Fast Calculation of the Density of States of a Fluid by Monte Carlo Simulations

Qiliang Yan and Juan J. de Pablo*
Department of Chemical Engineering, University of Wisconsin-Madison, Madison, Wisconsin 53706

(Received 11 October 2002; published 23 January 2003)

Two related methods are proposed to calculate the density of states of a fluid from Monte Carlo simulations. In contrast to previous approaches, which require that histograms be accumulated in a stochastic manner, the methods proposed here rely on evaluation of the instantaneous temperature. In the first method, the temperature is calculated from the gradient of the forces. In the second, it is estimated from the kinetic contribution to the total energy. The validity and usefulness of the new approaches are demonstrated by presenting results from simulations of a Lennard-Jones fluid. It is shown that the new methods are considerably faster than previously available techniques.

DOI: 10.1103/PhysRevLett.90.035701 PACS numbers: 64.60.Cn, 02.70.Uu, 05.10.Ln, 05.70.Ce

Molecular simulations of complex systems, such as proteins, polymers, or glasses, are often limited by the rugged nature of their energy landscape. Deep energy minima, separated by large barriers, make sampling of configuration space particularly challenging. Over the last decade, several simulation methods have been developed in attempts to systematically smooth out such landscapes. Umbrella sampling, multicanonical sampling, or parallel tempering formalisms provide examples of such methods [1–3].

The central element of those simulation techniques is the density of states Ω. Knowledge of the density of states would permit construction of stochastic algorithms designed to visit states with uniform probability $p = 1/\Omega$, regardless of whether a particular configuration of the system happens to correspond to an energy minimum or to the top of an energy barrier. Unfortunately the density of states is not known *a priori*; in fact, it is precisely the function that one would like to extract from a simulation, if it were at all possible.

Traditional multicanonical algorithms [2] provide a rough estimate of the density of states through "weights," constructed to promote or inhibit visits to particular state points according to a running estimate of the frequency with which they are visited. These algorithms are necessarily iterative, their use often requires considerable knowledge about the system under study and, for complex fluids, their convergence can be difficult.

Recently, a new, powerful algorithm has been proposed for direct calculation of the density of states from Monte Carlo simulations [4]. This algorithm has been shown to be particularly efficient in simulations of lattice system [5]. However, the convergence of that algorithm deteriorates with the size and complexity of the simulated system. Perhaps more seriously, as is shown in this work, it reaches an asymptotic accuracy beyond which additional calculations fail to improve the quality of the results.

The random-walk algorithm [4] of Wang and Landau relies on a histogram of energies to dictate the rate of convergence of a simulation. Random trial moves are accepted with probability $p = \min[1, \Omega(E^o)/\Omega(E^n)]$, where E^o and E^n denote the potential energy of the system before and after the move, respectively. Every time that an energy state E is visited, a running estimate of the density of states is updated according to $\Omega(E) = f\Omega(E)$, where f is an arbitrary convergence factor. The energy histogram is also updated; once it becomes sufficiently flat, a simulation "stage" is assumed to be complete, the energy histogram is reset to zero, and the convergence factor f is decreased according to some prescribed function (e.g., $f = \sqrt{f}$). The entire process is repeated until f is very close to 1 (the original literature recommends that $\ln f$ attain a value of 10^{-9}).

Two features of the above procedure must be emphasized. First, because the running estimate of the density of states changes at every step of the simulation, detailed balance is never satisfied. In practice, however, the convergence factor decreases exponentially and its final value can become so small as to render the violation of detailed balance inconsequential. The second feature, which has been overlooked in the literature, is more concerning: because of the decreasing convergence factor, configurations generated at different stages of the simulation do not contribute equally to the estimated density of states. In fact, in the late stages of the simulation, the convergence factor is so small that the corresponding configurations make virtually no contribution to the density of states estimate. In other words, many of the configurations generated by the simulation are not utilized effectively.

In this work, two new methods are proposed to determine the density of states of an arbitrary system. These techniques constitute a departure from earlier algorithms in several, important respects. First, and most importantly, a running estimate of the density of states is determined from the instantaneous temperature of the system, as opposed to a histogram of random visits to distinct energy of states. Second, for a given level of accuracy, the proposed methods are shown to be considerably faster than existing algorithms. Third, unlike

0031-9007/03/90(3)/035701(4)$20.00

previous techniques, the proposed methods are shown to be capable of providing estimates of the density of states of arbitrary accuracy.

The temperature of a system is related to its density of states $\Omega(N, V, E)$ by Boltzmann's equation [6]

$$\frac{1}{T} = \left(\frac{\partial S}{\partial E}\right)_V = k_B\left[\frac{\partial \ln\Omega(N, V, E)}{\partial E}\right]_V. \tag{1}$$

The above equation can be integrated to determine the density of states from knowledge of the temperature:

$$\ln\Omega(N, V, E) = \int_{E_0}^{E} \frac{1}{k_B T}\, dE. \tag{2}$$

Equation (2) requires that the temperature be known as a function of energy. In this manuscript, two completely different schemes are proposed to do that. The first involves use of the so-called configurational temperature [7]. The second relies on knowledge of the kinetic energy in a purposely designed statistical mechanical ensemble.

Recently, Evans *et al.* [7] have shown that an intrinsic temperature can be assigned to an arbitrary configuration of a system. This so-called "configurational temperature" is based entirely on configurational information and is given by

$$\frac{1}{k_B T_{\text{config}}} = \frac{\langle -\sum_i \nabla_i \cdot \mathbf{F}_i\rangle}{\langle \sum_i |\mathbf{F}_i|^2\rangle}, \tag{3}$$

where $\mathbf{F}_i$ represents the force acting on particle i. This configurational temperature can be particularly useful in Monte Carlo simulations, where kinetic energy is not explicitly involved.

In the first scheme proposed here, the temperature is calculated as a function of energy by introducing a configurational-temperature estimator. Four histograms are collected during a simulation; one for the density of states, one for the potential energy, one for the numerator of Eq. (3), and one for its denominator.

In the early stages of the simulation, detailed balanced is grossly violated as a result of the large value of f. The resulting estimates of configurational temperature are therefore incorrect. In order to avoid contamination of late stages by the transfer of incorrect information, the temperature accumulators can be reset at the end of the early stages, once the density of states is calculated from the temperature. For small enough convergence factors (e.g. $\ln f < 10^{-5}$), the slight violations of detailed balance incurred by the method become negligible, and the temperature accumulators need no longer be reset at the end of each stage.

In this proposed random-walk algorithm, the dynamically modified density of states only serves to guide a walker through configuration space. The "true," thermodynamic density of states is calculated from the configurational temperatures accumulated in the simulation. Because all configurations generated in the simulation contribute equally to the estimated configurational temperature, the new algorithm eliminates the problem alluded to earlier, namely, that of nonuniform contributions of different stages to Ω.

Most Monte Carlo simulations are conducted at constant temperature, and kinetic energy is not specifically involved. It is still possible, however, to construct a microcanonical-ensemble Monte Carlo simulation [8] in which a kinetic energy (and hence a temperature) is calculated for a specific value of the total energy.

Consider a microcanonical-ensemble comprising N particles, having volume V, and total energy E. The probability of observing a configuration $\mathbf{x}$ having total potential energy $U(\mathbf{x})$ is proportional to

$$\pi(\mathbf{x}) \propto [E - U(\mathbf{x})]^{(\ell-2)/2}\Theta[E - U(\mathbf{x})], \tag{4}$$

where ℓ is the total number of degrees of freedom of the system, and Θ is the unit step function. Temperature can then be estimated according to

$$T = \left\langle \frac{(E - U)/k}{\ell/2} \right\rangle_{NVE}. \tag{5}$$

To evaluate the density of states from temperature, T must be determined over a range of total energy. To do so, a new ensemble is constructed in which the total number of particles and the volume of the system are fixed. The total energy of the ensemble, however, is allowed to fluctuate in a specified range or "window." To facilitate sampling, we also require that the distribution of the total energy be uniform over the specified energy window. This can be achieved by introducing a weighting factor which is inversely proportional to the density of states. In such an ensemble, the probability to observe a configuration $\mathbf{x}$ having total energy E and total potential energy $U(\mathbf{x})$ is given by

$$\pi(E, \mathbf{x}) \propto \frac{[E - U(\mathbf{x})]^{(\ell-2)/2}\Theta[E - U(\mathbf{x})]}{\Omega(N, V, E)}. \tag{6}$$

In the remainder of this manuscript, we refer to this ensemble as "multimicrocanonical." In the simplest implementation of a simulation in this ensemble, the following two types of moves are employed: simple trial particle displacements, in which the coordinates of a randomly chosen particle are altered by a small random amount, and trial energy changes. The procedure to generate particle displacements is the same as that used in conventional Monte Carlo simulations, except that here the total energy of the system is conserved. A trial move is accepted with probability

$$p_1 = \min\left\{1, \frac{[E - U(\mathbf{x}^n)]^{(\ell-2)/2}\Theta[E - U(\mathbf{x}^n)]}{[E - U(\mathbf{x}^o)]^{(\ell-2)/2}\Theta[E - U(\mathbf{x}^o)]}\right\}. \tag{7}$$

Trial energy changes are proposed by introducing a random total energy change, uniformly distributed in an interval $[-\delta E_{\max}, \delta E_{\max}]$. The maximum change in energy $\delta E_{\max}$ is specified as an input parameter. Trial energy-change moves do not alter the configuration or the potential energy of the system. Such moves are accepted with probability

$$p_2 = \min\left\{1, \frac{\Omega(N, V, E^o)}{\Omega(N, V, E^n)} \times \frac{[E^n - U(\mathbf{x})]^{(\ell-2)/2}\Theta[E^n - U(\mathbf{x})]}{[E^o - U(\mathbf{x})]^{(\ell-2)/2}\Theta[E^o - U(\mathbf{x})]}\right\}. \quad (8)$$

The sought-after density of states enters the acceptance criteria. To gradually refine a running estimate of Ω, we adopt a scheme similar in spirit to that followed in multicanonical ensemble simulations. A kinetic energy accumulator is maintained for each energy level. At the beginning of the simulation, the density of states is initialized to be constant over the entire energy range of interest. The density of states is only updated at the end of each simulation stage, by integrating the temperatures accumulated up to that point. For a simple Lennard-Jones fluid, each simulation stage typically consists of several hundred thousand steps. Since the density of states remains unchanged in each stage, detailed balance is always fulfilled and the temperature accumulators need not be reset at the end of every stage. The simulation is terminated when the statistical errors in the accumulated temperature are sufficiently small.

In the remainder of this Letter, the results of the two proposed methods are compared to those of the random-walk algorithm introduced by Wang and Landau [4] for a truncated-and-shifted Lennard-Jones fluid (the potential energy is truncated at $r_c = 2.5\sigma$).

Our simulations are conducted in a cubic box of length $L = 8\sigma$, with $N = 400$ particles. The density of the system is $\rho = 0.78$, well within the liquid regime. For the configurational-temperature algorithm, the energy window is set to $-1930 \le E/\epsilon < -1580$; for the multimicrocanonical ensemble simulation, the energy window is $-1500 \le E/\epsilon < -500$. In both cases, the energy window corresponds roughly to temperatures in the range $0.85 < T^* < 1.5$ (i.e. above and below the critical point). In configurational-temperature simulations, the calculations are started with a convergence factor $f = \exp(0.1)$. The simulation is terminated when the convergence factor satisfies $f < \exp(10^{-10})$. After necessary transformations, the calculated configurational temperature and "kinetic" temperature are nearly indistinguishable.

To estimate the statistical errors in the estimated density of states, seven independent simulations are conducted, with exactly the same code, but with different random-number generator seeds. Figure 1 shows the statistical error of the density of states as a function of f. The diamonds show results from the Wang and Landau algorithm. These errors exhibit two distinct regimes. For large f (e.g. $f > 10^{-4}$), the statistical error is proportional to $\sqrt{f}$. In the small f regime ($f < 10^{-6}$), the error curve levels off, and asymptotically converges to a limiting value of approximately 0.1. Figure 1 is in direct contradiction with the common belief that the error in the density of states incurred by traditional random-walk algorithms is proportional to f. There are two reasons behind such an asymptotic behavior: first, in the Wang-Landau algorithm, the definition of "flatness" of a histogram is somewhat arbitrary. It is common practice to consider a histogram to be "sufficiently flat" when its minimum entry is no less than 80% of its average value. This assumption can cause problems, because an estimate of density of states with a minor error can result in a sufficiently flat histogram anyway. Second, configurations generated at different stages of the simulation do not contribute equally to the histograms. The Wang and Landau algorithm leaves the density of states essentially unchanged once f is reduced to less than 10^{-6}; additional simulations with smaller f only "polish" the results locally, but do little to improve the overall quality of the data. If phase space has not been ergodically sampled by the time f reaches about 10^{-6}, the final result of the simulation is likely to be inaccurate. Using a more stringent criterion for flatness only alleviates the problem partially, because the computational demands required to complete a stage increase dramatically. Figure 1 also shows that by increasing the flatness criterion to 90% one can decrease the asymptotic error down to about 0.04. However, this also increases the simulation time by a factor of 4.

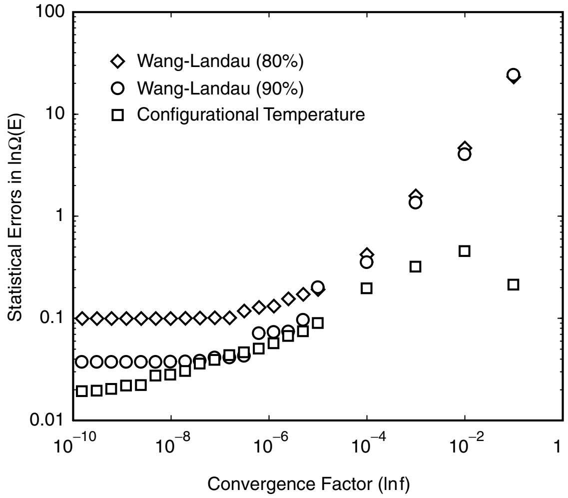

FIG. 1. Statistical errors in the density of states as a function of convergence factor.

In contrast to the curve generated through the conventional random-walk, the errors in the configurational-temperature calculation steadily decrease as the simulation proceeds. Figure 1 also shows that the statistical errors from the configurational-temperature method proposed here are always considerably smaller than those from the Wang and Landau technique. At the end of a

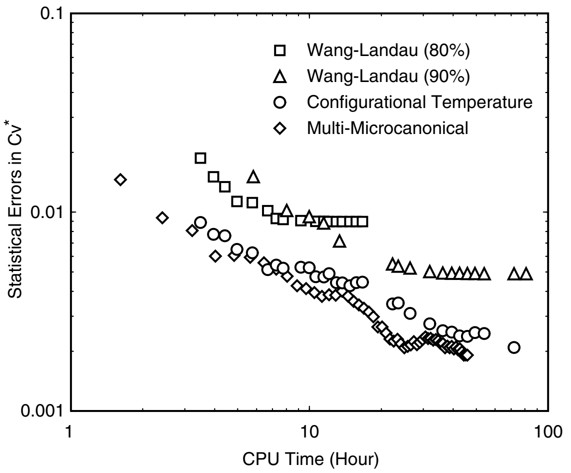

FIG. 2. Statistical errors in heat capacity as a function of CPU time.

simulation, the statistical error from a Wang-Landau calculation is approximately five times larger than that from configurational temperature. In other words, thermodynamic-property calculations of comparable accuracy would be much faster (approximately 25 times) in the proposed configurational algorithm than in existing random-walk algorithms.

Fig. 2 compares the average statistical errors in the heat capacity calculated from a Wang-Landau algorithm, from multimicrocanonical ensemble simulations, and from configurational temperature, as a function of actual CPU time. The error in the proposed two methods is significantly smaller than that of the original Wang-Landau algorithm. The error from multimicrocanonical ensemble simulations is slightly smaller than that from configurational-temperature calculations. The heat capacities (C_v/k_B) at $T^* = 1.0$ calculated from the three methods after a day of computer time are 2.35 ± 0.01, 2.342 ± 0.004, and 2.342 ± 0.003, respectively. Increasing the flatness parameter to 90% decreases the asymptotic error, but increases significantly the computational time required to achieve the same accuracy.

There are three main reasons behind the improved performance of the proposed methods. First, these two methods rely on direct configurational information to construct the density of states, as opposed to a histogram of stochastic visits to each distinct energy level. In other words, more information is utilized in these methods than in the original random- walk algorithm. Second, from the numerical point of view, the integration of temperature to determine a density of states eliminates much of the statistical noise in these calculations. Finally, as mentioned earlier, all stages of the simulation contribute equally to the construction of the density of states.

In summary, two new methods have been proposed to calculate the density of states of simple fluids from Monte Carlo simulations. These methods rely on knowledge of the instantaneous temperature to arrive at the density of states. In one implementation, the temperature is determined from the gradient of the forces. In a different implementation, the temperature is estimated from the kinetic contribution to the total energy. Both of these methods are considerably faster than recently proposed "random-walk" techniques for simulation of the density of states. Furthermore, they permit calculation of the density of states to arbitrary accuracy, something that was not possible in previous algorithms.

Of the two techniques proposed here, the multimicrocanonical ensemble formulation appears to be slightly superior. We believe this second implementation is more convenient in that the derivatives of the forces are not required, and it can also be applied to discontinuous potential energy functions or systems on a lattice, where Eq. (3) is not applicable.

The proposed methods are general and can be easily extended to multidimensional simulations by replacing the temperature with appropriate derivatives of the free energy with respect to the relevant extended variables [9]. For example, in a system having fluctuating energy and number of particles, a two-dimensional integration would be required over temperature and chemical potential of the system. Further studies are under way to explore the performance of such extensions.

The methods proposed here can also be combined with parallel tempering formalisms. Because temperature is an intensive quantity, it is possible to set up a global accumulator for all the replicas in a parallel simulation, thereby greatly simplifying the combination of simulation results in such calculations. The details of this and other combinations will be reported in a more extensive manuscript.

Financial support from the Department of Energy (DE-FG02-99ER14961) is gratefully acknowledged.

---

*Electronic address: depablo@engr.wisc.edu

[1] D. Frenkel and B. Smit, *Understanding Molecular Simulation—From Algorithms to Applications* (Academic Press, New York, 1996).
[2] B. Berg and T. Neuhaus, Phys. Lett. B **267**, 249 (1991).
[3] C. J. Geyer and E. A. Thompson, J. Am. Stat. Assoc. **90**, 909 (1995).
[4] F. Wang and D. P. Landau, Phys. Rev. Lett. **86**, 2050 (2001).
[5] F. Wang and D. P. Landau, Phys. Rev. E **64**, 056101 (2001).
[6] D. A. McQuarrie, *Statistical Mechanics* (Harper Collins Publishers Inc., New York, 1976).
[7] B. D. Butler, G. Ayton, O. G. Jepps, and D. J. Evans, J. Chem. Phys. **109**, 6519 (1998).
[8] R. Lustig, J. Chem. Phys. **109**, 8816 (1998).
[9] Q. Yan, R. Faller, and J. J. de Pablo, J. Chem. Phys. **116**, 8745 (2002).