CHAPTER 1

Computational Techniques and Strategies for Monte Carlo Thermodynamic Calculations, with Applications to Nanoclusters

Robert Q. Topper,[*] David L. Freeman,[†] Denise Bergin,[*] and Keirnan R. LaMarche[*]

[*]Department of Chemistry, The Cooper Union for the Advancement of Science and Art, 51 Astor Place, New York, New York 10003,[**] and [†]Department of Chemistry, University of Rhode Island, Kingston, Rhode Island 02881

[**]Present address: Department of Chemistry, Medical Technology, and Physics, Monmouth University, West Long Branch, New Jersey

INTRODUCTION

This chapter is written for the reader who would like to learn how Monte Carlo methods[1] are used to calculate thermodynamic properties of systems at the atomic level, or to determine which advanced Monte Carlo methods might work best in their particular application. There are a number of excellent books and review articles on Monte Carlo methods, which are generally focused on condensed phases, biomolecules or electronic structure theory.[2–13] The purpose of this chapter is to explain and illustrate some of the special techniques that we and our colleagues have found to be particularly

Reviews in Computational Chemistry, Volume 19
edited by Kenny B. Lipkowitz, Raima Larter, and Thomas R. Cundari
ISBN 0-471-23585-7

well suited for simulations of nanodimensional atomic and molecular clusters. We want to help scientists and engineers who are doing their first work in this area to get off on the right foot, and also provide a pedagogical chapter for those who are doing experimental work. By including examples of simulations of some simple, yet representative systems, we provide the reader with some data for direct comparison when writing their own code from scratch.

Although a number of Monte Carlo methods in current use will be reviewed, this chapter is not meant to be comprehensive in scope. Monte Carlo is a remarkably flexible class of numerical methods. So many versions of the basic algorithms have arisen that we believe a comprehensive review would be of limited pedagogical value. Instead, we intend to provide our readers with enough information and background to allow them to navigate successfully through the many different Monte Carlo techniques in the literature. This should help our readers use existing Monte Carlo codes knowledgably, adapt existing codes to their own purposes, or even write their own programs. We also provide a few general recommendations and guidelines for those who are just getting started with Monte Carlo methods in teaching or in research.

This chapter has been written with the goal of describing methods that are generally useful. However, many of our discussions focus on applications to atomic and molecular clusters (nanodimensional aggregates of a finite number of atoms and/or molecules).[14] We do this for two reasons:

1. A great deal of our own research has focused on such systems,[15] particularly the phase transitions and other structural transformations induced by changes in a cluster's temperature and size, keeping an eye on how various properties approach their bulk limits. The precise determination of thermodynamic properties (such as the heat capacity) of a cluster type as a function of temperature and size presents challenges that must be addressed when using Monte Carlo methods to study virtually any system. For example, analogous structural transitions can also occur in phenomena as disparate as the denaturation of proteins.[16,17] The modeling of these transitions presents similar computational challenges to those encountered in cluster studies.

2. Although cluster systems can present some unique challenges, their study is unencumbered by many of the technical issues regarding periodic boundary conditions that arise when solids, liquids, surface adsorbates, and solvated biomolecules and polymers are studied. These issues are addressed well elsewhere,[7,11,12] and can be thoroughly appreciated and mastered once a general background in Monte Carlo methods is obtained from this chapter.

It should be noted that "Monte Carlo" is a term used in many fields of science, engineering, statistics, and mathematics to mean entirely different things. The one (and only) thing that all Monte Carlo methods have in common is that they all use random numbers to help calculate something. What we mean by "Monte Carlo" in this chapter is the use of random-walk processes to draw samples from a desired probability function, thereby allowing one to

calculate integrals of the form $\int dq f(q)\,\rho(q)$. The quantity $\rho(q)$ is a normalized probability density function that spans the space of a many-dimensional variable q, and $f(q)$ is a function whose average is of thermodynamic importance and interest. This integral, as well as all other integrals in this chapter, should be understood to be a definite integral that spans the entire domain of q. Finally, we note that the inclusion of quantum effects through path-integral Monte Carlo methods is not discussed in this chapter. The reader interested in including quantum effects in Monte Carlo thermodynamic calculations is referred elsewhere.[15,18–22]

METROPOLIS MONTE CARLO

Monte Carlo simulations are widely used in the fields of chemistry, biology, physics, and engineering in order to determine the structural and thermodynamic properties of complex systems at the atomic level. Thermodynamic averages of molecular properties can be determined from Monte Carlo methods, as can minimum-energy structures. Let $\langle f \rangle$ represent the average value of some coordinate-dependent property $f(\boldsymbol{x})$, with $\boldsymbol{x}$ representing the $3N$ Cartesian coordinates needed to locate all of the N atoms. In the canonical ensemble (fixed N, V and T, with V the volume and T the absolute temperature), averages of molecular properties are given by an average of $f(\boldsymbol{x})$ over the Boltzmann distribution

$$\langle f \rangle = \frac{\int d\boldsymbol{x}\, f(\boldsymbol{x}) \exp[-\beta U(\boldsymbol{x})]}{\int d\boldsymbol{x} \exp[-\beta U(\boldsymbol{x})]} \qquad [1]$$

where $U(\boldsymbol{x})$ is the potential energy of the system, $\beta = 1/k_B T$, and k_B is the Boltzmann constant.[23] If one can compute the thermodynamic average of $f(\boldsymbol{x})$ it is then possible to calculate various thermodynamic properties. In the canonical ensemble it is most common to calculate E, the internal energy, and C_V, the constant-volume heat capacity (although other properties can be calculated as well). For example, if we average $U(\boldsymbol{x})$ over all possible configurations according to Eq. [1], then E and C_V are given by

$$E = \frac{3Nk_B T}{2} + \langle U \rangle \qquad [2]$$

$$C_V = \frac{3Nk_B}{2} + \frac{\langle U^2 \rangle - \langle U \rangle^2}{(k_B T^2)} \qquad [3]$$

The first term in each equation represents the contribution of kinetic energy, which is analytically integrable. In the harmonic (low-temperature) limit, E given by Eq. [2] will be a linear function of temperature and C_V from Eq. [3] will be constant, in accordance with the Equipartition Theorem.[10] For a small cluster of, say, 6 atoms, the integrals implicit in the calculation of Eqs. [1]

and [2] are already of such high dimension that they cannot be effectively computed using Simpson's rule or other basic quadrature methods.[2,24,25,26] For larger clusters, liquids, polymers or biological molecules the dimensionality is obviously much higher, and one typically resorts to either Monte Carlo, molecular dynamics, or other related algorithms.

To calculate the desired thermodynamic averages, it is necessary to have some method available for computation of the potential energy, either explicitly (in the form of a function representing the interaction potential as in molecular mechanics) or implicitly (in the form of direct quantum-mechanical calculations). Throughout this chapter we shall assume that U is known or can be computed as needed, although this computation is typically the most computationally expensive part of the procedure (because U may need to be computed many, many times). For this reason, all possible measures should be taken to assure the maximum efficiency of the method used in the computation of U.

Also, it should be noted that constraining potentials (which keep the cluster components from straying too far from a cluster's center of mass) are sometimes used.[27] At finite temperature, clusters have finite vapor pressures, and particular cluster sizes are typically unstable to evaporation. Introducing a constraining potential enables one to define clusters of desired sizes. Because the constraining potential is artificial, the dependence of calculated thermodynamic properties on the form and the radius of the constraining potential must be investigated on a case-by-case basis. Rather than diverting the discussion from our main focus (Monte Carlo methods), we refer the interested reader elsewhere for more details and references on the use of constraining potentials.[15,19]

Random-Number Generation: A Few Notes

Because generalized Metropolis Monte Carlo methods are based on "random" sampling from probability distribution functions, it is necessary to use a high-quality random-number generator algorithm to obtain reliable results. A review of such methods is beyond the scope of this chapter,[24,28] but a few general considerations merit discussion.

Random-number generators do not actually produce random numbers. Rather, they use an integer "seed" to initialize a particular "pseudorandom" sequence of real numbers that, taken as a group, have properties that leave them nearly indistinguishable from truly random numbers. These are conventionally floating-point numbers, distributed uniformly on the interval (0,1). If there is a correlation between seeds, a correlation may be introduced between the pseudorandom numbers produced by a particular generator. Thus, the generator should ideally be initialized only once (at the beginning of the random walk), and not re-initialized during the course of the walk. The seed should be supplied either by the user or generated arbitrarily by the program

using, say, the number of seconds since midnight (or some other arcane formula). One should be cautious about using the "built-in" random-number generator functions that come with a compiler for Monte Carlo integration work because some of them are known to be of very poor quality.[26] The reader should always be sure to consult the appropriate literature and obtain (and test) a high-quality random-number generator before attempting to write and debug a Monte Carlo program.

The Generalized Metropolis Monte Carlo Algorithm

The Metropolis Monte Carlo (MMC) algorithm is the single most widely used method for computing thermodynamic averages. It was originally developed by Metropolis et al. and used by them to simulate the freezing transition for a two-dimensional hard-sphere fluid.[1] However, Monte Carlo methods can be used to estimate the values of multidimensional integrals in whatever context they may arise.[29,30] Although Metropolis et al. did not present their algorithm as a general-utility method for numerical integration, it soon became apparent that it could be generalized and applied to a variety of situations. The core of the MMC algorithm is the way in which it draws samples from a desired probability distribution function. The basic strategies used in MMC can be generalized so as to apply to many kinds of probability functions and in combination with many kinds of sampling strategies. Some authors refer to the generalized MMC algorithm simply as "Metropolis sampling,"[31] while others have referred to it as the $M(RT)^2$ method[6] in honor of the five authors of the original paper (Metropolis, the Rosenbluths, and the Tellers).[1] We choose to call this the generalized Metropolis Monte Carlo (gMMC) method, and we will always use the term MMC to refer strictly to the combination of methods originally presented by Metropolis et al.[1]

In the literature of numerical analysis, gMMC is classified as an importance sampling technique.[6,24] Importance sampling methods generate configurations that are distributed according to a desired probability function rather than simply picking them at random from a uniform distribution. The probability function is chosen so as to obtain improved convergence of the properties of interest. gMMC is a special type of importance sampling method which asymptotically (i.e., in the limit that the number of configurations becomes large) generates states of a system according to the desired probability distribution.[6,9] This probability function is usually (but not always[6]) the actual probability distribution function for the physical system of interest. Nearly all statistical-mechanical applications of Monte Carlo techniques require the use of importance sampling, whether gMMC or another method is used (alternatively, "stratified sampling" is sometimes an effective approach[22,32]). gMMC is certainly the most widely used importance sampling method.

In the gMMC algorithm successive configurations of the system are generated to build up a special kind of random walk called a *Markov*

chain.[29,33,34] The random walk visits successive configurations, where each configuration's location depends on the configuration immediately preceding it in the chain. The gMMC algorithm establishes how this can be done so as to asymptotically generate a distribution of configurations corresponding to the probability density function of interest, which we denote as $\rho(\boldsymbol{q})$.

We define $K(\boldsymbol{q}_i \rightarrow \boldsymbol{q}_j)$ to be the conditional probability that a configuration at $\boldsymbol{q}_i$ will be brought to $\boldsymbol{q}_j$ in the next step of the random walk. This conditional probability is sometimes called the "transition rate." The probability of moving from $\boldsymbol{q}$ to $\boldsymbol{q}'$ (where $\boldsymbol{q}$ and $\boldsymbol{q}'$ are arbitrarily chosen configurations somewhere in the available domain) is therefore given by $P(\boldsymbol{q} \rightarrow \boldsymbol{q}')$:

$$P(\boldsymbol{q} \rightarrow \boldsymbol{q}') = K(\boldsymbol{q} \rightarrow \boldsymbol{q}')\,\rho(\boldsymbol{q}) \qquad [4]$$

For the system to evolve toward a unique limiting distribution, we must place a constraint on $P(\boldsymbol{q} \rightarrow \boldsymbol{q}')$. The gMMC algorithm achieves the desired limiting behavior by requiring that, on the average, a point is just as likely to move from $\boldsymbol{q}$ to $\boldsymbol{q}'$ as it is to move in the reverse direction, namely, that $P(\boldsymbol{q} \rightarrow \boldsymbol{q}') = P(\boldsymbol{q}' \rightarrow \boldsymbol{q})$. This likelihood can be achieved only if the walk is ergodic (an ergodic walk eventually visits all configurations when started from any given configuration) and if it is aperiodic (a situation in which no single number of steps will generate a return to the initial configuration). This latter requirement is known as the "detailed balance" or the "microscopic reversibility" condition:

$$K(\boldsymbol{q} \rightarrow \boldsymbol{q}')\,\rho(\boldsymbol{q}) = K(\boldsymbol{q}' \rightarrow \boldsymbol{q})\,\rho(\boldsymbol{q}') \qquad [5]$$

Satisfying the detailed balance condition ensures that the configurations generated by the gMMC algorithm will asymptotically be distributed according to $\rho(\boldsymbol{q})$.

The transition rate may be written as a product of a trial probability Π and an acceptance probability A

$$K(\boldsymbol{q}_i \rightarrow \boldsymbol{q}_j) = \Pi(\boldsymbol{q}_i \rightarrow \boldsymbol{q}_j)A(\boldsymbol{q}_i \rightarrow \boldsymbol{q}_j) \qquad [6]$$

where Π can be taken to be any normalized distribution that asymptotically spans the space of all possible configurations, and A is constructed so that Eq. [5] is satisfied for a particular choice of Π. The wonderful flexibility with which Π can be chosen is one of the reasons so many Monte Carlo methods are found in the literature. For example, Metropolis et al. used a uniform distribution of points about $\boldsymbol{x}_i$ to define the trial probability (we describe this in greater detail in the next section),[1] but a Gaussian distribution of points can also be used profitably in certain situations.[20] Many other distributions are possible and even desirable in different contexts.

From the detailed balance condition, it is straightforward to show that the ratio of acceptance probabilities, given by r, is

$$r = \frac{A(\boldsymbol{q}_i \rightarrow \boldsymbol{q}_j)}{A(\boldsymbol{q}_j \rightarrow \boldsymbol{q}_i)} = \frac{\Pi(\boldsymbol{q}_j \rightarrow \boldsymbol{q}_i)\,\rho(\boldsymbol{q}_j)}{\Pi(\boldsymbol{q}_i \rightarrow \boldsymbol{q}_j)\,\rho(\boldsymbol{q}_i)} \qquad [7]$$

where $r \geq 0$. From this, it can be seen that a rejection method can be used to effectively define the acceptance probability, A:

$$A(\boldsymbol{q}_i \rightarrow \boldsymbol{q}_j) = \min(1, r) \qquad [8]$$

Equation [8] is the heart of the gMMC algorithm.

In their original paper, Metropolis et al. considered systems represented within the canonical ensemble, for which the density is given by

$$\rho(\boldsymbol{q}_j) = \frac{\exp\left[-\beta U(\boldsymbol{q}_j)\right] J(\boldsymbol{q})}{Z(N, V, T)} \qquad [9]$$

where $Z(N,V,T)$ is the configuration integral, given by

$$Z(N, V, T) = \int d\boldsymbol{q}\, \exp[-\beta U(\boldsymbol{q})]\, J(\boldsymbol{q}) \qquad [10]$$

and $J(\boldsymbol{q})$ is the determinant of the Jacobian matrix defining a canonical transformation from Cartesian coordinates $\boldsymbol{x}$ to arbitrary coordinates $\boldsymbol{q}$ ($J(\boldsymbol{q}) = 1$ if Cartesian coordinates are used in the walk).[35] $J(\boldsymbol{q})$ must generally be included because all of the statistical-mechanical integrations are performed in Cartesian coordinates $\boldsymbol{x}$, in which the kinetic energy matrix is diagonal, and not in the arbitrary coordinates $\boldsymbol{q}$.[36] The ratio then becomes

$$r = \frac{\Pi(\boldsymbol{q}_j \rightarrow \boldsymbol{q}_i)}{\Pi(\boldsymbol{q}_i \rightarrow \boldsymbol{q}_j)} \exp\{-\beta \Delta U\} \frac{J(\boldsymbol{q}_j)}{J(\boldsymbol{q}_i)} \qquad [11]$$

where $\Delta U = U(\boldsymbol{q}_j) - U(\boldsymbol{q}_i)$. If Cartesian coordinates $\boldsymbol{x}$ are used in the walk, then $J = 1$ and we need not evaluate the Jacobian at each step to determine the acceptance probability. Metropolis et al. further chose the trial probability to be uniform so that the ratio of the trial probabilities Π cancel in the numerator and denominator of Eq. (11). Then the acceptance probability is simply given by

$$A(x_i \rightarrow x_j) = \min\{1, \exp\{-\beta \Delta U\}\} \qquad [12]$$

The implementation of this method is described in detail in the following section.

Metropolis Monte Carlo: The "Classic" Algorithm

Having established the preceding important framework, we now turn to the particulars of the original Metropolis Monte Carlo algorithm for sampling configurations from the canonical ensemble. As alluded to in the previous section, it is often most convenient to work in Cartesian coordinates for MMC calculations (with some exceptions discussed later). In this case, Eqs. [6]–[12] reduce to the algorithm described by the flowchart in Figure 1. First, an initial configuration of the system is established. The initial configuration can be generated randomly, although it is sometimes advantageous to start from an energy-minimized structure,[37] a crystalline lattice structure, or a structure obtained from experiment.

Next, a "trial move" is made to generate a new trial configuration, according to a rule that we call a "move strategy." The simple move strategy introduced by Metropolis et al.[1] is still the most widely used method. One

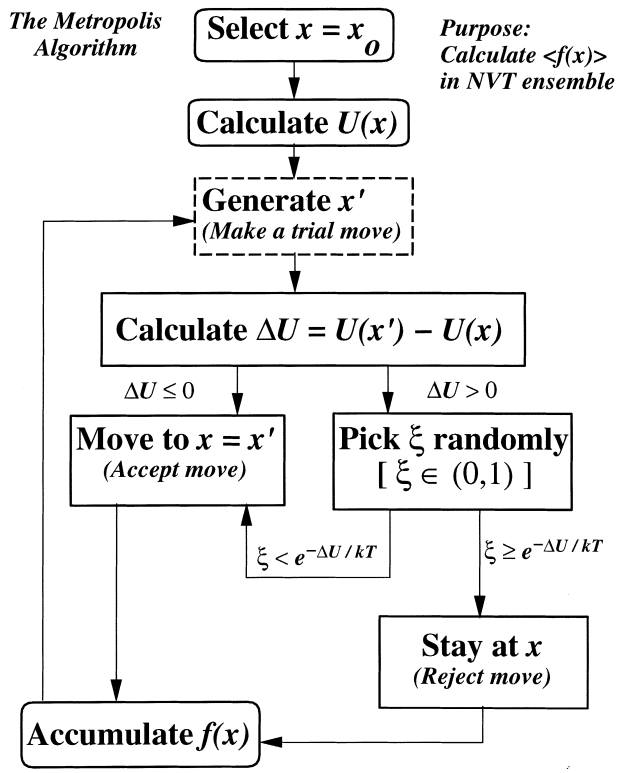

Figure 1 Flowchart of the "classic" Metropolis Monte Carlo algorithm for sampling in the canonical ensemble.[1] Note that samples of the property function $f(x)$ are always accumulated for averaging purposes, irrespective of whether a move is accepted or rejected.

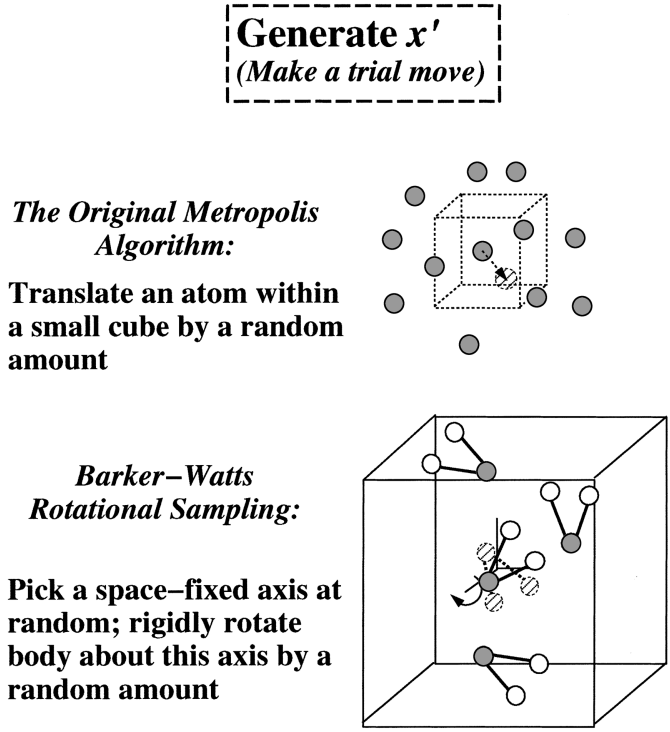

Figure 2 Single-particle Metropolis moves from the original MMC algorithm[1] and molecular rotation Barker–Watts moves[41] for generation of a trial Monte Carlo move.

preselects a maximum stepsize L and randomly moves each particle within a cube of length L centered on the atom's original position (see Figure 2). This procedure defines the Metropolis transition probability, Π_M. For a one-dimensional system moving along a single coordinate x:

$$\begin{aligned} \Pi_M &= \frac{1}{L}, \quad \frac{-L}{2} < x < \frac{L}{2} \\ &= 0 \qquad \text{elsewhere} \end{aligned} \tag{13}$$

The parameter L may have an optimum value for the particular type of atom of interest, as well as for the temperature and other variables studied. If L is too small, most moves will be accepted and a very large number of attempts will be required to move very far from the initial configuration. However, if L is too big very few trial moves will be accepted and again, the walker will require many steps to move away from the starting point. For this reason one generally chooses L so that between 30% and 70% of the moves are accepted (50% is a happy medium).[6,7] Each atom can be moved in sequence,

or an atom can be chosen randomly for each trial move, at the discretion of the programmer.

As noted previously, the Metropolis move strategy of "single-particle moves" is still the most widely used in the literature. The use of single-particle moves is often considered as being part and parcel of the Metropolis Monte Carlo method. However, a number of other move strategies in the canonical ensemble are possible, some of which are outlined later in this chapter. Metropolis Monte Carlo simulations in other ensembles require the use of other move strategies. For example, in the isothermal–isobaric ensemble (constant N,P,T), the volume and the configurations are perturbed.[7,13,38,39] In the grand canonical ensemble (constant chemical potential, V, and T) the number of particles N fluctuates, so a move may include randomly deleting or adding a particle.[11,13,39] The use of the Gibbs ensemble for phase equilibrium studies involves perturbations of volumes and configurations within each phase, as well as the transfer of particles between phases.[10,13,39,40] In all of these situations, suitable move strategies must be employed to ensure that detailed balance is satisfied for all variables involved.

Regardless how it is generated, the trial configuration does not automatically become the second step in the Markov chain. Within the canonical ensemble if the potential energy of the trial configuration is less than or equal to the potential energy of the previous configuration, that is, if $\Delta U \equiv U(\boldsymbol{x}') - U(\boldsymbol{x}) \leq 0$, the trial configuration is then "accepted." However, if $\Delta U > 0$ the trial move may still be conditionally accepted. A random number ζ between 0 and 1 is chosen and compared to $\exp(-\beta\Delta U)$. If ζ is less than or equal to $\exp(-\beta\Delta U)$ the trial move is accepted; otherwise, the trial configuration fails the "Boltzmann test," the move is "rejected," and the original configuration becomes the second step in the Markov chain (see Figure 1). The procedure is then repeated many times until equilibration is achieved (equilibration is defined in a later section). After equilibration, the procedure continues as before, but now the values of U and U^2 (and any other properties of interest, represented by f in Figure 1) are accumulated for each of the remaining n steps in the chain. Let n represent the number of samples used in computing the averages of U and U^2. The number of samples is chosen to be sufficiently large for convergence of $\langle U\rangle_n$ and $\langle U^2\rangle_n$, the n-point averages of U and U^2, to their true values $\langle U\rangle$ and $\langle U^2\rangle$. In the case of U,

$$\langle U\rangle = \lim_{n\to\infty} \langle U\rangle_n \qquad [14]$$

where

$$\langle U\rangle_n \equiv \frac{1}{n}\sum_{j=1}^{n} U(\boldsymbol{x}_j) \qquad [15]$$

Similar formulas hold for $\langle U^2 \rangle$ and $\langle U^2 \rangle_n$. It is important to emphasize that configurations are not excluded for averaging purposes simply because they have been most recently accepted or rejected. Both accepted and rejected configurations must be included in the average, or the potential energies will not be Boltzmann-distributed.

The Barker–Watts Algorithm for Molecular Rotations

The simplest way to generate trial configurations is to use single-particle moves, that is, to change the Cartesian coordinates of individual atoms according to the original MMC procedure presented by Metropolis et al. described in the preceding section.[1] For atomic clusters and liquids, this may be sufficient. However, single-particle moves alone are inefficient for molecular cluster systems, specifically clusters consisting of some finite number of molecules. For such systems it can be useful to also carry out Metropolis moves of each molecule's center of mass and to rotate each molecule within a cluster so as to efficiently generate all possible relative orientations. The Barker–Watts algorithm presented below is an example of such a "generalized Metropolis Monte Carlo" method. It is not limited to cluster systems, but is appropriate for any system for which rotational moves can be useful. Barker and Watts originally developed it for use in simulations of liquid water.[41]

In the Barker–Watts algorithm, a particular molecule is chosen either at random or systematically. One of the three Cartesian axes is selected at random and the molecule is rigidly rotated about the axis by a randomly chosen angle $\Delta\theta$ where $[-\Delta\theta_{MAX} \leq \Delta\theta \leq \Delta\theta_{MAX}]$, as shown in Figure 2. The Cartesian coordinates of each atom within the molecule after the trial move (x', y', z') are calculated from $\Delta\theta$ and the coordinates of that atom in the current configuration (x, y, z). For a rotation about the x axis, for example, the new coordinates would be given by

$$\begin{pmatrix} x' \\ y' \\ z' \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & \cos\Delta\theta & \sin\Delta\theta \\ 0 & -\sin\Delta\theta & \cos\Delta\theta \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix} \qquad [16]$$

with a similar expression for rotation about the y axis or the z axis. The new configuration's potential energy is calculated, and the move is accepted or rejected according to the Boltzmann probability. The transition probability is uniform in the chosen coordinate system for each trial move, in the same spirit as the original MMC algorithm.

Equilibration: Why Wait?

As a practical matter, one cannot immediately start accumulating U and U^2 for the computation of averages. The MMC algorithm does not instantly

sample configurations according to the Boltzmann distribution; the sampling is only guaranteed to be correct in the asymptotic limit of large n. It is therefore always necessary to allow the Markov walk to go through a large number of steps before beginning to accumulate samples to estimate U and U^2. This procedure is known as the "equilibration" of the walker.[6,7,10] The walker should undergo a sufficiently large number of iterations for there to be no "memory" of the system's initial configuration. The index value $j = 1$ in Eq. [15] refers to the first configuration after the system has equilibrated by cycling through n_{eq} steps; only states sampled after equilibration should be included in the average. n_{eq} is often referred to as the "equilibration period."

One way to estimate n_{eq} is to calculate the running averages of U and U^2 to make running estimates of E and C_V; these quantities are then plotted against m, the subtotal of all steps taken ($m < n$) at a few representative temperatures to determine when asymptotic, slowly varying behavior is attained.[7] Since E is given by Eq. [3], the running average of U may be plotted instead of E if desired. The running average of U is given by $\langle U \rangle_m$

$$\langle U \rangle_m \equiv \frac{1}{m} \sum_{j=1}^{m} U(\boldsymbol{x}_j) \qquad [17]$$

with a similar formula for $\langle U^2 \rangle_m$. These averages can then be used to form the running estimate of C_V. Alternatively, some authors have advocated the use of crystalline order parameters for this purpose in fluid simulations (but this is not generally practical for cluster simulations).[12] An example of a running-average equilibration study is shown in Figure 3 for a one-dimensional Lennard-Jones oscillator. The Lennard-Jones potential is given by

$$U(r) = 4\varepsilon \left[\left(\frac{\sigma}{r} \right)^{12} - \left(\frac{\sigma}{r} \right)^{6} \right] \qquad [18]$$

where r is the interparticle distance (varied uniformly during the walk according to Eq. [13]) and ε and σ are constants chosen appropriately to represent a particular system of interest (here we have chosen $\varepsilon/k_B = 124$ K and $\sigma = 0.3418$ nm, which would be appropriate for liquid argon[7]). Very low temperatures were considered and all calculations were initiated near the potential energy minimum, so it was not necessary to reject moves resulting in large interatomic separations. An imposed cutoff radius would be needed if higher temperatures were studied.[15,19]

In these plots the thermodynamic quantities go through some initial transient behavior, and then eventually settle down into small-amplitude oscillations. At this very low temperature both U and C_V settle down rapidly and they do so on similar "timescales." Typically, the running averages of U and C_V will not converge simultaneously. In fact C_V will usually be the slower of the two to converge, since its fluctuations arise from fluctuations of the

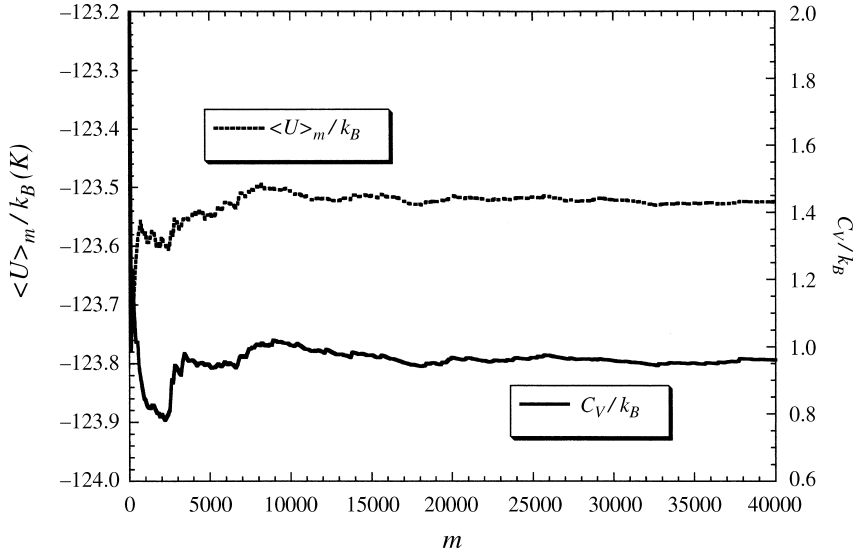

Figure 3 Cumulative values of $\langle U \rangle / k_B$ (dashed line) and C_V/k_B (solid line) for a Monte Carlo simulation of the one-dimensional Lennard-Jones oscillator (see text) at $T = 1$ K. Note that although C_V typically takes longer to equilibrate than does $\langle U \rangle$, there is no significant difference between the two for this simple system at this very low temperature.

potential energy (see Eq. [3]). By looking at these plots at representative temperatures, one chooses a conservative (but affordable) value of n_{eq}, for instance, 15,000 in the present example.

Note that if a different choice of initial conditions is used (e.g., starting from a lattice configuration rather than from a random one), n_{eq} might need to be adjusted (up or down). It is advisable to do a new equilibration study each time one substantially alters the way things are done, or when the system's properties become significantly different (e.g., when one moves substantially above or below a melting transition).

Error Estimation

It is expected that in the limit of large n, $\langle U \rangle_n$ will approach $\langle U \rangle$, that $\langle U^2 \rangle_n$ will approach $\langle U^2 \rangle$, and that both E and C_V will converge to their correct values. But what are the uncertainties in the calculated values of E and C_V? Because the Metropolis method is intrinsically based on the sampling of configurations from a probability distribution function, appropriate statistical error analysis methods can be applied. This fact alone is an improvement on most other numerical integration techniques, which typically lack such strict error bounds.

A significant correlation exists between successive configurations in the Markov walk.[42] In fact, some successive configurations are identical to one another (this happens when the walker rejects an uphill energy move). Because the configurations are correlated, the potential energies (and squared potential energies) that result from the sequence are likewise correlated. In practice, this makes the error analysis somewhat more complicated than it would otherwise be if the configurations were completely uncorrelated. Unfortunately, it is rarely possible to efficiently generate configurations in an uncorrelated manner; to do that, we would be restricted to potential energy functions for which the Boltzmann probability function is analytically integrable, with an integral that is analytically invertible.[5] Very few distribution functions have this property even in one dimension, although the Gaussian distribution function is one such exception.[43]

Fortunately, because there is so much randomness in the MMC algorithm, a configuration (A) spaced by n_{eq} iterations from another configuration (B) will have virtually no correlation with B. Even after equilibration there is a definite correlation length, which we can define qualitatively as the number of iterations required for the algorithm to "forget" where it was originally. We can take advantage of this to determine the errors of our estimates of E and C_V. It is worth noting that most of the following discussion can also be applied to "molecular dynamics" calculations, as well as virtually all algorithms used in molecular simulation work to calculate thermodynamic averages.

If the n samples were statistically independent of one another (which they are not), we could simply estimate σ_U, the error in $\langle U \rangle_n$, by using methods we all learned in our undergraduate laboratory courses.[44] Here we find that σ_U is not given by the standard error formula

$$\sigma_U \neq \sqrt{\frac{\langle U^2 \rangle_n - \langle U \rangle_n^2}{n-1}} \qquad [19]$$

and so one must work a little harder to assess the error. The following approach is effective for estimating the error. If Eq. [19] were an equality, σ_U would give the error at the 67% confidence level; multiplying σ_U by 2 would give a 95% confidence interval (there would be a 95% probability that $\langle U \rangle$ lies between $\langle U \rangle_n - 2\sigma_U$ and $\langle U \rangle_n + 2\sigma_U$). If the samples were uncorrelated, these error bars would be guaranteed to be accurate descriptors of the confidence interval in the limit of large n. A similar formula would hold for $\langle U^2 \rangle$.

However, in the case of a Markov process the variance is only the first (and largest) term in a series of terms, which must be added together. For a general property f, σ_f is given by

$$\sigma_f = \sqrt{\frac{\langle f^2 \rangle - \langle f \rangle^2}{n-1}} + \begin{pmatrix} \text{covariance terms due to the} \\ \text{correlations between } f \text{ values} \end{pmatrix} \qquad [20]$$

where n is the number of steps after equilibration. These covariance terms typically converge slowly; details for treating these kinds of problems can be found elsewhere.[45–48] Another commonly used method is "blocking."[7,12,49] For definiteness, let $n = 1{,}000{,}000$ samples. We divide these samples into $N_B = 100$ blocks of $s = 10{,}000$ samples each, with the first 10,000 samples in block 1, the second 10,000 samples in block 2, and so forth. Within the lth block we form estimates of $\langle U \rangle$ $[\langle U \rangle_s^{\{l\}}]$, and $\langle U^2 \rangle$ $[\langle U^2 \rangle_s^{\{l\}}]$. A list of 100 "block averages" for each property is thus established. The 100 block averages of each property are statistically independent of one another if s is substantially greater than the correlation length. This in turn means they can be analyzed as if they were 100 statistically independent objects. Considering the uncertainty in $\langle U \rangle_n$, we therefore have

$$\sigma_U = \sqrt{\frac{\mathrm{Var}\left[\langle U \rangle_s^{\{l\}}\right]}{N_B - 1}} \qquad [21]$$

where

$$\mathrm{Var}\left[\langle U \rangle_s^{\{l\}}\right] = \frac{1}{N_B} \sum_{l=1}^{N_B} \left[\langle U \rangle_s^{\{l\}} - \langle U \rangle_n\right]^2 \qquad [22]$$

with similar formulas defining σ_{U^2}. Note that σ_U decreases asymptotically as the inverse square root of the number of samples (even though only the number of blocks is explicit), and that $2\sigma_U$ defines the 95% confidence interval for $\langle U \rangle_n$. For published work we generally recommend that the 95% confidence level be reported. Since the thermodynamic energy E is given by Eq. [2] (which is linear in $\langle U \rangle$), Eqs. [21] and [22] also happen to give directly the error in E.

It should be noted that in practice the parameters s and N_B must generally be determined empirically by trial and error, until convergence of the error estimate is achieved at a single temperature. This is usually straightforward in practice, and the prior determination of the equilibration period makes this task considerably easier to achieve because the correlation length is likely to be less than or equal to the equilibration period.

Similar equations and considerations hold for the average of the squared potential energy, but the propagation of the error through Eq. [3] is somewhat more involved. It is simplest to compute N_B values of $(C_V)_s^{\{\ell\}}$, the heat capacity within each block, using Eq. [3], and then using Eq. [23] to obtain the standard error in C_V from the variance of the block heat capacities:

$$\sigma_{C_V} = \sqrt{\frac{\mathrm{Var}\left[(C_V)_s^{\{\ell\}}\right]}{N_B - 1}} \qquad [23]$$

Table 1 Convergence of Approximate 95% Confidence Level Error Estimates of a Metropolis Monte Carlo Estimate of $\langle U \rangle$ and C_V with Respect to the Number of Blocks Used for the One-Dimensional Lennard-Jones Oscillator at $T = 1$ K[a,b]

N_B	$\langle U \rangle / k_B$ (K)	$2\sigma_U$/abs($\langle$U$\rangle$)	C_V/k_B	$2\sigma_{C_V}/C_V$
2	−123.4958	3.3636×10^{-5}	1. 0044	0.0072
4	—	4.2259×10^{-5}	—	0.0109
5	—	1.8810×10^{-5}	—	0.0055
8	—	3.5502×10^{-5}	—	0.0118
10	—	3.3166×10^{-5}	—	0.0102
16	—	2.9819×10^{-5}	—	0.0106
20	—	2.8031×10^{-5}	—	0.0088
25	—	2.6449×10^{-5}	—	0.0093
32	—	2.6552×10^{-5}	—	0.0093
40	—	2.3754×10^{-5}	—	0.0076
50	—	2.2375×10^{-5}	—	0.0083
80	—	2.5355×10^{-5}	—	0.0087

[a]See Eq. [18] and accompanying text.
[b]All calculations given are for a single Metropolis Monte Carlo calculation in which 20,000 equilibration cycles were followed by 500,000 data collection cycles. The stepsize was 0.01 nm, which produced an acceptance ratio of approximately 50%.

In Table 1 we show how the error in the average potential energy and the error in C_V (at the 95% confidence level) depend upon the number of blocks in a sample of fixed size (500,000 samples after 20,000 equilibration cycles) for the one-dimensional Lennard-Jones oscillator at 1 K. As noted previously,

Table 2 Average Potential Energy, Heat Capacity, and Fractional 95% Confidence Level Statistical Errors for the One-Dimensional Lennard-Jones Oscillator[a,b]

$T(K)$	$\langle U \rangle / k_B(K)$	$2\sigma_U$/abs($\langle U \rangle$)	C_V/k_B	$2\sigma_{C_V}/C_V$
0.1	−123.9499	5×10^{-6}	0.999	0.01
0.5	−123.7495	2×10^{-5}	1.002	0.01
1.0	−123.4938	4×10^{-5}	1.008	0.01
1.5	−123.2367	5×10^{-5}	1.020	0.01
2.0	−122.9741	9×10^{-5}	1.022	0.02
2.5	−122.7207	8×10^{-5}	1.027	0.01
3.0	−122.4455	2×10^{-4}	1.043	0.02
3.5	−122.1789	2×10^{-4}	1.048	0.02
4.0	−121.6182	3×10^{-4}	1.053	0.02
4.5	−121.6182	3×10^{-4}	1.063	0.02
5.0	−121.3596	3×10^{-4}	1.060	0.02
5.5	−121.0521	4×10^{-4}	1.076	0.02
6.0	−120.7361	5×10^{-4}	1.100	0.03
6.5	−120.4725	5×10^{-4}	1.104	0.03

[a]See Eq. [18] and accompanying text.
[b]All calculations given are for Metropolis Monte Carlo calculations in which 20,000 equilibration cycles were followed by 500,000 data collection cycles. The stepsize was 0.01 nm, which produced an acceptance ratio of approximately 50%. Ten blocks were used to estimate the statistical errors.

the relative error in C_V is significantly larger than the error in E for the same value of n. This is because the heat capacity represents the mean-squared fluctuations of the energy, and its error is proportional to the root mean square (RMS) fluctuations of the energy's fluctuations. In some cases the blocking technique may fail to effectively "break" the correlations; in such a situation other methods may be employed to estimate the covariance directly.[44–48]

It is important to emphasize that different thermodynamic properties generally converge at different rates. Those rates also generally depend on the temperature and can be quite strong functions of temperature in certain cases (for example, when phase transitions are possible). Table 2 shows how the error in E and the error in C_V vary with temperature for the Lennard-Jones oscillator. For this simple system the temperature dependence of the error is weak, but the trend of gradual increases in the error is evident. Also note that the average potential energy is approximately a linear function of temperature at low temperatures with a slope equal to $k_B/2$ (Figure 4), and that the reduced heat capacity C_V/k_B approaches 1.0 as the temperature approaches zero (Figure 5), both in agreement with the Equipartition Theorem.[10]

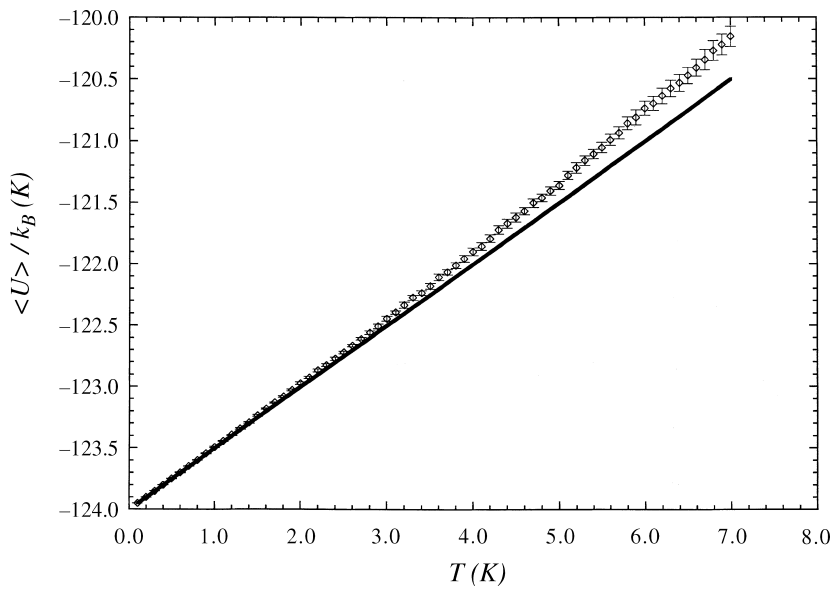

Figure 4 $\langle U \rangle / k_B$ as a function of temperature (diamonds) for the one-dimensional Lennard-Jones oscillator (see text). The Equipartition Theorem requires that the slope approach the harmonic limit ($= \frac{1}{2}$) as the temperature approaches zero (solid line).

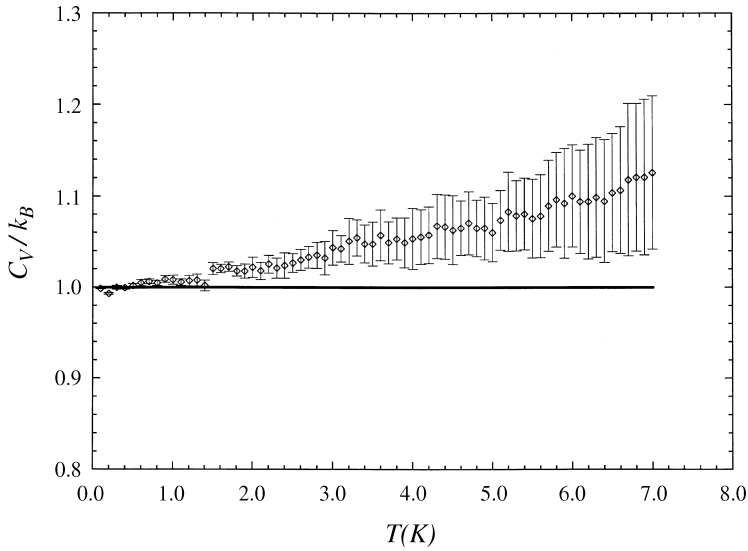

Figure 5 C_V/k_B as a function of temperature for the one-dimensional Lennard-Jones oscillator (see text). The Equipartition Theorem requires that the ratio of the heat capacity to k_B approach the harmonic limit (= 1) as the temperature approaches zero (solid line).

QUASI-ERGODICITY: AN INSIDIOUS PROBLEM

While the methods discussed thus far are sufficiently powerful to allow the simulation of many complex phenomena, there are circumstances where their direct implementation must be modified for efficient sampling. In many important cases the direct application of the preceding strategies with a (necessarily) finite set of points can give misleading or incorrect results.

In Monte Carlo computations of thermodynamic properties, it is often desirable that the sampling be ergodic.[50] In the present context, we define an ergodic random walk as one that can eventually reach every possible state from every possible initial state. A simulation that samples ergodically is typically characterized by low asymptotic variance and by rapid convergence. However, for systems for which there are wells in the potential energy surface that are separated by high barriers, sampling that is confined to a subset of the wells becomes problematic. This type of sampling is termed *quasi-ergodic sampling*.[15,51] When a system is quasi-ergodic it usually appears to be ergodic, exhibiting low asymptotic variance and rapid convergence, thus making quasi-ergodicity particularly difficult to detect. This problem is not unique to Monte Carlo methods, but is also a feature of molecular dynamics calculations and virtually all other molecular simulation methods.

Cluster melting simulations are situations in which special effort must be taken to ensure ergodic sampling. The melting transitions typically occur over a range of temperatures, and clusters can often coexist in solid-like and liquid-like forms (or in several solidlike forms) within this range. The large energy barriers characteristic of "magic number clusters" lead to quasi-ergodic sampling.[15]

To illustrate the kinds of problems to which we refer, we make use of a classical double-well potential that has proved to be useful[52] for the study of quasi-ergodicity:

$$U(x) = \frac{3x^4}{2a+1} + \frac{4(a-1)x^3}{2a+1} - \frac{6ax^2}{2a+1} + 1 \qquad [24]$$

This potential has a minimum of zero energy at $x = 1$, a second minimum of variable energy at $x = -a$ and a barrier of unit height separating the minima at $x = 0$. For $0 \leq a \leq 1$ it is useful to define the relative well depth by

$$\gamma = \frac{U(0) - U(-a)}{U(0) - U(1)} = a^3 \frac{a+2}{2a+1} \qquad [25]$$

For the case $\gamma = 0.9$ ($a = 0.961261$), $U(x)$ is shown in Figure 6. This model potential has many features generic to molecular potential surfaces, in particular

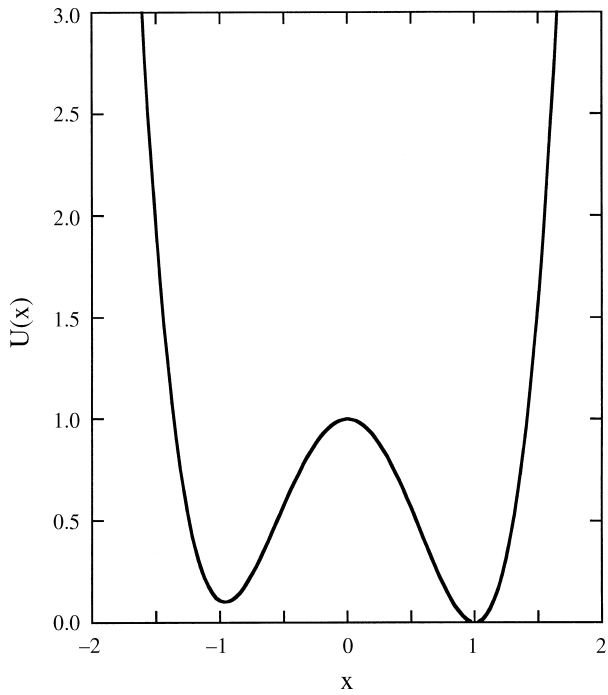

Figure 6 Asymmetric double-well potential $U(x)$ with $\gamma = 0.9$ (see Eqs. [24,25]).

the multiple minima separated by energy barriers. In general we might not know the locations and depths of the minima nor the heights or locations of the barriers that separate the minima on a potential surface. In thinking about the potential shown in Figure 6 it is helpful to imagine what happens in a simulation if the actual structure of the potential surface is not known in advance.

In a Metropolis Monte Carlo calculation, the simulation must be initiated at some configuration. We imagine the result of some random process initiating the walk in the highest-energy well (on the left side in Figure 6). At the lowest temperatures a finite Monte Carlo walk may never leave this well. Thermodynamic properties calculated from such a walk would appear reasonable with the statistical fluctuations in the computed values falling, as expected, with the number of points included. However, such computed thermodynamic properties would reflect contributions only from the left well and thereby be incorrect having ignored the contributions from the lower energy potential minimum. At somewhat higher simulation temperatures, a finite Metropolis Monte Carlo walk may visit both wells, hopping between the wells infrequently. The infrequency implies contributions from the two wells may not be properly weighted, and the resulting calculated thermodynamic property may still be incorrect. Contrary to the situation at low temperatures, the fluctuations of calculated properties at intermediate temperatures may not decrease properly with increasing Monte Carlo points, a behavior that is indicative of a problem in attaining an ergodic result. At sufficiently high temperatures, a Metropolis Monte Carlo walk can be expected to execute transitions over the barrier with sufficient frequency that a correct and well-behaved result follows.

We illustrate the behavior discussed in the previous paragraph with a simulation of the heat capacity associated with an assembly of one-dimensional particles subject to the potential energy expressed in Eq. [24]. We choose to examine the heat capacity because, as a fluctuation quantity, it is particularly sensitive to sampling errors. In Figure 7 we present the computed heat capacity of the double-well potential as a function of temperature. The solid dark line represents the exact results obtained from one-dimensional numerical quadrature. The dotted line with error bars marked "Metropolis" represents MMC data. The MMC calculations were initiated from the leftmost (highest-energy) well represented in Figure 6, and the simulation consists of 10^8 Metropolis Monte Carlo points. The Metropolis box size for the simulation was chosen so that approximately 50% of the attempted Metropolis moves are accepted. At the lowest temperatures both the exact and simulated results agree with the value anticipated from the Equipartition Theorem. However, the exact result rises at low temperatures owing to fluctuations between the two potential wells. The simulated MMC results remain flat because at low temperatures hops between the wells are not observed for a finite set of

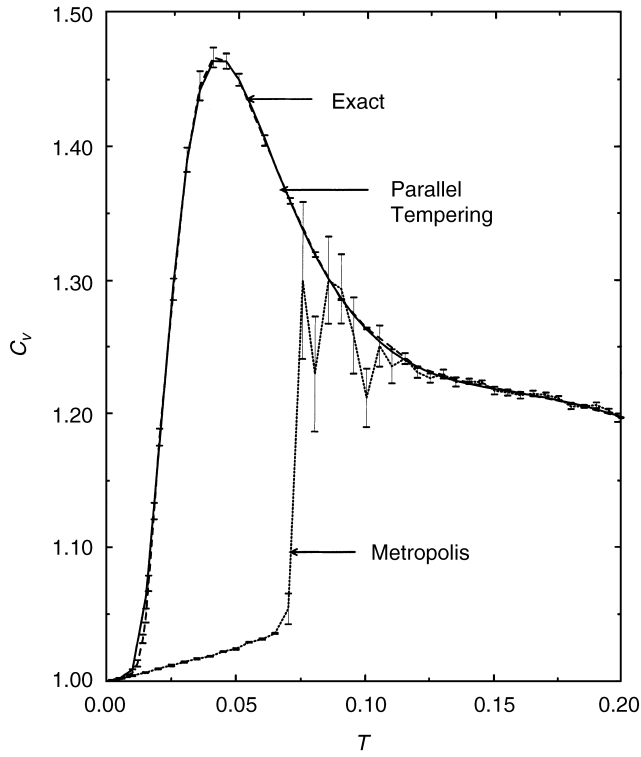

Figure 7 C_V as a function of T for the asymmetric double-well potential shown in Figure 6. The "exact" result (solid line) is obtained by direct integration of the Boltzmann average. "Metropolis" Monte Carlo results (dotted line) are seen to be in sharp disagreement with the exact result until temperatures are sampled that are well above the transition temperature for motion between the two potential energy wells. Parallel tempering Monte Carlo results (dashed line) are in all cases within the 95% confidence interval of the exact result.

Monte Carlo points. As the temperature is increased beyond the heat capacity maximum in the exact result, the MMC simulated points begin to rise but they are in poor agreement with the exact data. Additionally, the calculated error bars increase, but are artificially large in this calculation and do not accurately reflect the true asymptotic fluctuations of the heat capacity. Finally, at the highest calculated temperatures, both the Metropolis Monte Carlo and exact data are in agreement.

Several methods have been developed to remove such difficulties from MMC simulations. One obvious but not very useful method is to include more Metropolis Monte Carlo points. In the limit of an infinite simulation

adding more points is guaranteed to work, but in many cases this is impractical. Another approach, which does work for the present one-dimensional example problem, is to simply extend the original Metropolis scheme. In the Metropolis method one usually chooses a single maximum displacement for each Monte Carlo attempted move (the Metropolis step size). As in the example displayed in Figure 7, the step size is selected so that approximately 50% of the attempted moves are accepted. A simple modification to this procedure is to include two stepsizes. The first, used for most moves, has a size chosen to meet the usual 50% criterion. The second maximum stepsize is chosen to be two units of distance in length (the distance between the two minima in Figure 6). By using this magnified stepsize for a portion of the moves, the barrier between the wells is overcome even at low temperatures. It is not difficult to verify that the occasional inclusion of a magnified stepsize, which we call "mag-walking," satisfies detailed balance.[52]

Mag-walking is sufficiently simple that it can be useful for some many-particle applications where the locations of the barriers between the minima are known in advance. Unfortunately this simple extension, based on two (or more) maximum stepsizes, fails for many important applications of Monte Carlo to interacting many-particle systems. For many-body applications the potential minima are generally separated not just by a distance but also by one or more curvilinear directions. To move from one potential well to another in a many-particle application, specifying the distance between the wells is insufficient. Because the locations of the wells are seldom known in advance, more sophisticated approaches are usually necessary.

From the previous discussion, we see that the problem of adequately sampling a potential energy surface can often be solved with a strategy having two distinct parts. The first part concerns locating the relevant potential minima in the energy surface along with the transition state barriers that separate those minima. The second part concerns the development of a strategy to sample the important minima with the correct statistics. Simulation methods based on this two-part separation strategy have limited utility in cluster studies because the number of minima on a potential surface grows extremely rapidly with the size of the system studied (at a rate believed to be exponential). For example, a cluster of 13 atoms interacting via Lennard-Jones forces is known to have more than 1500 minima,[53] while nanodimensional Lennard-Jones clusters containing 147 atoms are believed to have about 10^{60} minima.[54] It is clearly impossible to enumerate all the minima for the latter cluster, even if a practical procedure were available to sample all such minima. Of course, we are really concerned with only those minima that are actually accessible with a reasonable probability at a given temperature, but the number of such minima can also be unreasonably large. The most successful approaches to proper sampling of complex potential surfaces involve methods that solve both parts in a single step.

OVERCOMING QUASI-ERGODICITY

Mag-Walking

As mentioned previously, the simplest method for addressing quasi-ergodicity is to occasionally magnify the stepsize in a Metropolis walk.[52] A probability P_m is specified, and a magnified stepsize is used if P_m is greater than or equal to a random number ξ. The probability that a magnified move is accepted is the same as that for a regular move. For example, if P_m is chosen to be 0.1 then magnified steps are possible 10% of the time.

This very simple method has a chance of being effective only when the magnified stepsize corresponds to a displacement that is known to carry a molecule from one conformation to another. For example, we have found mag-walking to be useful in certain contexts when applied to Barker–Watts rotational moves; $\Delta\theta_{MAX}$ may be equal to one radian for a regular rotational step, and π radians for a magnified step. We have used this method successfully in simulations of order–disorder transitions in solid ammonium chloride,[55] and in some preliminary investigations of cationic ammonium chloride clusters.[56] However, because one does not usually know a priori what kind of displacement will tend to generate new conformers, mag-walking is of limited utility.

Subspace Sampling

The subspace sampling method developed by Shew and Mills uses transitions among subspaces of the configuration space to overcome quasi-ergodicity.[57] Configuration space is divided into subspaces based on the potential energy surface. For example, for a double-well potential each well would be a subspace defined by the location of the local maximum separating the two wells (subspaces A and B hereafter). A transfer probability P_{AB} is specified and a transfer of the system from configuration i in subspace A to configuration j in subspace B is attempted if P_{AB} is greater than a random number ξ. The transfer is accepted by the probability $A(\boldsymbol{q}_{i,A} \rightarrow \boldsymbol{q}_{j,B})$,

$$A(\boldsymbol{q}_{i,A} \rightarrow \boldsymbol{q}_{j,B}) = c\frac{V_B}{V} \qquad \text{if } U(\boldsymbol{q}_{i,A}) < U(\boldsymbol{q}_{j,B})$$

$$= c\frac{V_B}{V}\exp[-\beta(U(\boldsymbol{q}_{i,A}) - U(\boldsymbol{q}_{j,B}))] \qquad \text{if } U(\boldsymbol{q}_{i,A}) > U(\boldsymbol{q}_{j,B}) \quad [26]$$

where V is the total volume, V_B is the volume occupied by subspace B and c is an empirically chosen constant. As in mag-walking, the subspace sampling

method requires prior knowledge of the potential energy surface to divide configuration space into subspaces. Moreover, the determination of each subspace's volume is only possible when definite rectilinear or spherical boundaries can be drawn which separate and enclose all the subspaces (in which case the volume calculations are trivial). Identifying definite boundaries that correctly separate the relevant subspaces can be challenging. Subspace sampling is therefore generally practical only for low-dimensional systems.

Jump Between Wells Method

The jump between wells (JBW) method developed by Senderowitz, Guarnieri, and Still[58] involves repeated sampling from known low energy conformers to generate a Markov chain. Each trial move is based on the transformation of one low-energy conformer to another and subsequent perturbation of the selected conformer. This method has been shown to work well for relatively small organic molecules, but because it requires knowledge of all low energy conformers, it may not be practical for larger molecules. This limitation applies especially to atomic and molecular clusters, since the location of all low energy conformers may itself be a significant computational challenge.

Atom-Exchange Method

The atom-exchange method was developed by Tsai, Abraham, and Pound[59] to speed barrier crossing in binary (two types of atoms) alloy cluster simulations. During the Metropolis walk two different types of atoms are periodically chosen, and their positions are exchanged. The exchange is accepted or rejected by the standard Metropolis acceptance probability. The utility of this method is naturally limited to systems of this particular type, namely, binary atomic clusters and liquids.

Histogram Methods

The single histogram method[17,60] involves constructing a histogram of energies $h(U)$ that are obtained at an elevated sampling temperature T_s, with $T_s > T$. For continuous systems a large number of bins should be set up to discretize the energy. Because the probability distribution function is known, it can be used to calculate ensemble averages at any temperature T

$$\langle f \rangle = \frac{\sum_U f(U)\, h(U)\, \exp[-U(\beta - \beta_s)]}{\sum_U h(U)\, \exp[-U(\beta - \beta_s)]} \qquad [27]$$

where β_s is $1/k_B T_s$. It is not necessary to actually save the histogram of energies if the summations are carried out over configurations instead of energies:[61]

$$\langle f \rangle = \frac{\sum_i (f(\boldsymbol{x}_i)/\{\exp[-\beta_s U(\boldsymbol{x}_i)]\}) \exp[-\beta U(\boldsymbol{x}_i)]}{\sum_i (1/\{\exp[-\beta_s U(\boldsymbol{x}_i)]\}) \exp[-\beta U(\boldsymbol{x}_i)]} \quad [28]$$

In addition to improving the simulation efficiency, the single-histogram method is designed to alleviate quasi-ergodicity by increasing the effective sampling temperature. The square of the error is proportional to the number of entries in the histogram, so the accuracy is greatest where $h(U)$ is largest and the error is greatest at the wings of the histogram.[62] Thus the limitation of this method is that it may not be accurate for temperatures far below the sampling temperature T_s, whereas T_s must be much higher than T for many cases to overcome quasi-ergodicity.

The accuracy of the histogram method can be improved and its limitations largely overcome by combining multiple histograms with overlapping temperature ranges.[60,63] The density is estimated from a linear combination of the estimates from multiple histograms $h_i(U)$ measured at temperatures T_i. The weight assigned to each estimate is optimized to reduce error through an iterative procedure. The simulated annealing–optimal histogram method takes a different approach wherein simulations at various temperatures are used to generate optimized sampling of energy bins.[17] A more detailed discussion is beyond the scope of this chapter.

Umbrella Sampling

Umbrella sampling is a technique that facilitates barrier crossing by introducing an artificial bias potential, called the "umbrella potential."[4] The umbrella potential optimally biases the sampling toward important regions of configuration space that might otherwise be rarely visited. The probability distribution of the physical system can then be extracted from the probability distribution of the unphysical system. The probability distributions are often analyzed along a "reaction coordinate" ζ, which can be one- or multidimensional and is expressed as a function of the coordinates of the system.[64] The reaction coordinate is often taken to be a distance or angle, or a linear combination of distances or angles.

A modified potential energy function $U_{\text{mod}} = U + E_u(\zeta)$ is constructed, where U is the true potential energy of the system and $E_u(\zeta)$ is the umbrella sampling potential [note that $\zeta = \zeta(\boldsymbol{q})$ in general]. The Boltzmann average of any property f can be computed by using $U_{\text{mod}} = U + E_u$ in place of U in the Metropolis algorithm and calculating the average as

$$\langle f \rangle = \frac{\langle f \exp(\beta E_u) \rangle_{U_{\text{mod}}}}{\langle \exp(\beta E_u) \rangle_{U_{\text{mod}}}} \quad [29]$$

where $\langle B \rangle_{U_{\text{mod}}}$ is the average of B taken using the modified potential function in a Markov walk where moves are accepted or rejected with probability

$$\min\{\, 1, e^{-\beta \Delta U_{\text{mod}}} \}$$

The efficiency of the method depends on the choice of umbrella potential. The limitation of the umbrella sampling method is that the umbrella potential must be appropriately chosen for each system of interest. There are many methods of calculating an appropriate umbrella potential depending on the goals of the simulation.[65] A number of procedures have been developed to automatically and iteratively calculate an umbrella potential for certain types of situations.[66–70] We will also see in a later section that umbrella sampling is used within the multicanonical ensemble. Finally, Valleau has recently reviewed the thermodynamic-scaling Monte Carlo method, which uses umbrella sampling to guide a system between thermodynamic states.[71]

J-WALKING, PARALLEL TEMPERING, AND RELATED METHODS

There is a class of generally applicable Monte Carlo methods that have been developed to address the problem of quasi-ergodicity. These methods involve allowing Metropolis walkers at different temperatures to either "jump" to one another or exchange configurations with one another. These methods generally result in greatly improved sampling, and they are based on two key concepts: (1) at temperatures that are sufficiently high to overcome the barriers between basins on the potential surface, Metropolis Monte Carlo methods are free of serious sampling problems; and (2) the particle density at high (but not infinite) temperatures has a form that reflects the structure of the underlying potential energy surface. Figure 8 illustrates this latter point.

The heavy solid line in Figure 8 represents the double-well potential, redrawn from Figure 6. The light solid line in Figure 8 is the particle density $\exp[-\beta U(x)]$ for $\beta = 1$ and the dashed line is the particle density for $\beta = 5$. The low temperature particle density ($\beta = 5$) is nearly zero in the barrier region near $x = 0$, signaling difficulties in accessing both potential wells in a Metropolis Monte Carlo calculation. At higher temperatures ($\beta = 1$) there is significant particle density in the barrier region, and Metropolis Monte Carlo simulations at such temperatures do not suffer from the sampling problems discussed here. Furthermore, the particle density for $\beta = 1$ reflects the structure of the underlying potential energy surface. The maxima in the particle density are located at the same places as the potential energy minima, with the deeper well associated with the largest particle density. It is desirable to exploit the information contained in the particle density at high temperatures,

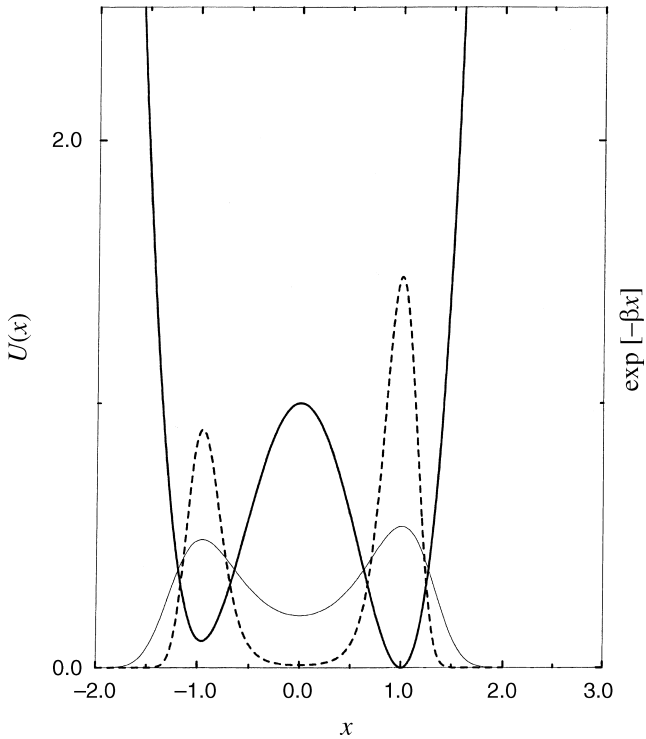

Figure 8 Boltzmann particle densities $\exp[-\beta U(x)]$ at $\beta = 1$ (light, solid line) and $\beta = 5$ (dashed line) for the asymmetric double-well potential $U(x)$ with $\gamma = 0.9$, which is also shown (dark, solid line; see Figure 6).

which is easily generated from a Metropolis calculation, to enable an ergodic simulation at low temperatures where the problems are more serious.

J-Walking

The first method to have been developed using the concepts discussed in the previous paragraph is called J-walking, or alternatively, jump-walking.[52] In the J-walking method a Metropolis calculation is performed at a temperature sufficiently high to avoid ergodic sampling problems. A Monte Carlo walk is then performed at a low temperature that periodically makes moves to the configurations determined by the high-temperature walk. The configurations can be taken from either a walk carried out simultaneously (often called *tandem J-walking*) or configurations that have been stored from a previous high-temperature walk. In the low temperature simulation, where achieving ergodic sampling is problematic, moves are made periodically to the configurations from the high-temperature walk (see Figure 9). Attempts to jump to these

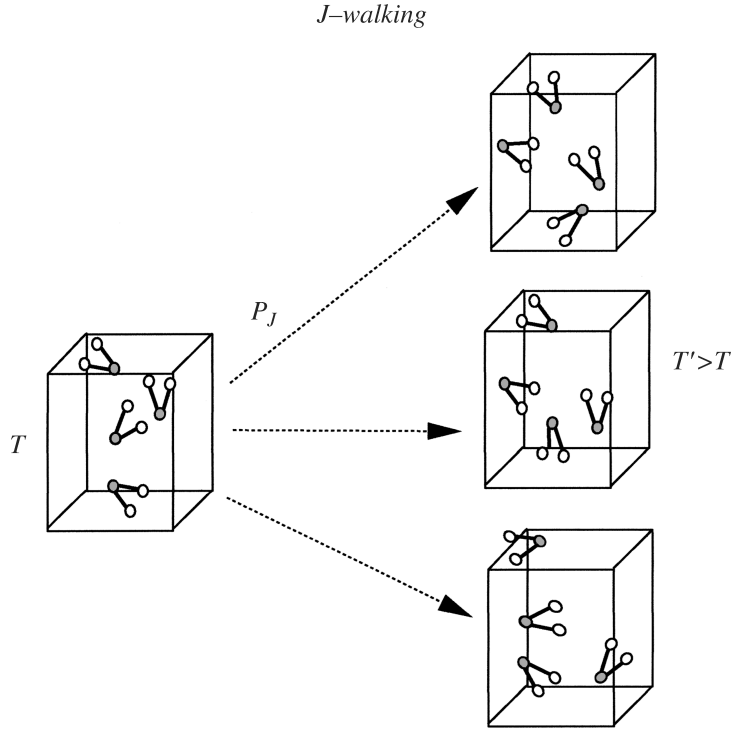

Figure 9 Schematic illustration of J-walking. A Monte Carlo walk at temperature T occasionally "jumps" to a configuration selected (not necessarily sequentially) from a configuration generated by a Monte Carlo walk at a higher temperature T'.

high-temperature configurations are accepted or rejected in a fashion that satisfies the detailed balance relation given in Eq. [5]. Recall that the transition rate may be written as a product of a trial probability Π and an acceptance probability A

$$K(\boldsymbol{x}_i \rightarrow \boldsymbol{x}_j) = \Pi(\boldsymbol{x}_i \rightarrow \boldsymbol{x}_j)A(\boldsymbol{x}_i \rightarrow \boldsymbol{x}_j) \qquad [30]$$

where Π can be taken to be any normalized distribution and A is constructed so that Eq. [30] is satisfied. For example, in Metropolis Monte Carlo the trial probability is usually taken to be a uniform distribution of points about the starting point of width L, as discussed previously. In J-walking the trial probability is taken to be that of the high-temperature configurations, which for classical canonical simulations takes the form

$$\Pi(\boldsymbol{x}_i \rightarrow \boldsymbol{x}_j) = Z_J^{-1} \exp[-\beta_J U(\boldsymbol{x}_i)] \qquad [31]$$

where Z_J is a normalization factor, $\beta_J = 1/k_B T_J$ and T_J is the temperature used to generate the high-temperature configurations. An acceptance probability ensuring satisfaction of detailed balance is given by

$$A(\boldsymbol{x}_i \rightarrow \boldsymbol{x}_j) = \min\left(1, \frac{\rho(\boldsymbol{x}_j)\,\Pi(\boldsymbol{x}_j \rightarrow \boldsymbol{x}_i)}{\rho(\boldsymbol{x}_i)\,\Pi(\boldsymbol{x}_i \rightarrow \boldsymbol{x}_j)}\right) \qquad [32]$$

which, for jumps to the high-temperature configurations in a classical canonical simulation, takes the form

$$A(\boldsymbol{x}_i \rightarrow \boldsymbol{x}_j) = \min(1, \exp(-\Delta\beta\Delta U)) \qquad [33]$$

where $\Delta\beta = \beta - \beta_J$ and $\Delta U = U(\boldsymbol{x}_j) - U(\boldsymbol{x}_i)$. Again, $\beta = 1/k_B T$, where T is the temperature of interest (associated with the low-temperature walk) and β_J is the inverse temperature used to generate the high-temperature external distribution.

The barriers that separate potential wells and make sampling difficult at low temperatures are overcome by jumping periodically to the high-temperature configurations. In this J-walking procedure detailed balance is satisfied to the extent that the configurations in the high-temperature walk are an exact representation of the actual high-temperature probability function (Eq. [31]). Such an exact representation of the actual probability function is possible only in the limit of an infinite external distribution, or when jumps to the tandem distribution are taken after an infinite number of steps. Systematic errors are difficult to remove in the tandem method, so most published applications have used external distributions. In practice a large (but necessarily finite) high-temperature distribution is generated. In the low-temperature simulation most Monte Carlo moves are generated using the Metropolis method with moves to the external high-temperature distribution at some prescribed frequency. In most applications these jump attempts are made about 10% of the simulation time, although higher jump attempt rates have been shown to be optimal in some cases.[72] To ensure that the distribution is sufficiently large (so as to be an accurate representation of the true high-temperature probability function), the distribution size must be larger than the number of jumps attempted in a simulation.

When using finite high-temperature external distributions, another issue arises when applying Eq. [33], because this equation assumes the exact probability is used. This issue is the origin of tandem J-walking problems. When generating the external high-temperature distribution using the Metropolis method, it is important to not store every configuration from that Metropolis walk. As noted previously, configurations generated using a Metropolis walk are correlated,[6] and the correlations can introduce systematic errors when moves are accepted and rejected on the basis of the criterion of Eq. [33]. To avoid these systematic errors it is necessary to store configurations from a

high-temperature Metropolis walk only after a sufficient number of Monte Carlo steps have been taken to break the correlations.

There is a final issue in J-walking that occurs in related simulation methods as well. If the difference in temperature between that used to generate the jumping distribution T_J and the temperature of interest T is sufficiently large, the probability that an attempted jump is accepted can become vanishingly small. When (effectively) no attempted jumps are accepted, the J-walking algorithm is equivalent to a Metropolis walk. Rather than jumping from temperature T to the distribution generated at temperature T_J, a series of temperatures are chosen between T and T_J, and configurations are stored at each of the series of temperatures to circumvent this problem. The temperatures are chosen to be sufficiently close to each other so that jump attempts between adjacent temperatures are accepted with reasonable probability (e.g., at least 10%). J-walking between adjacent temperatures is used to ensure ergodic distributions at each temperature.

To decrease the size of the external distributions needed in J-walking, an alternative method known as "smart walking" has been developed.[73] In smart walking, the energy of each high-temperature configuration is minimized before a jump is attempted. However, smart walking does not satisfy detailed balance.[74] A new technique called "smart darting" has recently been developed that modifies the smart-walking procedure by including the use of "darts," which are displacement vectors connecting the minimum-energy configurations.[74] Smart darting satisfies detailed balance, and it has been applied successfully to calculations of an 8-atom Lennard-Jones cluster and also for the alanine dipeptide.

While J-walking has proved to be useful for many applications,[75–82] the need for a series of large external distributions limits its application to many-body systems having modest numbers of particles. The needed configurations for J-walking can require a prohibitively large portion of both computer memory and disk space. In the next section we describe another approach that obviates the need for external distributions.

Parallel Tempering

Like J-walking, parallel tempering[83–90] addresses sampling problems by using information about the underlying potential surface obtained from a high-temperature walk where sampling is not problematic to ensure proper sampling by a lower-temperature simulation. While high-temperature configurations are fed to a low-temperature walk in J-walking, parallel tempering uses configurations that are exchanged between the high- and low-temperature walkers (see Figure 10).

To understand the basis of parallel tempering, we let

$$\rho_2(\boldsymbol{x}, \boldsymbol{x}') = F^{-1} \exp[-\beta U(\boldsymbol{x})] \exp[-\beta_J U(\boldsymbol{x}')] \qquad [34]$$

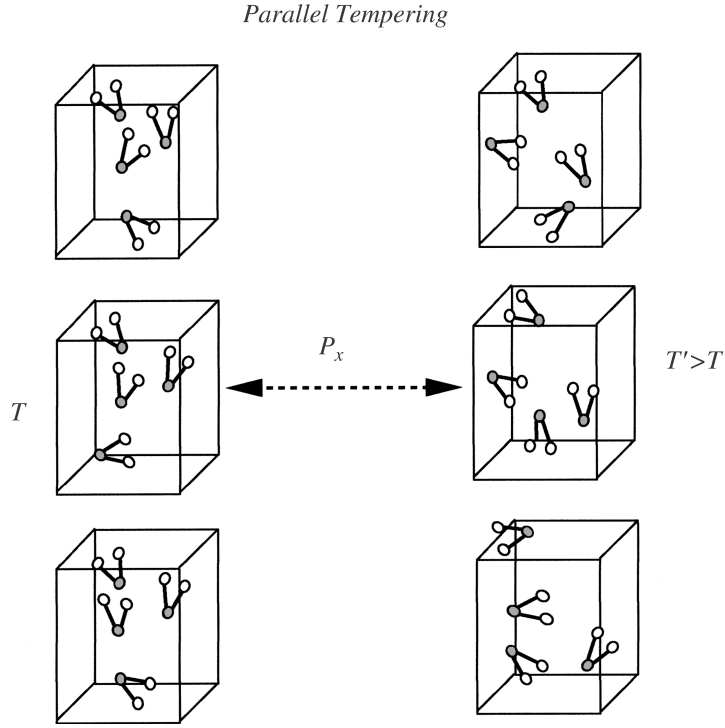

Figure 10 Schematic illustration of parallel tempering. Two Monte Carlo walks at two different temperatures (T and T' with $T' > T$) are executed in parallel, and occasionally the two walks exchange configurations with one another.

be the joint probability density that the low-temperature walker is at configuration $\boldsymbol{x}$ and the high-temperature walker is at configuration $\boldsymbol{x}'$. The detailed balance condition for exchanging configurations between the two walkers is given by

$$\rho_2(\boldsymbol{x}, \boldsymbol{x}')K(\boldsymbol{x} \to \boldsymbol{x}', \boldsymbol{x}' \to \boldsymbol{x}) = \rho_2(\boldsymbol{x}', \boldsymbol{x})K(\boldsymbol{x}' \to \boldsymbol{x}, \boldsymbol{x} \to \boldsymbol{x}') \qquad [35]$$

By solving for the ratio of the conditional transition probabilities

$$\frac{K(\boldsymbol{x} \to \boldsymbol{x}', \boldsymbol{x}' \to \boldsymbol{x})}{K(\boldsymbol{x}' \to \boldsymbol{x}, \boldsymbol{x} \to \boldsymbol{x}')} = \exp\{-(\beta - \beta_J)[U(\boldsymbol{x}') - U(\boldsymbol{x})]\} \qquad [36]$$

it is evident that if exchanges are accepted with the same probability as the acceptance criterion used in J-walking (see Eq. [33]), detailed balance is satisfied.

Like the Metropolis method, and unlike J-walking (which satisfies detailed balance only in the limit that the external distributions required are of infinite size), the parallel tempering approach satisfies detailed balance once the random walk has reached the asymptotic limit. Consequently, no external distributions are required, and parallel tempering can be organized in the same simple fashion as tandem J-walking. The simple organization implies that parallel tempering can be applied to large scale problems.

We illustrate an application of the parallel tempering method again using the double-well potential represented in Eq. [24]. The data have been generated using 28 temperatures roughly equally spaced between $T = 0.002$ and $T = 0.2$ in reduced units. Including a series of temperatures in the simulation ensures adequate acceptance of exchange attempts between adjacent temperatures. As in a J-walking simulation, if the gaps in temperature between adjacent temperatures are too large, exchanges can be accepted too infrequently for parallel tempering to provide an improvement over Metropolis simulations. Each temperature point plotted in Figure 7 is the result of 10^7 equilibration moves followed by 10^8 Monte Carlo points with the accumulation of data. Both the equilibration and Metropolis moves with the accumulation of data used box sizes satisfying the usual 50% acceptance criterion, and parallel tempering exchanges were attempted with a 10% probability. The parallel tempering data in Figure 7, displayed with the dashed line that is labeled "parallel tempering," are essentially indistinguishable from the exact results. Results using the J-walking method would be essentially identical to the parallel tempering data, but the J-walking results require significantly more effort to obtain.

Jumping to Tsallis Distributions

Both J-walking and parallel tempering depend on: (1) a set of configurations generated ergodically at a high temperature and (2) attempted moves to those configurations in a manner that satisfies detailed balance. The use of temperature as a parameter to generate ergodic configurations is not required. Instead, jumping approaches can be developed that use configurations generated in any fashion that are simultaneously ergodic and overcome energy barriers in the potential surface. Tsallis statistics[91] have been used to generate such ergodic configurations. Tsallis distributions are based on the probability function

$$\rho_q(\boldsymbol{x}) = N_q^{-1}[1 - (1 - q)\beta U(\boldsymbol{x})]^{1/(1-q)} \qquad [37]$$

where q is a parameter whose value is optimized for the system of interest. N_q is a normalization defined so that

$$N_q = \int d\boldsymbol{x}\,[1 - (1 - q)\beta U(\boldsymbol{x})]^{1/(1-q)} \qquad [38]$$

The Tsallis probability distribution is equivalent to the classical Boltzmann distribution in the limit $q \to 1$. As shown numerically elsewhere,[92] for $q > 1$ the Tsallis distribution broadens, overcoming energy barriers, and like high-temperature Boltzmann particle densities, the Tsallis distribution has maxima at the coordinates of potential minima. The Tsallis distribution can be used like high-temperature Boltzmann distributions if Tsallis configurations are accepted or rejected with probability

$$\min\left\{1,\ e^{-\beta\Delta U}\left[\frac{\rho_q(\boldsymbol{x})}{\rho_q(\boldsymbol{x}')}\right]^q\right\}$$

Successful applications of jumps to Tsallis distributions have included cluster systems.[92]

Applications to Microcanonical Simulations

Monte Carlo methods can also be applied to systems in the microcanonical ensemble,[7,10] and the techniques just described can be extended to this ensemble as well. For a cluster of N atoms with the same mass m, microcanonical averages of observables are obtained by calculating statistical expectations using the density function

$$\rho_E(\boldsymbol{x}) = \frac{\left(2\pi m/h^2\right)^{3N/2}\{1/[N!\,\Gamma(3N/2)]\}}{\Omega(E)}\,\Theta[E - U(\boldsymbol{x})]\,[E - U(\boldsymbol{x})]^{(3N/2)-1} \qquad [39]$$

where $\Gamma(y)$ is the gamma function, $\Theta(y)$ is the Heaviside (step) function and Ω(E) is the classical density of states[93]

$$\Omega(E) = \frac{1}{h^{3N}N!}\int d\boldsymbol{x}\,d\boldsymbol{p}\,\delta[E - H(\boldsymbol{x},\boldsymbol{p})] \qquad [40]$$

where $\boldsymbol{p}$ is the linear momentum vector, H the classical Hamiltonian, and $\delta[y]$ the Dirac delta function.[35] For example, the average kinetic energy K of a system can be obtained from the expression

$$\langle K\rangle = \frac{\int d\boldsymbol{x}\,\Theta(E-U)\,(E-U)^{(3N/2)-1}(E-U)}{\int d\boldsymbol{x}\,\Theta(E-U)\,(E-U)^{(3N/2)-1}} = \langle (E-U)\rangle \qquad [41]$$

Either parallel tempering[94,95] or J-walking[82] in the microcanonical ensemble consists of taking configurations from a walk at high energy E_h, rather than

from a high temperature. Exchanges or jumps are accepted or rejected with probability

$$\min\left\{1, \frac{\rho_E(\boldsymbol{x}')\rho_{E_b}(\boldsymbol{x})}{\rho_E(\boldsymbol{x})\rho_{E_b}(\boldsymbol{x}')}\right\}$$

Microcanonical parallel tempering has been extended to the molecular dynamics ensemble by introducing the appropriate center of mass and angular momentum constraints, the details of which can be found elsewhere.[94]

MULTICANONICAL ENSEMBLE / ENTROPY SAMPLING

Another class of methods that has been used to remove sampling difficulties is based on what is often called the *multicanonical ensemble*.[96] These methods have also been called "entropy sampling" methods,[97] for reasons that are made clear below.[98] It is easiest to understand the multicanonical methods by considering the full classical canonical coordinate–momentum distribution

$$\rho(\boldsymbol{x}, \boldsymbol{p}) = M^{-1} \exp[-\beta H(\boldsymbol{x}, \boldsymbol{p})] \qquad [42]$$

where M is a normalization factor. The canonical density can also be expressed as a function of the total energy E rather than the coordinates and momenta

$$\rho(\beta, E) = M^{-1}\Omega(E) \exp[-\beta E] \qquad [43]$$

As is discussed in many elementary treatments of statistical mechanics, $\Omega(E)$ is a rapidly increasing function of the energy.[99] Owing to the decay of the exponential factor in Eq. [42], $\rho(\beta, E)$ is a sharp Gaussian distribution, and is peaked about the mean energy associated with the inverse temperature β.

For a Metropolis walk using Eq. [42] the energy of each configuration generated can be tabulated to generate $\Omega(E)$. The sorting of configurations into energy bins is the basis of histogram sampling methods,[17,60,100,101] which (as mentioned previously) are beyond the scope of this chapter. Accurate estimates of $\Omega(E)$ can be achieved only for energies about the average energy of the temperature used to generate the density of states. Consequently, to generate $\Omega(E)$ accurately, calculations must be performed at a series of temperatures. The normalization for each energy distribution so generated must be chosen carefully such that the distributions created for each temperature match. The subtleties associated with this procedure are also beyond the scope of this chapter; for the purposes of this section, we only remark that $\Omega(E)$ can be constructed using a canonical Boltzmann Monte Carlo walk.[17]

Earlier we made clear that Monte Carlo walks can become trapped in regions of space owing to energy barriers that separate the disconnected configurations. In multicanonical methods the energy barriers are overcome by performing a walk in a space with a uniform energy distribution. The energy distribution is defined by the relation

$$\rho_M(E) = M^{-1}\Omega(E)w(E) = \bar{\kappa} \qquad [44]$$

where $\bar{\kappa}$ is a constant. We then sample with respect to the distribution

$$w(E) = \frac{\kappa}{\Omega(E)} \qquad [45]$$

where $\kappa = \bar{\kappa}M$. It is of interest to write

$$\begin{aligned} \ln w(E) &= \ln \kappa - \frac{k_B}{k_B} \ln \Omega(E) \\ w(E) &= K \exp[-k_B \ln \Omega(E)] \end{aligned} \qquad [46]$$

where K is a constant. Using the standard microcanonical expression for the entropy $S = k_B \ln \Omega$,[99] w becomes

$$w(E) = K \exp[-S(E)] \qquad [47]$$

Because of expression Eq. [47], multicanonical sampling is often called "entropy sampling."

To calculate canonical averages of an observable $f(\boldsymbol{x}, \boldsymbol{p})$ using the entropy distribution (Eq. [47]), we can implement umbrella sampling methods

$$\langle f \rangle = \frac{\int d\boldsymbol{x}\, d\boldsymbol{p}\, f(\boldsymbol{x}, \boldsymbol{p}) \exp(-\beta H)}{\int d\boldsymbol{x}\, d\boldsymbol{p}\, \exp(-\beta H)} \qquad [48]$$

$$= \frac{\left\langle f(\boldsymbol{x}, \boldsymbol{p}) \frac{\exp(-\beta H)}{w} \right\rangle_S}{\left\langle \frac{\exp(-\beta H)}{w} \right\rangle_S} \qquad [49]$$

where the subscript S below each average in Eq. [49] implies sampling with respect to the distribution w.

J-walking ideas have also been used in multicanonical approaches[102] as an alternative to calculating expectation values using Eq. [49]. In the multicanonical J-walking approach, periodic jumps are made to the distribution of configurations associated with the entropy distribution w rather than periodic jumps to a set of high-temperature ergodic configurations. To satisfy detailed balance, such jumps are accepted with probability

$$\min\left\{ 1, \frac{\exp[-\beta H(\boldsymbol{x}', \boldsymbol{p}')]\, w[H(\boldsymbol{x}, \boldsymbol{p})]}{\exp[-\beta H(\boldsymbol{x}, \boldsymbol{p})]\, w[H(\boldsymbol{x}', \boldsymbol{p}')]} \right\}$$

From an operational point of view, an important issue when using multicanonical methods is the generation of the multicanonical weight w. Because w is not in general known analytically, w must be generated numerically. Some of the most successful methods for generating w are based on iterative methods, the discussion of which can be found elsewhere.[103] These iterative procedures suffer from oscillatory convergence characteristics, but these convergence problems have been solved using other J-walking based approaches.[104] In practice, the generated multicanonical weights are accurately described at intermediate energies but often provide poor representations at low energies. Consequently, multicanonical methods can be deficient for systems having important low temperature structures that are difficult to access.[88]

CONCLUSIONS

As you perhaps may have noticed, many Monte Carlo methods can be found in the literature. We have not described all of the Monte Carlo methods developed to date. Indeed, new methods are being developed every day, and methods better than those that now exist can be expected to appear in the future. Nonetheless we can make a few definite recommendations to the readers interested in writing their own Monte Carlo programs (especially for studying atomic and molecular clusters) based on existing methodology.

First, it is often useful to study a simple potential such as that given in Eq. [29], and use it as a *minimal* proving ground for testing and implementation of new Monte Carlo algorithms. Such potentials contain some (but not all) of the features of more realistic and complex problems. However, some methods will work well for an asymmetric double-well potential but fail when applied to more complex systems. We therefore warn the reader: let the buyer of a "foolproof" Monte Carlo algorithm beware!

Second, it is advisable to carry out ordinary Metropolis Monte Carlo calculations using Metropolis sampling[1] (with Barker–Watts rotational sampling[41] and center-of-mass moves in addition to single-particle moves if the cluster contains molecules) as a function of temperature before moving to more complex methods. This will help identify the temperature range of greatest importance and indicate how high the temperature must be for quasi-ergodic sampling problems to be avoided. This information can then be used in parallel tempering, J-walking, and other calculations.

Parallel tempering works, and works well in a wide variety of contexts (although the reader should be warned that we are even now finding particularly nasty examples where even parallel tempering converges at an unacceptably slow rate). In our experience, parallel tempering is much simpler to

implement than many of the alternatives. Thus, our final recommendation is that parallel tempering is strongly recommended whenever quasi-ergodicity problems are possible. Parallel tempering can be extremely useful in Monte Carlo studies of atomic and molecular clusters. It also can be used to excellent effect in molecular dynamics simulations, and thus it constitutes an extremely general and powerful technique. We are hopeful that the reader who uses this chapter as a starting point to develop new and improved Monte Carlo methods will find the work as rewarding and interesting as we have.

ACKNOWLEDGMENTS

We are grateful for the efforts of all of our colleagues and students who have helped us understand and use the Monte Carlo method, and other stochastic methods, in our studies of clusters and condensed phases. Our sincere thanks to Sarah Kopel-Fink for assisting with some of the calculations presented here, and to Gayle G. Topper for her help with the literature search. RQT and KRL's work on this chapter was supported in part by the Petroleum Research Fund of the American Chemical Society under grant PRF 33601-B6. DLF's work on this chapter has been supported in part by NSF grant CHE-0095053. DB acknowledges fellowship support from the Enders Memorial Fund, as well from as the Supercomputer Research Institute (SCRI) at Florida State University for a travel grant to attend the 1999 SCRI Monte Carlo Workshop.

REFERENCES

1. N. Metropolis, A. W. Rosenbluth, M. N. Rosenbuth, A. H. Teller, and E. Teller, *J. Chem. Phys.,* **21,** 1087 (1953). Equation of State Calculations by Fast Computing Machines.
2. J. M. Hammersley and D. C. Handscomb, *Monte Carlo Methods,* Methuen, London, 1964.
3. J. P. Valleau and S. G. Whittington, in *Statistical Mechanics Part A: Equilibrium Techniques,* B. J. Berne, Ed., Plenum Press, New York, 1977, pp. 137–168. A Guide to Monte Carlo for Statistical Mechanics: 1. Highways.
4. J. P. Valleau and G. M. Torrie, in *Statistical Mechanics Part A: Equilibrium Techniques*; B. J. Berne, Ed., Plenum Press, New York, 1977, pp. 169–191. A Guide to Monte Carlo for Statistical Mechanics: 2. Byways.
5. K. Binder, in *Monte Carlo Methods in Statistical Physics,* K. Binder. Ed., Springer-Verlag, Berlin, 1979, pp. 1–45. Introduction: Theory and "Technical" Aspects of Monte Carlo Simulations.
6. M. H. Kalos and P. A. Whitlock, *Monte Carlo Methods,* Wiley, New York, 1986.
7. M. P. Allen and D. J. Tildesley, *Computer Simulation of Liquids,* Clarendon Press, Oxford, 1992.
8. J. M. Lluch, M. Moreno, and J. Bertran, in *Computational Chemistry: Structure, Interaction, and Reactivity*; S. Fraga, Ed., Elsevier Science Publishers, Amsterdam, 1992, p. 311. Monte Carlo Simulations for Solutions.
9. D. J. Tildesley, *NATO ASI Ser., Ser. C* (*Computer Simulations in Chemical Physics*) **397,** 1 (1993). The Monte Carlo Method.
10. R. L. Rowley, *Statistical Mechanics for Thermophysical Property Calculations,* PTR Prentice-Hall, Englewood Cliffs, NJ, 1994.
11. D. Frenkel and B. Smit, *Understanding Molecular Simulation: From Algorithms to Applications,* Academic Press, San Diego, 1996.

12. A. R. Leach, *Molecular Modelling: Principles and Applications,* Longman, London, 1996.
13. J. I. Siepmann, *Adv. Chem. Phys.,* **105,** 1–12 (1999). An Introduction to the Monte Carlo Method for Particle Simulations.
14. G. Benedek, T. P. Martin, and G. Pacchioni, Eds., *Elemental and Molecular Clusters: Proc. 13th International School,* Springer-Verlag, Berlin, 1988.
15. D. L. Freeman and J. D. Doll, *Ann. Rev. Phys. Chem.,* **47,** 43–80 (1996). Computational Studies of Clusters: Methods and Results.
16. D. Eisenberg and D. Crothers, *Physical Chemistry with Applications to the Life Sciences,* Benjamin/Cummings, Menlo Park, CA, 1979.
17. D. M. Ferguson and D. G. Garrett, *Adv. Chem. Phys.,* **105,** 311 (1999). Simulated Annealing—Optimal Histogram Methods.
18. B. J. Berne and D. Thirumalai, *Ann. Rev. Phys. Chem.,* **37,** 401–424 (1986). On the Simulation of Quantum Systems: Path Integral Methods.
19. D. L. Freeman and J. D. Doll, *Adv. Chem. Phys.,* **70B,** 139 (1988). The Quantum Mechanics of Clusters.
20. J. D. Doll, D. L. Freeman, and T. L. Beck, *Adv. Chem. Phys.,* **78,** 61–128 (1990). Equilibrium and Dynamical Fourier Path Integral Methods.
21. G. A. Voth, *Adv. Chem. Phys.,* **93,** 135–218 (1996). Path Integral Centroid Methods in Quantum Statistical Mechanics and Dynamics.
22. R. Q. Topper, *Adv. Chem. Phys.,* **105,** 117–170 (1999). Adaptive Path-Integral Monte Carlo Methods for Accurate Computation of Molecular Thermodynamic Properties.
23. J. W. Gibbs, *Elementary Principles in Statistical Mechanics,* Yale Univ. Press, New Haven, CT, 1902.
24. J. H. Halton, *SIAM Rev.,* **12,** 1 (1970). A Retrospective and Prospective Survey of the Monte Carlo Method.
25. F. James, *Rep. Prog. Phys.,* **43,** 73 (1980). Monte Carlo Theory and Practice.
26. W. H. Press, S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery, *Numerical Recipes in FORTRAN: The Art of Scientific Computing,* 2nd ed., Cambridge Univ. Press, Cambridge UK, 1992.
27. J. K. Lee, J. A. Barker, and F. F. Abraham, *J. Chem. Phys.,* **58,** 3166 (1973). Theory and Monte Carlo Simulation of Physical Clusters in the Imperfect Vapor.
28. A. Srinivasan, D. M. Ceperley, and M. Mascagni, *Adv. Chem. Phys.,* **105,** 13–36 (1999). Random Number Generators for Parallel Applications.
29. R. W. Wilde and S. Singh, *Statistical Mechanics: Fundamentals and Modern Applications,* Wiley, New York, 1998.
30. W. L. McMillan, *Phys. Rev.,* **138,** 442 (1965). Ground State of Liquid ^{4}He.
31. J. B. Anderson, in *Reviews in Computational Chemistry,* K. B. Lipkowitz and D. B. Boyd, Eds., Wiley-VCH, New York, 1999, Vol. 13, pp. 133–182. Quantum Monte Carlo: Atoms, Molecules, Clusters, Liquids, and Solids.
32. A. Riganelli, W. Wang, and A. J. C. Varandas, *J. Phys. Chem. A,* **103,** 8303 (1999). Monte Carlo Approach to Internal Partition Functions for Van der Waals Molecules.
33. U. Grenander, *Ark. Matemat.,* **1,** 195 (1950). Stochastic Processes and Statistical Inference.
34. W. K. Hastings, *Biometrika,* **5,** 97 (1970). Monte Carlo Sampling Methods Using Markov Chains and Their Applications.
35. H. Goldstein, *Classical Mechanics,* 2nd ed., Addison-Wesley, Reading MA, 1980.
36. D. A. McQuarrie, *Statistical Mechanics,* HarperCollins, New York, 1976.
37. D. D. Frantz, D. L. Freeman, and J. D. Doll, *J. Chem. Phys.,* **93,** 2769 (1990). Reducing Quasi-Ergodic Behavior in Monte Carlo Simulations by J-Walking: Applications to Atomic Clusters.

38. I. R. McDonald, *Molec. Phys.,* **23,** 41 (1972). NpT-Ensemble Monte Carlo Calculations for Binary Liquid Mixtures.

39. J. I. Siepmann, *Adv. Chem. Phys.,* **105,** 443 (1999). Monte Carlo Methods for Simulating Phase Equilibria of Complex Fluids.

40. A. Z. Panagiotopoulos, *Molec. Phys.,* **61,** 813 (1987). Direct Determination of Phase Coexistence Properties of Fluids by Monte Carlo Simulation in a New Ensemble.

41. J. A. Barker and R. O. Watts, *Chem. Phys. Lett.,* **3,** 144 (1969). Structure of Water: A Monte Carlo Calculation.

42. W. J. Dixon, *Ann. Math. Stat.,* **15,** 119 (1944). Further Contributions to the Problem of Serial Correlation.

43. G. E. P. Box and M. E. Müller, *Ann. Math. Statist.,* **29,** 610 (1958). A Note on the Generation of Random Normal Deviates.

44. J. R. Taylor, *An Introduction to Error Analysis: The Study of Uncertainties in Physical Measurements,* University Science, Mill Valley, CA, 1982.

45. E. B. Smith and B. H. Wells, *Molec. Phys.,* **53,** 701 (1984). Estimating Errors in Molecular Simulation Calculations.

46. T. P. Straatsma, H. J. C. Berendsen, and A. J. Stam, *Molec. Phys.,* **57,** 89 (1986). Estimation of Statistical Errors in Molecular Simulation Calculations.

47. R. Q. Topper and D. G. Truhlar, *J. Chem. Phys.,* **97,** 3647 (1992). Quantum Free-Energy Calculations: Optimized Fourier Path-Integral Monte Carlo Computation of Coupled Vibrational Partition Functions.

48. J. D. Doll and D. L. Freeman, *Chem. Phys. Lett.,* **227,** 436 (1994). Decisions, Decisions: Noise and Its Effect on Integral Monte Carlo Algorithms.

49. J. M. Haile, *Molecular Dynamics Simulation: Elementary Methods,* Wiley, New York, 1992.

50. R. W. Wilde and S. Singh, *Statistical Mechanics: Fundamentals and Modern Applications,* Wiley, New York, 1998.

51. J. I. Siepmann and I. R. McDonald, *Molec. Phys.,* **75,** 255 (1992). Monte Carlo Simulations of Mixed Monolayers.

52. D. D. Frantz, D. L. Freeman, and J. D. Doll, *J. Chem. Phys.,* **93,** 2769 (1990). Reducing Quasi-Ergodic Behavior in Monte Carlo Simulations by J-Walking: Applications to Atomic Clusters.

53. D. B. Faken, A. F. Voter, D. L. Freeman, and J. D. Doll, *J. Phys. Chem.,* **103,** 9521 (1999). Dimensional Strategies and the Minimization Problem: Barrier-Avoiding Algorithms.

54. D. J. Wales and J. P. K. Doye, *J. Phys. Chem.,* **101,** 5111 (1997). Global Optimization by Basin-Hopping and the Lowest Energy Structures of Lennard-Jones Clusters Containing up to 110 Atoms.

55. R. Q. Topper and D. L. Freeman, *Los Alamos Chemical Physics Preprint Database,* #9403002, http://xxx.lanl.gov/archive/physics (1994). Monte Carlo Studies of the Orientational Order-Disorder Phase Transition in Solid Ammonium Chloride.

56. D. Bergin, P. Sweeney, S. Fink, and R. Q. Topper, work in progress.

57. C.-Y. Shew and P. Mills, *J. Phys. Chem.,* **97,** 13824 (1993). A Monte Carlo Method to Simulate Systems with Barriers: Subspace Sampling.

58. H. Senderowitz, F. Guarnieri, and W. C. Still, *J. Am. Chem. Soc.,* **117,** 8211 (1995). A Smart Monte-Carlo Technique for Free-Energy Simulations of Multiconformational Molecules: Direct Calculations of the Conformational Populations of Organic Molecules.

59. N.-H. Tsai, F. F. Abraham, and G. M. Pound, *Surf. Sci.,* **77,** 465 (1978). The Structure and Thermodynamics of Binary Microclusters: A Monte Carlo Simulation.

60. A. M. Ferrenberg and R. H. Swendsen, *Phys. Rev. Lett.,* **61,** 2635 (1988). New Monte Carlo Technique for Studying Phase Transitions.

61. C. J. Tsai and K. D. Jordan, *J. Chem. Phys.,* **99,** 6957 (1993). Use of the Histogram and Jump-Walking Methods for Overcoming Slow Barrier Crossing Behavior in Monte Carlo Simulations: Applications to the Phase Transitions in the $(Ar)_{13}$ and $(H_2O)_8$ Clusters.
62. R. H. Swendsen, *Physica A,* **194,** 53 (1993). Modern Methods of Analyzing Monte Carlo Computer Simulations.
63. P. Labastie and R. L. Whetten, *Phys. Rev. Lett.,* **65,** 1567 (1990). Statistical Thermodynamics of the Cluster Solid-Liquid Transition.
64. O. Engkvist and G. Karlström, *Chem. Phys.,* **63,** 213 (1996). A Method to Calculate the Probability Distribution for Systems with Large Energy Barriers.
65. T. C. Beutler and W. F. van Gunsteren, *J. Chem. Phys.,* **100,** 1492 (1994). The Computation of a Potential of Mean Force: Choice of the Biasing Potential in the Umbrella Sampling Technique.
66. M. Mezei, *J. Comput. Phys.,* **68,** 237 (1987). Adaptive Umbrella Sampling: Self-Consistent Determination of the Non-Boltzmann Bias.
67. J. E. Hunter III and W. P. Reinhardt, *J. Chem. Phys.,* **103,** 8627 (1995). Finite-Size-Scaling Behavior of the Free-Energy Barrier between Coexisting Phases—Determination of the Critical Temperature and Interfacial Tension of the Lennard-Jones Fluid.
68. R. W. W. Hooft, B. P. van Eijck, and J. Kroon, *J. Chem. Phys.,* **97,** 6690 (1992). An Adaptive Umbrella Sampling Procedure in Conformational Analysis Using Molecular Dynamics and Its Application to Glycol.
69. C. Bartels and M. Karplus, *J. Comput. Chem.,* **18,** 1450 (1997). Multidimensional Adaptive Umbrella Sampling: Applications to Main Chain and Side Chain Peptide Conformations.
70. C. Bartels and M. Karplus, *J. Phys. Chem. B,* **102,** 865 (1998). Probability Distributions for Complex Systems: Adaptive Umbrella Sampling of the Potential Energy.
71. J. P. Valleau, *Adv. Chem. Phys.,* **105,** 369 (1999). Thermodynamic-Scaling Methods in Monte Carlo and Their Application to Phase Equilibria.
72. M. Eleftheriou, D. Kim, J. D. Doll, and D. L. Freeman, *Chem. Phys. Lett.,* **276,** 353 (1997). Information Theory and the Optimization of Monte Carlo Simulations.
73. R. Zhou and B. J. Berne, *J. Chem. Phys.,* **107,** 9185 (1997). Smart Walking: A New Method for Boltzmann Sampling of Protein Conformations.
74. I. Andricioaei, J. E. Straub, and A. F. Voter, *J. Chem. Phys.,* **114,** 6994 (2001). Smart Darting Monte Carlo.
75. M. A. Strozak, G. E. Lopez, and D. L. Freeman, *J. Chem. Phys.,* **97,** 4445 (1992). Gibbs Free-Energy Changes for the Growth of Argon Clusters Adsorbed on Graphite.
76. G. E. Lopez and D. L. Freeman, *J. Chem. Phys.,* **98,** 1428 (1993). A Study of Low Temperature Heat Capacity Anomalies in Bimetallic Alloy Clusters Using J-Walking Monte Carlo Methods.
77. D. D. Frantz, *J. Chem. Phys.,* **102,** 3747 (1995). Magic Numbers for Classical Lennard-Jones Cluster Heat-Capacities.
78. C. J. Tsai and K. D. Jordan, *J. Chem. Phys.,* **99,** 6957 (1993). Use of the Histogram and Jump-Walking Methods for Overcoming Slow Barrier Crossing Behavior in Monte Carlo Simulations: Applications to the Phase Transitions in the $(Ar)_{13}$ and $(H_2O)_8$ Clusters.
79. A. Matro, D. L. Freeman, and R. Q. Topper, *J. Chem. Phys.,* **104,** 8690 (1996). Computational Study of the Structures and Thermodynamic Properties of Ammonium Chloride Clusters Using a Parallel Jump-Walking Approach.
80. A. J. Acevedo, L. M. Caballero, and G. E. Lopez, *J. Chem. Phys.,* **106,** 7257 (1997). Phase Transitions in Molecular Clusters.
81. N. Y. Matos and G. E. Lopez, *J. Chem. Phys.,* **109,** 1141 (1998). Classical Monte Carlo Study of Phase Transitions in Rare-Gas Clusters Adsorbed on Model Surfaces.
82. E. Curotto, D. L. Freeman, and J. D. Doll, *J. Chem. Phys.,* **109,** 1643 (1998). A J-Walking Algorithm for Microcanonical Simulations: Applications to Lennard-Jones Clusters.

83. E. Marinari and G. Parisi, *Europhys. Lett.,* **19,** 451 (1992). Simulated Tempering: A New Monte Carlo Scheme.
84. C. J. Geyer and E. A. Thompson, *J. Am. Stat. Assoc.,* **90,** 909 (1995). Annealing Markov-Chain Monte-Carlo with Applications to Ancestral Inference.
85. M. C. Tesi, E. J. Janse van Rensburg, E. Orlandini, and S. G. Whittington, *J. Stat. Phys.,* **82,** 155 (1996). Monte Carlo Study of the Interacting Self-Avoiding Walk Model in Three Dimensions.
86. M. Falcioni and M. W. Deem, *J. Chem. Phys.,* **110,** 1754 (1999). A Biased Monte Carlo Scheme for Zeolite Structure Solution.
87. Q. Yan and J. J. de Pablo, *J. Chem. Phys.,* **111,** 9509 (1999). Hyper-Parallel Tempering Monte Carlo: Application to the Lennard-Jones Fluid and the Restricted Primitive Model.
88. J. P. Neirotti, F. Calvo, D. L. Freeman, and J. D. Doll, *J. Chem. Phys.,* **112,** 10340 (2000). Phase Changes in 38-Atom Lennard-Jones Clusters. I. A Parallel Tempering Study in the Canonical Ensemble.
89. M. R. Ghayal and E. Curotto, *J. Chem. Phys.,* **113,** 4298 (2000). The Melting of Ar_{54}–HF: A Canonical Parallel Tempering Simulation.
90. D. D. Frantz, *J. Chem. Phys.,* **115,** 6136 (2001). Magic Number Behavior for Heat Capacities of Medium-Sized Classical Lennard-Jones Clusters.
91. C. Tsallis, *J. Stat. Phys.,* **52,** 479 (1988). Possible Generalization of Boltzmann-Gibbs Statistics.
92. I. Andricioaei and J. E. Straub, *J. Chem. Phys.,* **107,** 9117 (1997). On Monte Carlo and Molecular Dynamics Methods Inspired by Tsallis Statistics: Methodology, Optimization, and Application to Atomic Clusters.
93. J. I. Steinfeld, J. S. Francisco, and W. L. Hase, *Chemical Kinetics and Dynamics,* Prentice-Hall, Englewood Cliffs NJ, 1989.
94. F. Calvo, J. P. Neirotti, D. L. Freeman, and J. D. Doll, *J. Chem. Phys.,* **112,** 10350 (2000). Phase Changes in 38-Atom Lennard-Jones Clusters. II. A Parallel Tempering Study of Equilibrium and Dynamic Properties in the Molecular Dynamics and Microcanonical Ensembles.
95. E. Curotto, *J. Chem. Phys.,* **114,** 4533 (2001). The HF Stretch Red Shift as a Function of Internal Energy in Ar_n – HF ($n = 12,54$): Comparisons in the Microcanonical Ensemble.
96. B. A. Berg and T. Neuhaus, *Phys. Lett. B,* **267,** 249 (1991). Multicanonical Algorithms for First Order Phase Transitions.
97. J. Lee, *Phys. Rev. Lett.,* **71,** 211 (1993). New Monte Carlo Algorithm: Entropic Sampling.
98. B. A. Berg, U. H. E. Hansmann, and Y. Okamoto, *J. Phys. Chem.,* **99,** 2236 (1995). Comment on "Monte Carlo Simulation of a First-order Transition for Protein Folding."
99. R. S. Berry, S. A. Rice, and J. Ross, *Matter in Equilibrium: Statistical Mechanics and Thermodynamics,* 2nd ed., Oxford Univ. Press, New York, 2002.
100. R. H. Swendsen and J. S. Wang, *Phys. Rev. Lett.,* **58,** 86 (1987). Nonuniversal Critical Dynamics in Monte Carlo Simulations.
101. P. Labastie and R. L. Whetten, *Phys. Rev. Lett.,* **65,** 1567 (1990). Statistical Thermodynamics of the Cluster Solid-Liquid Transition.
102. H. Xu and B. J. Berne, *J. Chem. Phys.,* **110,** 10299 (1999). Multicanonical Jump Walking: A Method for Efficiently Sampling Rough Energy Landscapes.
103. Y. Okamoto and U. H. E. Hansmann, *J. Phys. Chem.,* **99,** 11276 (1995). Thermodynamics of Helix-Coil Transitions Studied by Multicanonical Algorithms.
104. M-H. Hao and H. A. Scheraga, *J. Phys. Chem.,* **98,** 9882 (1994). Statistical Thermodynamics of Protein-Folding Sequence Dependence.