where separate simulations with different coupling parameters $\{\lambda\}$ are carried out to sample different regions of the reaction path. The reaction coordinate (termed ξ here) will be a function of x. By adjusting the values of λ_i in eq. (1) any region of interest along the coordinate ξ can be preferentially sampled. Free energies (or PMF values) can then be obtained after corrections for the restraining potential; relative free energies can also be obtained as a function of the coupling parameters.‡

In the problem discussed here, ξ is the Pseudorotation Phase Angle[17-19] of the sugar ring in the nucleic acid base deoxyadenosine. The Hamiltonian is written as

$$\hat{H}_\lambda(x) = \hat{H}_0(x) + \lambda \sum_{i=0}^{3} [1.0 + \cos(\nu_i - \alpha_i + \pi)] \quad (2)$$

The ν_i in eq. (2) refer to the usual sugar torsion angles and are restrained to the values α_i. Here, $\alpha_0 = 36.14°$, $\alpha_1 = 337.6°$, $\alpha_2 = 0.0°$, and $\alpha_3 = 22.34°$; $\hat{H}_0(x)$ is the AMBER all-atom force field of Kollman and coworkers.[20,21] The Hamiltonian $\hat{H}_\lambda(x)$ of eq. (2) has only one restraining potential with

$$\hat{V}(x) = \sum_{i=0}^{3} [1.0 + \cos(\nu_i - \alpha_i + \pi)] \quad (3)$$

The α_i in eqs. (2) and (3) have been chosen so as to bias the sampling toward the energetically unfavorable region in the vicinity of the O_4'-exo conformation. The restraint is on the torsion angles that determine the pseudorotation phase angle and is chosen to enhance sampling in the neighborhood of $\xi = 270°$. The pseudorotation phase angle is not a simple function of the coordinates, thus requiring a complicated restraining potential. Simulations can be carried out with the coupling parameter λ set at various values so as to minimize statistical errors.

The Umbrella Sampling and FEP equations for simulations carried out with multiple restraining potentials as given in eq. (1) are given below primarily to explain the notations used here.

The probability density $P_{\{\lambda\},\beta}(\xi)$ obtained from a simulation with the Hamiltonian $\hat{H}_{\{\lambda\}}$ [as in eq. (1)] can be written as

$$P_{\{\lambda\},\beta}(\xi) = \exp[-\beta W_{\{\lambda\},\beta}(\xi)] = \langle \delta[\xi - \hat{\xi}(x)] \rangle_{\{\lambda\},\beta} \quad (4)$$

The angular brackets denote ensemble averages and the subscripts refer to the values of the coupling parameters λ_i and to the parameter β given by

*We are following terminology currently in use in the field of biomolecular simulations when we refer to methods described here as "Umbrella Sampling." These same methods are sometimes referred to as "Multistage Sampling" because of historical distinctions between the original Umbrella Sampling and Multistage Sampling methods.

†Thus $\hat{V}_i(x)$ denotes the function and V_i a particular *value* the function takes; circumflexes will be used only where ambiguities might arise.

‡A recent method for calculating PMFs along "internal coordinates" of interest is due to Tobias and Brooks.[16] In this method a holonomic constraint is used to fix the coordinate (analogous to the SHAKE algorithm) at a series of values at which the relative free energies (or PMFs) are calculated. This method is well suited to simple reaction coordinates such as a hydrogen bonding distance. However, it is not clear how to apply this method to situations where the coordinate of interest is a complicated function of internal coordinates as in the case of the pseudorotation phase angle that is discussed here; applying constraints to many internal coordinates could lead to improper sampling of conformational space.

$\beta = 1/k_B T$ where k_B is the Boltzmann constant and T is the temperature. $W_{\{\lambda\},\beta}(\xi)$ is the PMF associated with ξ when the simulation is carried out with the coupling parameters $\{\lambda\}$ at temperature T.

If $P_{\{0\},\beta}(\xi)$ is the probability density obtained from an unbiased sampling, i.e., with all the λ_i (except λ_0 which is equal to one) set to zero, then

$$P_{\{0\},\beta}(\xi) = \exp[-\beta W_{\{0\},\beta}(\xi)] \tag{5}$$

or

$$P_{\{0\},\beta}(\xi) = \frac{Z_{\{\lambda\},\beta}}{Z_{\{0\},\beta}} \times \left\langle \delta[\xi - \hat{\xi}(x)] \prod_{i=1}^{L} \exp[\beta \lambda_i \hat{V}_i(x)] \right\rangle_{\{\lambda\},\beta} \tag{6}$$

where Z is the partition function. If we restrict the restraining potential $\hat{V}_i(x)$ to be functions of the coordinate ξ only—that is if

$$\hat{V}_i(x) \equiv \hat{V}_i[\hat{\xi}(x)] \tag{7}$$

then $P_{\{\lambda\},\beta}(\xi)$ will be related to $P_{\{0\},\beta}(\xi)$ by

$$P_{\{\lambda\},\beta}(\xi) = D(\{\lambda\}, \beta) P_{\{0\},\beta}(\xi) \exp\left[-\sum_{k=1}^{L} \lambda_k \beta \hat{V}_k(\xi) \right] \tag{8}$$

and $W_{\{\lambda\},\beta}$ is related to $W_{\{0\},\beta}$ by[8,9]

$$W_{\{0\},\beta}(\xi) = -\sum_{j=1}^{L} \lambda_j \hat{V}_j(\xi) + W_{\{\lambda\},\beta}(\xi) + C(\{\lambda\}, \beta) \tag{9}$$

where the functions $D(\{\lambda\}, \beta)$ and $C(\{\lambda\}, \beta)$ are given by

$$D(\{\lambda\}, \beta) = \frac{Z_{\{0\},\beta}}{Z_{\{\lambda\},\beta}},$$
$$C(\{\lambda\}, \beta) = \beta^{-1} \ln D \tag{10}$$

The equations given above can be extended to situations where the parameter β is also varied. Equation (9) is the form that has been used most often in estimating free-energy differences and generating PMF profiles by using the Umbrella Sampling method. The method can also be used to calculate free-energy differences as a function of any coupling parameter λ_k. The Umbrella Sampling method, for instance, can be used to estimate the free energy of binding between receptor and ligand molecules, when the binding takes place along a suitable path of approach (or a reaction coordinate). By choosing a different set of λ_i for each simulation such that successive simulations sample overlapping regions along ξ, the function $C(\{\lambda\}, \beta)$ in eq. (9) can be determined so as to make $W_{\{0\},\beta}(\xi)$ agree in the regions of overlap.[12-14]

The standard FEP equations can be readily generalized to the case of multiple restraining potentials as follows:

$$\frac{\partial W_{\{\lambda\},\beta}}{\partial \lambda_k} = \frac{\langle \hat{V}_k \delta[\xi - \hat{\xi}(x)] \rangle_{\{\lambda\},\beta}}{\langle \delta[\xi - \hat{\xi}(x)] \rangle_{\{\lambda\},\beta}} - \langle \hat{V}_k \rangle_{\{\lambda\},\beta} \tag{11}$$

The FEP methods are generally used in situations where the Hamiltonian is changed in small steps so that a given molecule can be "mutated" to a desired end state gradually. By calculating the free-energy changes that occur at each step and by finally summing these free-energy changes, the total free-energy change can be obtained. As a typical example, consider a Hamiltonian of the form

$$\hat{H}_\lambda = (1 - \lambda)\hat{H}_i + \lambda \hat{H}_f = \hat{H}_i + \lambda(\hat{H}_f - \hat{H}_i) \tag{12}$$

where $\hat{H}_i$ and $\hat{H}_f$ could be the Hamiltonian for a "wild-type" and mutated biomolecule; here, λ is a coupling parameter and by varying λ slowly from 0 to 1 the system can be taken from its initial state to its desired end state. Equation (12) is a special case of eq. (1). For the special case of eq. (12) the discretized forms of the FEP equations for the free energy A are

$$\beta[A(\lambda = 1) - A(\lambda = 0)] = -\sum_{i=1}^{i=n} \ln\langle \exp(-\beta[\hat{H}_{\lambda_{i+1}} - \hat{H}_{\lambda_i}]) \rangle_{\lambda_i} \tag{13}$$

and

$$A(\lambda = 1) - A(\lambda = 0) = \sum_{\lambda=0}^{\lambda=1} \left\langle \frac{\partial \hat{H}_\lambda}{\partial \lambda} \right\rangle_\lambda \Delta\lambda = \sum_{\lambda=0}^{\lambda=1} \langle \hat{H}_f - \hat{H}_i \rangle_\lambda \Delta\lambda \tag{14}$$

n in eq. (13) is the number of intervals between $\lambda = 0$ and $\lambda = 1$ over which the summation is carried out. Equations (13) and (14) are the basic FEP equations. Sometimes the implementation of eq. (13) has been referred to as the "Windowing" method and that of eq. (14) as the "Integration" method in the literature. The FEP equations do not have an in-built estimate of errors which makes it difficult for estimating statistical errors in the results. The WHAM algorithm does provide for objective estimation of statistical errors [see eq. (22)].

To summarize: PMF profiles and free-energy differences have been calculated thus far generally by using Umbrella Sampling techniques that use eqs. (7), (8), and (9) and by using FEP methods that utilize eqs. (12), (13), and (14).

Normal Mode analyses[7-22] have also been used in the investigation of the relative stabilities of different conformational states of a molecule.[23] However, conformational states of a biomolecule are characterized by transitions across several energy minima and therefore Normal Mode methods can-

not give a reasonable estimate of the entropy of the biomolecule.

The basic problem then is this: What happens to the free energy as some parameter (or set of parameters) is varied? We present below the Single Histogram and Multiple Histogram equations, which we can use to study the behavior of the free energy as some parameter—either a "coupling" parameter λ or the temperature T—is changed. The WHAM equations presented here are essentially those of Ferrenberg and Swendsen,[1-3] but have been extended to the case of molecular mechanics potentials that characterize biomolecules and can readily be applied to situations where free energies and PMFs are needed as a function of the coupling parameter(s) λ_i and/or the temperature T.

We have tested the SH and WHAM equations on the problem of generating the PMF profile of the pseudorotation phase angle of the sugar ring in deoxyadenosine, the main purpose of this study being to test the feasibility and robustness of the histogram equations when applied to molecular mechanics type potentials that characterize biomolecules. Although the AMBER "All-Atom" force-field of Kollman and coworkers was used in this study the efficiency of the method should not depend upon the particular Hamiltonian that is being used. Applications of these methods to larger systems are in progress.*

SINGLE AND MULTIPLE HISTOGRAM METHODS

The partition function $Z_{\{\lambda\},\beta}$ of a system whose Hamiltonian is given by eq. (1) is

$$Z_{\{\lambda\},\beta} = \sum_{\{V\},\xi} \Omega(\{V\}, \xi) \prod_{i=0}^{L} e^{-\lambda_i \beta V_i} \qquad (15)$$

where $\Omega(\{V\}, \xi)$ is a generalized density of states given by

$$\Omega(\{V\}, \xi) = \int dx\, \delta[\xi - \hat{\xi}(x)] \prod_{i=0}^{L} \delta[V_i - \hat{V}_i(x)] \qquad (16)$$

$\Omega(\{V\}, \xi)$ is independent of $\{\lambda\}$ and β. The SH and WHAM methods can be applied when the partition function is of the form given in eqs. (15) and (16).

An outline of the derivation of the SH and WHAM equations is given in the Appendix. In this section, we will first describe how to obtain PMFs and probability densities from a single simulation using SH equations before generalizing to the case of multiple simulations.

*For an interesting account of the history of histogram techniques see Ferrenberg's thesis.[1]

Single Histogram Equations

The first description of the SH equations dates back to 1959 and is due to Salsburg, Jacobsen, Fickett, and Wood.[24] We will present the "operational" form of the SH equations as applied to biomolecular systems here. Using these equations, the objective generally is to generate the PMF profile of the coordinate ξ from a single simulation (and hence the term "Single Histogram"). Let us suppose that a simulation was carried out at temperature $T_1 = 1/k_B\beta_1$ with λ_0 set to one and with the restraining potentials appropriately weighted by the coupling parameters $\lambda_1, \lambda_2, \ldots, \lambda_L$ (to enhance sampling in high energy regions). The quantity of interest is then the probability $\tilde{P}_{\beta_2}(\xi)$ that the coordinate ξ would take if a simulation were done with $\lambda_0 = 1$ and all the other coupling parameters set to zero at a temperature $T_2 = 1/k_B\beta_2$. Generally, $T_1 > T_2$ so as to enhance conformational sampling in high energy regions along ξ. By taking the logarithm of the probabilities PMF profiles can be generated. The data is put into "bins" to generate histograms and the "operational" form of the SH equations becomes

$$\tilde{P}_{\beta_2}[\xi \in (\xi_m, \xi_{m+1})] = \frac{\sum_{j=1}^{\eta(m)} \exp\left[(\beta_1 - \beta_2)\tilde{V}_{0,j}^{(m)} + \sum_{i=1}^{L} \lambda_i \beta_1 \tilde{V}_{i,j}^{(m)}\right]}{\sum_{k=1}^{B} \sum_{j=1}^{\eta(k)} \exp\left[(\beta_1 - \beta_2)\tilde{V}_{0,j}^{(k)} + \sum_{i=1}^{L} \lambda_i \beta_1 \tilde{V}_{i,j}^{(k)}\right]} \qquad (17)$$

where the expression now gives the probability that ξ has the value between ξ_m and ξ_{m+1}—the mth bin—at the temperature T_2. $\tilde{V}_{i,j}^{(k)}$ is the value that the restraining potential V_i takes at the jth snapshot of the kth bin. $\eta(k)$ is the total number of data points that the simulation yielded in the kth bin; it is just the value taken on by the histogram at the bin numbered k. B is the total number of bins that the data has been divided into.

Equation (17) can also be expressed in terms of $N_{\{\lambda\},\beta_1}(\{V\}, \xi)$ where $N_{\{\lambda\},\beta_1}(\{V\}, \xi)$ is the value taken by the histogram at $\{V\}$ and ξ during the simulation at temperature $T_1 = 1/k_B\beta_1$ and with the coupling parameters set to $\{\lambda\}$. Again, $\tilde{P}_{\beta_2}(\xi)$ refers to the probability of occurrence of the coordinate ξ during a simulation performed at temperature T_2 with no restraints. In terms of $N_{\{\lambda\},\beta_1}(\{V\}, \xi)$ we have†

†Summation over $\{V\}$ as in eq. (18) denotes summation over the possible values of $V_1, V_2, \ldots, V_L$. Similar remarks apply to summation over $\{\lambda\}$.

$$\tilde{P}_{\beta_2}(\xi) = \frac{\sum_{\{V\}} N_{\{\lambda\},\beta_1}(\{V\}, \xi) \exp\left[(\beta_1 - \beta_2)V_0 + \sum_{i=1}^{L} \lambda_i \beta_1 V_i\right]}{\sum_{\{V\},\xi} N_{\{\lambda\},\beta_1}(\{V\}, \xi) \exp\left[(\beta_1 - \beta_2)V_0 + \sum_{i=1}^{L} \lambda_i \beta_1 V_i\right]} \quad (18)$$

WHAM Equations

The WHAM equations are a natural generalization of the SH equations. Simulations are carried out with various sets of coupling parameters to enhance conformational sampling. PMFs are then calculated for the case when a simulation is done with the desired set of coupling parameters at a specified temperature. We will state the main results first and an outline of the derivation of the WHAM equations will be presented in the next section.

Consider R simulations with the ith simulation being carried out at temperature $T_i = 1/k_B\beta_i$ with the coupling parameters in eq. (1) set to $\{\lambda\}_i$*; also, let the number of snapshots taken from the ith simulation be n_i. Then the (unnormalized) probability histogram $P_{\{\lambda\},\beta}(\{V\}, \xi)$ is given by[1-3]†

$$P_{\{\lambda\},\beta}(\{V\}, \xi) = \frac{\sum_{k=1}^{R} N_k(\{V\}, \xi) \exp\left(-\beta \sum_{j=0}^{L} \lambda_j V_j\right)}{\sum_{m=1}^{R} n_m \exp\left(f_m - \beta_m \sum_{j=0}^{L} \lambda_{j,m} V_j\right)} \quad (19)$$

and

$$\exp(-f_j) = \sum_{\{V\},\xi} P_{\{\lambda\}_j,\beta_j}(\{V\}, \xi) \quad (20)$$

where $N_i(\{V\}, \xi)$ is the value taken by the histogram at $\{V\}$ and ξ during the ith simulation, and f_j is the (dimensionless) free energy of the system described by the Hamiltonian of eq. (1) with coupling parameters $\{\lambda\}_j$; $f_j = \beta_j A_j$ where A_j is identical to the (Helmholtz) free energy of the system during the jth simulation. Equations (19) and (20) were derived by minimizing the errors (see Appendix) in the overlapping probability distributions.[1-3] By iterating eqs. (19) and (20) the f_i and, therefore, the free energies, can be determined self-consistently. For the case of a single simulation the WHAM equations reduce to the SH equations except for a normalization factor.

We can compute the f_i directly from the data (to reduce computational errors) by using the following expression:

$$\exp(-f_i) = \sum_{k=1}^{R} \sum_{t=1}^{n_k} \frac{\exp\left[-\beta_i \sum_{j=0}^{L} \lambda_{j,i} V_{j,t}^{(k)}\right]}{\sum_{m=1}^{R} n_m \exp\left[f_m - \beta_m \sum_{j=0}^{L} \lambda_{j,m} V_{j,t}^{(k)}\right]} \quad (21)$$

In this expression $V_{j,t}^{(i)}$ is the value that the restraining potential V_j takes at the tth snapshot of the lth simulation.

One can start with an arbitrary (but not too unreasonable) set of values for the f_i; a good starting point would be to set all the f_i to zero initially. Convergence was generally very fast for the problem discussed here with the number of iterations being less than 10 and no special care was needed for the initial assignment of values for the f_i to accelerate convergence. However, it is quite possible that free-energy calculations for some systems could benefit from acceleration techniques; the interested reader is referred to Ferrenberg's thesis.[1]

The relative error, $\delta\Omega(\{V\}, \xi)/\Omega(\{V\}, \xi)$, (see Appendix) can be shown to be[1-3]

$$\frac{\delta\Omega(\{V\}, \xi)}{\Omega(\{V\}, \xi)} = \left[g^{-1} \sum_{i=1}^{R} N_i(\{V\}, \xi)\right]^{-1/2} \quad (22)$$

Thus, by knowing where the $\delta\Omega/\Omega$ are high more simulations can be done with the appropriate value of the coupling parameters thus reducing the error by increasing the statistics obtained from the simulations. An overall factor g, with $g = 1 + 2\tau$ where τ is an integrated correlation time[25] for the simulations has been included in the equation; since only the relative magnitudes of the quantity $\delta\Omega(\{V\}, \xi)/\Omega(\{V\}, \xi)$ are of interest the quantity g may safely be omitted.

APPLICATION OF THE HISTOGRAM EQUATIONS

We will now demonstrate the use of Single and Multiple Histogram equations by applying them to estimate the PMF of the Pseudorotation Phase Angle of the sugar ring in deoxyadenosine. While this system is small its Hamiltonian contains most of

*$\{\lambda\}_k$ refers to the value of the coupling parameters during the kth simulation, that is, $\{\lambda\}_k$ denotes the set $\{\lambda_1, \lambda_2, \ldots, \lambda_L\}_k$ which is identical to $\{\lambda_{1,k}, \lambda_{2,k}, \ldots, \lambda_{L,k}\}$.

†In the original formulation by Ferrenberg and Swendsen the equations for the probability distributions contained factors, g_i, that depended upon the integrated correlation times of a simulation; these have been omitted here in eq. (19). For biomolecular systems these factors are approximately equal for each simulation and therefore cancel out of eq. (19). In fact, for biomolecular systems, we have ascertained that these factors make negligible difference to the results even if they differed by factors of 9 or 10. The g_i, however, should not be neglected when phase transitions are involved (see the Appendix and next section).

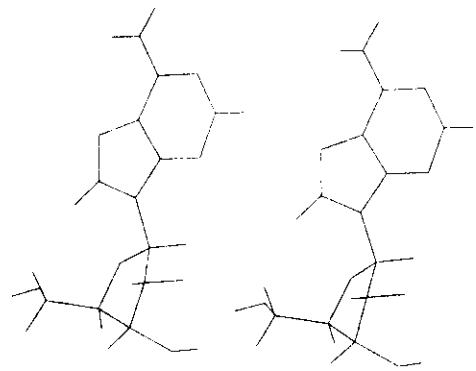


Figure 1. Stereoview of deoxyadenosine.

the complexity of larger molecular systems and thus presents a good test case for the WHAM method.

The system (Fig. 1) consists of 31 atoms; its Hamiltonian is given in eq. (2) with $\hat{H}_0(x)$ being the AMBER "All-Atom" force-field. The restraining potential $\hat{V}_i(x)$ is necessary, for without it very poor statistics are obtained for the pseudorotation phase angle around the $O_{4'}$-exo region ($\xi = 270°$) (Fig. 2). Figure 3 shows the corresponding histogram from a simulation carried out at 298 K with $\lambda = 1.4$; the sampling in the $O_{4'}$-exo region is seen to be better than when there was no restraining potential.

Data from different MD simulations were taken. One simulation was carried out at 250 K; the rest were done at either 298 or 350 K. To eliminate the high frequency bond vibrations, bond lengths were constrained to the values in the AMBER[20] database using SHAKE.[26] All the simulations were done with the restraint given in eq. (2) but with a different value of the coupling parameter λ [see eq. (1)]. The starting coordinates of the molecule were obtained from the AMBER database. Prior to the MD phase of each simulation, the molecular structure was relaxed using the method of Conjugate Gradients[27,28] to an energy gradient of the order of 10^{-2} kcal/Å^2 mol. The MD updates were done with the AMBER program using the leap-frog algorithm[35]; temperature was maintained constant by coupling the system to a heat bath as proposed by Berendsen et al.[30] A distance-dependent dielectric function[21] was used in this study. The details of the MD runs are summarized in Table I.

RESULTS AND DISCUSSION

The first test of the WHAM equations was to see whether or not the calculated free energies were independent of the arbitrarily assigned initial values for the f_j. Since the correlation times were about the same in all the MD runs each g_i (see second footnote on p. 1015 and the Appendix) was set equal to one. Convergence was very fast and was achieved in less than 20 iterations irrespective of the initial values of f_j. The free energies obtained from different starting values of f_i are identical. It can be seen that the "All-Atom" force field of Kollman and coworkers gives a barrier of about 2.5 kcal/mol for a $C_{2'}$-endo ($\xi \approx 144°$) to $C_{3'}$-endo ($\xi \approx 36°$) transition via the $O_{4'}$-exo ($\xi \approx 270°$) region; about 0.5 kcal/mol for the $C_{2'}$-endo → $C_{3'}$-endo

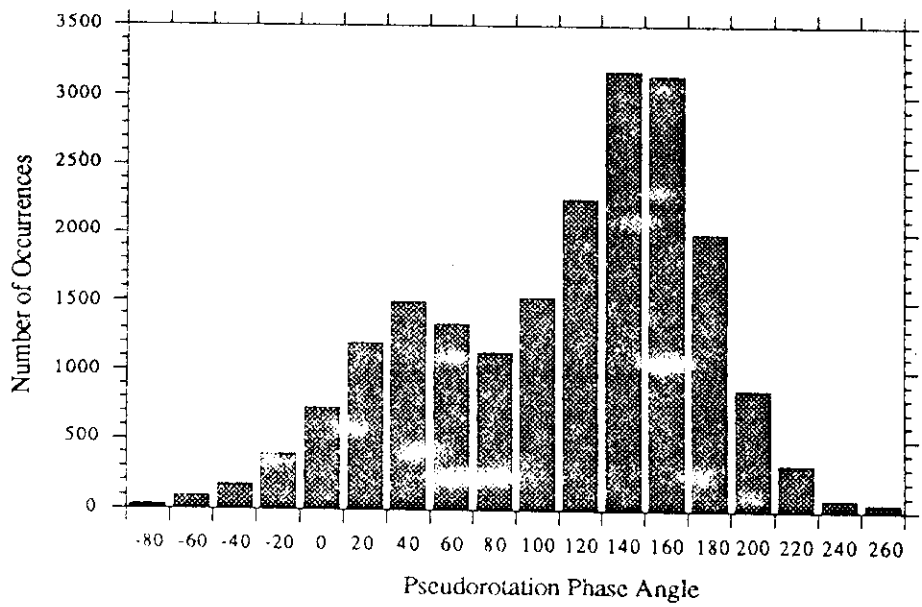


Figure 2. Histogram of the pseudorotation phase angle from simulation 1 (see Table I).

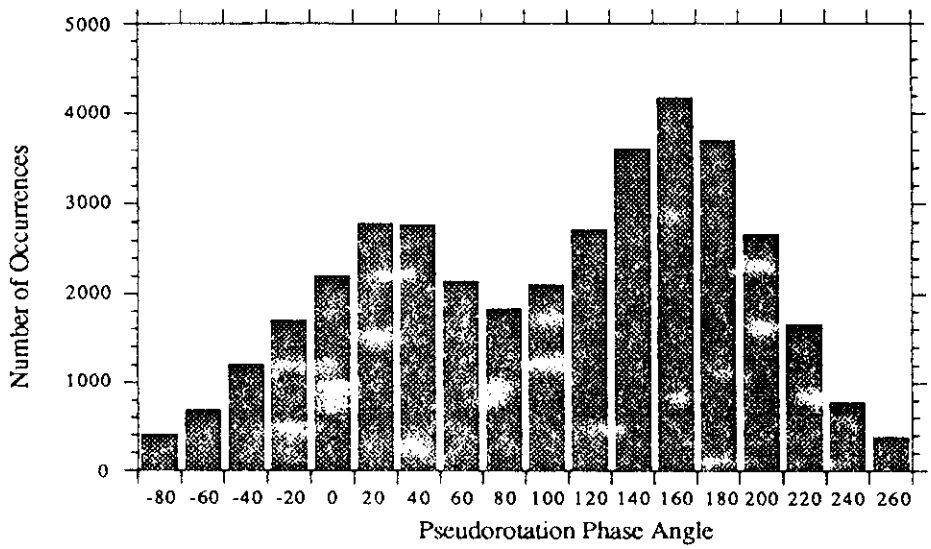


Figure 3. Histogram of the pseudorotation phase angle with $\lambda = 1.4$.

transition via the O_4-endo ($\xi \approx 90°$) region; and almost no barrier for the C_3-endo → C_2-endo transition via the O_4-endo region (see Fig. 4).

Due to the difficulty in measuring correlation times accurately it is important that the free-energy differences do not depend strongly on the relative magnitudes of g_i. The values of g_i were now varied over a wide range (from 1–10) (keeping the initial values of f_i equal to zero) to ascertain that the calculated free energies would not differ greatly from the values determined with the g_i set to one. In spite of the wide variation in the ratios of the g_i, we found that the discrepancies in the free energies were negligible. For example, when the WHAM calculations were carried out on simulation 2–7 with $\{g_2 = 3.6, g_3 = 1.8, g_4 = 1.0, g_5 = 1.0, g_6 = 1.0, g_7 = 1.0\}$ and with $\{g_2 = 10.0, g_3 = 1.8, g_4 = 1.0, g_5 = 1.0, g_6 = 1.0, g_7 = 1.0\}$ the maximum discrepancy in the relative values of the free energies f_i was less than 2%. Therefore, differences in the PMF profiles were also negligible. This aspect of the method makes it particularly suitable for free-energy calculations using the Hamiltonian of eq. (1) even if the correlation times tend to vary with the coupling parameter λ; correlation times can be easily determined to within a factor of 2 or 3 (and certainly to within a factor of 10.0). The g_i reflect the weights assigned to each of the histograms; under conditions of biomolecular simulations where phase transitions do not occur, the ratio of the g_i should not differ significantly from one. The last column of Table I gives the approximate correlation time g_i for the simulations.

Qualitative behavior of systems under investigation can also be obtained from histogram techniques. We have used the WHAM equations to obtain the PMF of the pseudorotation phase angle at temperatures 350 and 250 K (Fig. 5); these PMFs were calculated for the case when λ is zero. It can

Table I. Summary of the simulations.

Simulation no.	n	λ_0	λ_1	T(K)	g
1[a]	20,000	1.00	0.00	298	4.0
2	20,000	1.00	0.20	298	4.0
3	20,000	1.00	0.40	298	1.7
4	20,500	1.00	0.50	298	1.1
5	45,000	1.00	1.00	350	1.0
6	48,750	1.00	1.20	350	1.0
7	52,500	1.00	1.40	298	1.1
8	45,000	1.0	1.40	250	1.5
9[a]	60,000	1.0	0.00	298	1.7
10	60,000	1.0	0.50	298	1.7

[a]The difference in the correlation times between runs 1 and 9 is due to the difference in the time step used in the Verlet algorithm[33] and to the difference in the archival rates of the snapshots.

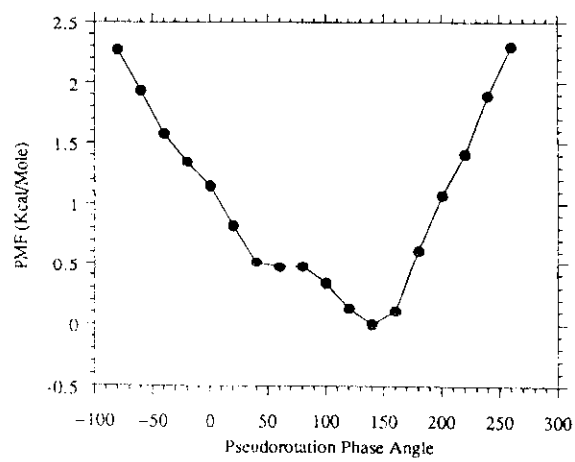


Figure 4. PMF of ξ at 298 K from all simulations.

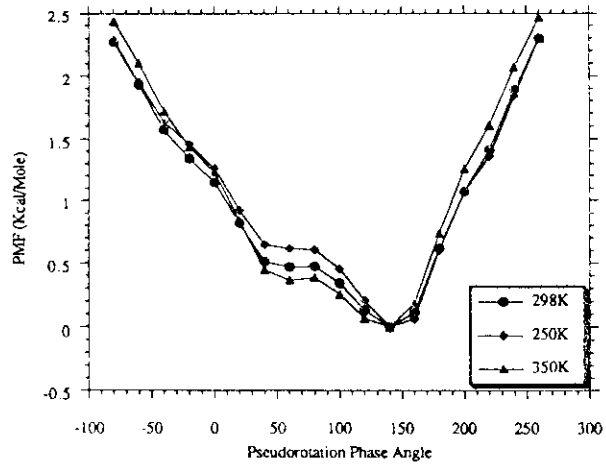


Figure 5. PMF profiles of the pseudorotation phase angle at 250, 298, and 350 K.

be seen that the PMF varies much more with temperature between $\xi = 18°$ and $\xi = 144°$ than around $\xi = -90°$ (the $O_{4'}$-exo region). This suggests that entropy contributes more to the PMF in the former region than the latter. That is expected since the energetics of the $O_{4'}$-exo ($\xi = -90°$) conformation is dominated by steric clash between the $C_{5'}$ hydrogen atoms and the base.[17] Kollman and coworkers[21] found by energy minimization that the difference in energy between the $C_{3'}$-endo and $C_{2'}$-endo region to be -0.6 kcal/mol. The difference in the PMF values obtained in this work is seen to tend toward this value as the temperature is lowered (Fig. 5). Kollman et al. also report an energy barrier of about 3.7 kcal/mol for the $C_{2'}$-endo to $C_{3'}$-endo transition via the $O_{4'}$-exo region. While the results of Kollman and coworkers are from energy minimizations keeping the sugar puckering amplitude[17-19] fixed, the results obtained here include entropic effects also. This is the main cause for the apparent discrepancies between the two results.

The PMF profile of the pseudorotation phase angle ξ depends on the size, nature, etc. of the molecule comprising the sugar ring. For instance, the $C_{2'}$-endo $\rightarrow$ $C_{3'}$-endo transition via the $O_{4'}$-exo region for sugar rings in the dodecamer CGCGAATTCGCG is greater than the barrier reported here by about 2 kcal/mol.[31] These will be reported in a future communication. Comparative studies between the WHAM, FEP, and Umbrella Sampling methods will also be reported in a future communication.

Initially only four simulations (1–4 in Table I) were carried out. However, to decrease the relative errors $\delta\Omega/\Omega$ [see eq. (22)] in the "outlying" bins ($\xi \approx 270°$) of the histograms six more simulations (5–10 in Table I) were carried out with increased values for the coupling constant λ. The error propagation from individual bins to the final PMF is not straightforward; however, one can look for errors by breaking all or some of the simulation runs into multiple runs and carrying out the WHAM calculations. We carried out a variety of such calculations and the resulting PMF of ξ was always found to be in agreement with Figure 4.

GENERAL COMPARISONS

As stated in the previous section we will report quantitative comparisons between the FEP, Umbrella Sampling, and WHAM techniques; nevertheless, with the experience to date on the WHAM method certain general comparisons between the methods can be made and will be outlined in this section.

The WHAM method is an extension of the Umbrella Sampling method but it has a number of advantages over the conventional Umbrella Sampling method. The WHAM method, in addition to optimizing the links between simulations, also allows multiple overlaps of probability distributions for obtaining better estimates of the free-energy differences. The older method of obtaining a single distribution function by requiring that the probability distributions agree at some point in the overlap region will fail to yield unique free-energies if three or more distributions are involved in the overlap region.[32] This algorithm provides a built-in estimate of errors that give investigators objective estimates of the optimal location and length of additional simulations to improve the accuracy of their results. With only two simulations, the WHAM method is still better than the conventional Umbrella Sampling, and actually reduces to Bennett's optimal solution for this special case.[33]

Umbrella Sampling methods that rely on eq. (9) cannot use the most general form of the restraining potential (and it is this special form with all but one of the λ_i set to zero that has generally been used so far by researchers). The WHAM method, however, can be used with the most general form of the restraining potential given in eq. (1); it lends itself particularly well to situations where the potential energy and/or the restraining potentials cannot be expressed as a direct function of the parameter(s) of interest. One of the limitations of the Umbrella Sampling method is in the determination of the value that the function $C(\{\lambda\}, \beta)$ so as to make $W_{\{0\},\beta}(\xi)$ agree in the regions of overlap[12-14]; the accuracy of $C(\{\lambda\}, \beta)$ is limited by the statistical errors in the distributions that are "stitched" together. To achieve the same level of accuracy conventional Umbrella Sampling would require much longer simulations than the WHAM method presented here. The WHAM method overcomes this

difficulty by taking into account all the simulations that produce overlapping distributions.

The calculation of free energies and the PMF of reaction or conformation coordinates using the FEP or the conventional Umbrella Sampling methods are computationally expensive. This is a consequence of the convergence problem associated with these computational techniques where many simulations have to be carried out as the Hamiltonian is gradually changed to propel the system along a certain coordinate. When using the FEP method $\Delta\lambda$ in eq. (14) or $(\lambda_{i+1} - \lambda_i)$ in eq. (13) has to be made small to assure convergence and to control the errors of discretization; moreover, errors propagate when connecting distributions at each step. The WHAM method is not a discretization. It uses multiple overlaps that do not have to be as close together as they have to be if the FEP method is used. The WHAM method links the different simulations through the overlapping histograms in an optimal manner. The FEP equations do not have a built-in estimate of errors, which makes it difficult for estimating statistical errors in the results, while the WHAM algorithm does provide for objective estimation of statistical errors [see eq. (22)].

The WHAM equations can also be readily used to generate PMFs and free energies as a function of the coupling parameter(s) λ_i and/or the temperature. This is useful as simulations can be carried out at a range of temperatures to improve conformational sampling and the results extrapolated (or interpolated) to the desired temperature.

This work was supported by grants from the National Institute of General Medicine, NIH (GM25671), the division of Advanced Scientific Computing of the National Science Foundation (ASC-9015310), and the Pittsburgh Supercomputing Center (DMB 90026P). The authors thank Yong Duan of the University of Pittsburgh for helpful suggestions and discussions.

APPENDIX: DERIVATION OF THE WHAM EQUATIONS

Consider R constant temperature simulations with the ith simulation being carried out at temperature T_i and with coupling parameters $\{\lambda\}_i$. Let the number of snapshots of the system taken from the ith simulation be n_i. The objective of the WHAM equations is then to obtain the best estimates of the probability density $P_{\{\lambda\},\beta}(\{V\}, \xi)$ at some $\{\lambda\}$ and β. The WHAM equations also yield the R free energies—$A_1, A_2, \ldots, A_R$—of the system associated with the R simulations.

An estimation of the generalized density of states from the kth simulation, $\Omega_k(\{V\}, \xi)$, can be written as

$$\Omega_k(\{V\}, \xi) = N_k(\{V\}, \xi) \exp\left[\left(\sum_{i=0}^{L} \beta_k \lambda_{i,k} V_i\right) - f_k\right] \quad (k = 1, 2, \ldots, R) \quad (23)$$

where $N_{\{\lambda\}_k,\beta_k}(\{V\}, \xi)$ has been shortened to $N_k(\{V\}, \xi)$ and $f_k = \beta_k A_k$. There will be R such estimates. The best value for the density of states, $\Omega(\{V\}, \xi)$ is written as a weighted sum of the R estimates $\Omega_i(\{V\}, \xi)$ $(i = 1, 2, \ldots, R)$, that is

$$\Omega(\{V\}, \xi) = \sum_{j=1}^{R} \omega_j(\{V\})\Omega_j(\{V\}, \xi) \quad (24)$$

subject to the condition

$$\sum_{j=1}^{R} \omega_j(\{V\}) = 1 \quad (25)$$

The set of ω_i that yield the best estimate of $\Omega(\{V\}, \xi)$ is derived by minimizing the statistical error, $\delta^2\Omega(\{V\}, \xi)$, in the best estimate of $\Omega(\{V\}, \xi)$. If the restraining potentials V_i are functions of the coordinate ξ, then the weights ω_i will depend on ξ through the restraining potentials. Now, the error, $\delta^2\Omega(\{V\}, \xi)$, arises out of the errors, $\delta^2\Omega_1(\{V\}, \xi)$, $\delta^2\Omega_2(\{V\}, \xi)$, $\delta^2\Omega_3(\{V\}, \xi)$, $\ldots$, $\delta^2\Omega_R(\{V\}, \xi)$, in the R estimates $\Omega_1(\{V\}, \xi)$, $\Omega_2(\{V\}, \xi)$, $\Omega_3(\{V\}, \xi)$, $\ldots$, $\Omega_R(\{V\}, \xi)$, which in turn depend upon the errors in the histograms, $\delta^2 N_1(\{V\}, \xi)$, $\delta^2 N_2(\{V\}, \xi)$, $\delta^2 N_3(\{V\}, \xi)$, $\ldots$, $\delta^2 N_R(\{V\}, \xi)$. Equations (26) and (27) summarize this:

$$\delta^2\Omega(\{V\}, \xi) = \sum_{j=1}^{R} \omega_j^2(\{V\})\delta^2\Omega_j(\{V\}, \xi) \quad (26)$$

and

$$\delta^2\Omega_k(\{V\}, \xi) = n_k^{-2} \exp\left[\left(2\sum_{i=0}^{L} \beta_k \lambda_{i,k} V_i\right) - 2f_k\right] \times \delta^2 N_k(\{V\}, \xi) \quad (k = 1, 2, \ldots, R) \quad (27)$$

Following Ferrenberg and Swendsen,[1-3] the error in $N_i(\{V\}, \xi)$ is written as

$$\delta^2 N_i(\{V\}, \xi) = g_i \overline{N_i(\{V\}, \xi)} \quad (i = 1, 2, \ldots, R) \quad (28)$$

where the bar indicates the expectation value with respect to all simulations of length n_i and $g_i = 1 + 2\tau_i$ where τ_i is the integrated correlation time of the ith simulation.[29] It should be noted that for biomolecular systems the g_i $(i = 1, 2, \ldots, R)$ are roughly equal to each other and hence cancel each other out of the WHAM equations.

We now make an estimate of the $\overline{N_i(\{V\}, \xi)}$ as follows:

$$\overline{N_i(\{V\}, \xi)} = n_i\Omega(\{V\}, \xi) \exp\left(f_i - \beta_i \sum_{j=0}^{L} \lambda_{j,i} V_j\right) \quad (i = 1, 2, \ldots, R) \quad (29)$$

From eqs. (26), (27), (28), and (29) we can obtain an expression for $\delta^2\Omega(\{V\}, \xi)$. The error is then minimized by setting the partial derivatives $\partial[\delta^2\Omega(\{V\}, \xi)]/\partial\omega_i (i = 1, 2, \ldots, R)$ equal to zero subject to eq. (25). From the resulting expression the WHAM equations

$$P_{\{\lambda\},\beta}(\{V\}, \xi) = \frac{\sum_{k=1}^{R} g_k^{-1} N_k(\{V\}, \xi) \exp\left(-\beta \sum_{j=0}^{L} \lambda_j V_j\right)}{\sum_{m=1}^{R} n_m g_m^{-1} \exp\left(f_m - \beta_m \sum_{j=0}^{L} \lambda_{j,m} V_j\right)} \quad (30)$$

and

$$\exp(-f_j) = \sum_{\{V\},\xi} P_{\{\lambda\}_j,\beta_j}(\{V\}, \xi) \quad (31)$$

can be derived. The density of state $\Omega(\{V\}, \xi)$ can also be determined by setting $\partial[\delta^2\Omega(\{V\}, \xi)]/\partial\omega_i$ to zero and is given by

$$\Omega(\{V\}, \xi) = \frac{\sum_{k=1}^{R} g_k^{-1} N_k(\{V\}, \xi)}{\sum_{m=1}^{R} n_m g_m^{-1} \exp\left(f_m - \beta_m \sum_{j=0}^{L} \lambda_{j,m} V_j\right)} \quad (32)$$

By inserting the expression for $\Omega(\{V\}, \xi)$ into the expression for $\delta^2\Omega(\{V\}, \xi)$ the relative error in $\delta\Omega/\Omega$ can be determined to be

$$\frac{\partial\Omega(\{V\}, \xi)}{\Omega(\{V\}, \xi)} = \left[\sum_{k=1}^{R} g_k^{-1} N_k(\{V\}, \xi)\right]^{-1/2} \quad (33)$$

When the restraining potential is a function of the coordinate ξ only the dimensionality of the histograms reduces from $L + 2$ to 2 and eqs. (30) and (31) simplify to

$$P_{\{\lambda\},\beta}(V_0, \xi) = \frac{\sum_{k=1}^{R} g_k^{-1} N_k(V_0, \xi) \exp\left[-\beta\lambda_0 V_0 - \beta \sum_{j=1}^{L} \lambda_j \hat{V}_j(\xi)\right]}{\sum_{m=1}^{R} n_m g_m^{-1} \exp\left[f_m - \beta_m\lambda_0 V_0 - \beta_m \sum_{j=1}^{L} \lambda_{j,m} \hat{V}_j(\xi)\right]} \quad (34)$$

and

$$\exp(-f_j) = \sum_{V_0,\xi} P_{\{\lambda\}_j,\beta_j}(V_0, \xi) \quad (35)$$

The WHAM equations can easily be generalized to situations where the objective is to generate multidimensional PMF profiles of multiple reaction coordinates.

References

1. A.M. Ferrenberg, PhD Thesis, Carnegie Mellon University, Pittsburgh, PA, 1989.
2. A.M. Ferrenberg and R.H. Swendsen, *Phys. Rev. Lett.*, **61**, 2635 (1988).
3. A.M. Ferrenberg and R.H. Swendsen, *Phys. Rev. Lett.*, **63**, 1195 (1989).
4. D. Bouzida, S. Kumar, and R.H. Swendsen, in *Computer Simulation Studies in Condensed Matter Physics III, Springer Proceedings in Physics 53*, D.P. Landau, K.K. Mon, and H.-B. Schuttler, Eds., Springer-Verlag, Berlin, 1991.
5. R.H. Swendsen, D. Bouzida, and S. Kumar, Almost Markov Processes in Monte Carlo Simulation of Biological Molecules, Technical Report 91-121-NAMS-27, Carnegie Mellon University, Pittsburgh, PA, 1991.
6. D. Bouzida, S. Kumar, and R.H. Swendsen, *Phys. Rev. A* (to appear).
7. D.A. McQuarrie, *Statistical Mechanics*, Harper & Row, New York, 1976.
8. J.A. McCammon and S.C. Harvey, *Dynamics of Proteins and Nucleic Acids*, Cambridge University Press, Cambridge, UK, 1989.
9. C.L. Brooks, M. Karplus, and M. Pettitt, *Proteins: A Theoretical Perspective of Dynamics, Structure and Thermodynamics, Advances in Chemical Physics*, vol. **LXXI**, John Wiley & Sons, New York, 1989.
10. P.A. Bash, U.C. Singh, R. Langridge, and P.A. Kollman, *Science*, **236**, 564 (1987).
11. S.H. Northrup, M.R. Pear, C.-Y. Lee, J.A. McCammon, and J. Karplus, *Proc. Natl. Acad. Sci. USA*, **79**, 4035 (1982).
12. C. Pangali, M. Rao, and B.J. Berne, *J. Chem. Phys.*, **71**, 2975 (1979).
13. G.N. Patey and J.P. Valleau, *J. Chem. Phys.*, **63**, 2334 (1975).
14. J.P. Valleau and D.N. Card, *J. Chem. Phys.*, **57**, 5457 (1972).
15. M. Mezei, P.K. Mehrotra, and D.L. Beveridge, *J. Am. Chem. Soc.*, **107**, 2239 (1985).
16. D.J. Tobias, C.L. Brooks III, *Chem. Phys. Lett.*, **142**, 472 (1987); D.J. Tobias, C.L. Brooks III, *J. Chem. Phys.*, **89**, 5115 (1988).
17. W. Saenger, *Principles of Nucleic Acid Structure*, Springer-Verlag, New York, 1984.
18. C. Altona and M. Sundaralingam, *Tetrahedron*, **24**, 13 (1968).
19. D. Cremer and J.A. Pople, *J. Am. Chem. Soc.*, **97**, 1354 (1975).
20. P.K. Weiner and P.A. Kollman, *J. Comp. Chem.*, **2**, 287 (1980).
21. S.J. Weiner, P.A. Kollman, D.T. Nguyen, and D.A. Case, *J. Comp. Chem.*, **7**, 230 (1986).
22. H. Goldstein, *Classical Mechanics*, Addison-Wesley, Reading, MA, 1980.
23. K.K. Irikura, B. Tidor, B.R. Brooks, and M. Karplus, *Science*, **229**, 571 (1985).
24. Z.W. Salsburg, J.D. Jacobson, W. Fickett, and W.W. Wood, *J. Chem. Phys.*, **30**, 65 (1959).
25. H. Muller-Krumbhaar and K. Binder, *J. Stat. Phys.*, **8**, 1 (1973).
26. J.-P. Ryckaert, G. Ciccotti, and H.J.C. Berendsen, *J. Comp. Phys.*, **23**, 327 (1977).
27. W.H. Press, B.P. Flannery, S.A. Teukolsky, and W.T. Vetterling, *Numerical Recipes. The Art of Scientific Computing*, Cambridge University Press, New York, 1986.
28. D.H.J. Mackay, A.J. Cross, and A.T. Hagler, in *Prediction of Protein Structure and the Principles of Protein Conformation*, G.D. Fasman, Ed., Plenum Press, New York, 1989.
29. R.W. Hockney and J.W. Eastwood, *Computer Simu-*

lation Using Particles, Adam Hilger, Bristol, UK, 1988.
30. H.J.C. Berendsen, J.P.M. Postma, W.F. van Gunsteren, A. DiNola, and J.R. Haak, *J. Chem. Phys.*, **81**, 3684 (1984).
31. S. Kumar, PhD thesis, University of Pittsburgh, Pittsburgh, PA, 1990.
32. G.M. Torrie and J.P. Valleau, *Chem. Phys. Lett.*, **28**, 578 (1974).
33. C.H. Bennett, *J. Comp. Phys.*, **22**, 245 (1976).