

A method to bridge different-level coarse-grained models by estimating free energies of high-dimensional conformations: jump-in-sample simulations

Xin Zhou ^{*,1,2} Yi Jiang ^{†,2} Arthur F. Voter,² Steen Rasmussen,¹ and Hans Ziock¹

¹*Earth and Environmental Science Division,*

Los Alamos National Laboratory, Los Alamos, NM 87545, USA

²*Theoretical Division, Los Alamos National Laboratory, Los Alamos, NM 87545, USA*

(Dated: November 13, 2007)

Abstract

We present an efficient algorithm, which uses a random jumping walk in the sample of coarse-grained (CG) conformational spaces, to bridge CG models of different levels. The method iteratively estimates the free energies of these high-dimensional CG conformations. The effective potential of the CG model is fitted from the free energies. The method can be used to construct CG models, as well as to evaluate and correct any existing CG models. We test the method in a Tetrahedral Molecular fluid to evaluate the existing CG molecular model and to construct the intermolecular effective potential. This method not only works more efficiently than the Wang-Landau algorithm for calculating the density of state or free energy, but also can work in high-dimensional spaces where the existing methods fail.

PACS numbers: 02.70.Tt, 61.20.Ja, 61.25.Em

* Email: xzhou@lanl.gov

† Email: jjiang@lanl.gov

I. INTRODUCTION

Computational simulations including Monte Carlo (MC) and molecular dynamics (MD) simulations are standard tools for providing equilibrium and dynamical properties of systems in materials science, chemistry, physics and biology. The fundamental limitation of these simulations is that the accessible time and spatial scales are too small to study many interesting macroscopic phenomena and processes. Recent attempts to overcome this limitation include the enhanced sampling techniques [1–9] for achieving more complete equilibrium samples, the slow-dynamic techniques [10–15] for long-time kinetics or dynamics, and many kinds of multiscale methods [16–21] for studying large-spatial (sometimes also large-time) properties. The coarse-graining method, which averages some degrees of freedom of the original system, is one of the multiscale techniques. For example, in the simulations of polymers, some bonded atoms may be mapped into superatoms that interact through an effective potential [18, 22–27], such that larger systems can be simulated. This coarse-graining method by lumping atoms is found to not only quantitatively reproduce the equilibrium properties of the polymers, but also qualitatively (even semi-quantitatively) reproduce the kinetics and/or dynamics of the original systems.

The coarse-graining method can be formally described as follows. We have an original atomic system with N atoms, whose potential energy is $V(\mathbf{R}^N)$, and \mathbf{R}^N is the $3N$ -dimensional conformation vector. By lumping some of the atoms into superatoms, we have a coarse-grained (CG) model with n superatoms, $n < N$. The CG model has an effective potential $V_{eff}(\mathbf{X}^n)$ in a CG-variable space \mathbf{X}^n , where \mathbf{X}^n is the $3n$ -dimensional conformation vector. The position of a superatom is usually defined as the position of the center of mass of the corresponding groups of atoms, $\mathbf{X}_i = \frac{1}{M_i} \sum_{\alpha} m_{i\alpha} \mathbf{R}_{i\alpha}$, where $M_i = \sum_{\alpha} m_{i\alpha}$, $m_{i\alpha}$ and $\mathbf{R}_{i\alpha}$ are the mass and position of the α th atom of the i th superatom, respectively. Thus, the original atomic conformation \mathbf{R}^N can be rewritten as the CG conformation \mathbf{X}^n and the remaining coordinates \mathbf{Y}^{N-n} under a linear transformation. The CG variables x and the corresponding effective potential $V_{eff}(x)$ form a complete CG model of the original system. To reproduce the thermodynamics of the original system, the effective potential should be equal to the Landau free energy $F(x)$,

$$F(x) = -k_B T \ln \int d r e^{-\beta V(r)} \delta(x - x(r)), \quad (1)$$

where $\beta = 1/(k_B T)$. However, $F(x)$ is often hard to obtain in the $3n$ -dimensional space x , if

n is large. Canonically, we presume some formula for the effective potential with a few free parameters, then determine these parameters by requiring that the CG model reproduces some important thermodynamic properties of the original system. The most commonly used thermodynamic properties are the equation of states, and the radial distribution function (RDF). These properties of the original system can be obtained either from atomistic simulations of the original system or from experiments.

We aim to improve this method for generating the effective potentials of the CG models for the following reasons. (1) The choice of the thermodynamic properties is arbitrary. Usually, only a few properties are used in the construction of the CG models, thus the fitted effective potentials depend on the particular choice. (2) These fitted properties are the result of macroscopic (*i.e.* ensemble) averaging. Thus the effective potential might not correctly reproduce the microscopic distribution of the corresponding physical variables. (3) These thermodynamic properties result mainly from important (stable) conformations, the CG model might not characterize interesting transition regions. In principle, we do not expect the CG models to reproduce exactly the kinetics or dynamics of the original systems. However, we hope that the constructed CG models can detect the kinetics or dynamics at least qualitatively. (4) There is as yet no general method to evaluate and refine the CG models.

In the paper, we present a Monte Carlo method, the jump-in-sample (JIS) method, to bridge different-levels of CG models. The method can efficiently calculate the free energies of any high-dimensional samples over very large free-energy range. By fitting the effective potential from these free energies, we can construct the CG model from the original system or the CG models with finer resolution. The JIS method can evaluate the CG model and refine the results of the CG model. We test the method in both a low-dimensional (1D) system and a high-dimensional tetrahedral molecular liquid.

II. THEORIES AND METHODS

A. Free Energies in High-Dimensional CG Space

Take any sample of the CG model, $\{x^i\}, i = 1, \dots, M$, where each x^i is a conformation of the $3n$ -dimensional CG space. No matter what distribution the sample satisfies, it is possible

to organize a MC random walk in accordance with the standard Metropolis algorithm to calculate their free energies, $F(x^i)$. Recall that the original conformation R^N can be divided into the CG conformation x and the remaining coordinates y . Two types of MC steps could occur: (1) random-walk in y with the fixed x^i , with the standard acceptance criterion: $acc(x^i, y \rightarrow y') = \min(1, \exp\{-\beta[V(x^i, y') - V(x^i, y)]\})$; (2) jump from x^i to x^j with fixed y . The acceptance probability of the jump transition is

$$acc((x^i, y) \rightarrow (x^j, y)) = \min(1, \exp\{-\beta[(V(x^j, y) - V(x^i, y)) - (U_j - U_i)]\}). \quad (2)$$

Here $V(x, y)$ is the potential of the original system, and U_i is the potential defined in the CG sample space $\{x^i\}$. The initial conformation is randomly chosen from $\{x^i\}$ and a random y .

In the simulation, the probability of visits in a sample space (*i.e.* the visit histogram) is, $p_i \propto \exp\{-\beta[F(x^i) - U_i]\}$, thus we have the free energy of x^i

$$F(x^i) = U_i - k_B T \ln p_i + const. \quad (3)$$

In principle, $F(x^i)$ is independent of the choice of U_i . In practice, however, the free energies from different samples may be very different, thus finite-length simulation might not be enough to get a good estimate of $F(x^i)$ because of the exponential dependence of p_i on $F(x^i)$. To improve this situation, we can choose U_i such that all p_i are in the similar order. For example, if we choose $U_i = F(x^i)$, all conformations in the sample will be visited equally, and p_i will be completely flat. But since $F(x^i)$ is unknown, we will gradually and iteratively adjust U_i to make p_i flat, so that all x^i can be visited enough times and that $F(x^i)$ can be well estimated. We call this method, a random jump in sample combined with an iterative method for estimating the free energy, jump-in-sample (or JIS) method.

Many methods can be used to determine U_i . In the beginning of the simulations, we can choose a simple initial U_i , e.g., $U_i \equiv 0$, and then iteratively improve the choice. We first run a segment of simulation to collect the histogram of the visited states, p_i . We then increase U_i by $\delta V = -k_B T \ln p_i$ for all nonzero p_i . Under the new U_i , we run a new segment of simulation, collect the histogram, and update U_i . Repeat this process, the histogram p_i will become flatter and flatter. We will have an acceptable U_i when the flatness of p_i is acceptable. This iterative method to determine U_i has already been used in the generalized hyperdynamics to capture the slow dynamics of entropy-dominated systems [14]. Here, because our aim is

to improve sampling and not to keep dynamical information, we can adjust U_i with more flexibility. We can run a segment of simulation and decrease U_i by a positive δV after each trial jump (or even after a few trial jumps). When a trial jump from x^i to x^j is accepted, we decrease U_j by δV ; otherwise, we decrease U_i by δV . Repeat this process till the histogram is flat. This algorithm is similar to both the Wang-Landau algorithm [28, 29] for calculating the density of state in the energy space and the metadynamics method [6] for calculating the free energy in low-dimensional collective-variable space. While both these methods only work in low-dimension spaces, our JIS method can calculate the free energies of any sample points in high-dimensional spaces.

Similar to the Wang-Landau algorithm [28], in the JIS simulation, we first use a large δV , e.g. $1k_B T$, in order to quickly visit all conformations x^i . If some conformations x^i are never visited even in a very long simulation under a large δV , the free energies of these conformations must be far larger than that of others and than the range we are interested in. We simply discard these conformations that are useless in the construction of our CG model. When all the relevant x^i are visited sufficiently and we have collected a reasonable histogram, we terminate the short run and decrease the δV to, e.g. $\delta V/4$, and repeat the process. In the standard Wang-Landau algorithm, δV keeps decreasing as the iteration proceeds until δV approaches zero ($10^{-8}k_B T$), when the simulation is terminated and the final U_i corresponds to the free energies $F(x^i)$. This method takes increasingly long simulation time for each iteration and the last few iterations are the bottleneck. In our simulations, after a few beginning Wang-Landau-like strages by using some larger δV (e. g. $\delta V = 1, 1/4, 1/16$ or $1/64$ of $k_B T$), we then run segments of normal MC simulations under the fixed U_i and update the free energy $F(x^i)$ by using $\delta V = -k_B T \ln p_i$ after each segment of simulation. This updating scheme is more efficient than the standard Wang-Landau algorithm [30]. More importantly, it satisfies the detailed-balance condition. Our algorithm does not require that the final histogram p_i be very flat unlike the Wang-Landau method. We use a more relaxed criterion to determine the end of the simulation: the minimal histogram p_{min} is larger than a preselected threshold $p_c = 5 \sim 100$. This criterion significantly reduces the simulation time.

The simulation efficiency depends on the acceptance probability of the jump between sample points. In some cases, the acceptance probability of the jump move $(x^i, y) \rightarrow (x^j, y)$ may be very low. For example, when a jump from (x^i, y) to (x^j, y) occurs, we change the

positions of the centers-of-mass of superatoms (x^i to x^j) while keeping y , the relative coordinates of atoms to the centers of mass of their corresponding superatoms, unchanged. Some atoms from different superatoms might overlap and the conformation has to be discarded. The Wang-Landau algorithm can not work for these cases because of the strong correlation between the successive conformations [30]. Our JIS scheme avoids the difficulties of Wang-Landau algorithm by updating U_i after each a segment of simulation rather than each simulation step. However, if the acceptance probability is too small, the current JIS method would require too long simulations to be practical. We can further refine the JIS to overcome this difficulty. For example, we can use intermediate potentials to increase the acceptance probability of the jump move. An intermediate potential is defined as:

$$V(x, y; \alpha) = V_{intra}(y) + \alpha V_{inter}(x, y), \quad (4)$$

with $0 \leq \alpha \leq 1$. Here $V_{intra}(y)$ is the intra-superatom potential, which depends only on the relative coordinates of atoms y ; $V_{inter}(x, y)$ is the inter-superatom interaction. $\alpha = 1$ corresponds to the original system. The acceptance probability of jumps increases as α decreases. An extreme case is when $\alpha = 0$, all jumps will be accepted. With this intermediate potential, the free energy becomes

$$F(x^i) = U_i - k_B T \ln Z(x^i; \alpha) + \text{const}, \quad (5)$$

where

$$Z(x^i; \alpha) = \sum_{k=1}^{n_i} \exp\{-\beta(1 - \alpha)V_{inter}(x^i, y_k)\}. \quad (6)$$

Here n_i is the total number of conformations in $x = x^i$. Since smaller α brings larger statistic errors of free energies, we adjust α until getting enough acceptance probability and smaller statistic errors in special systems.

B. Effective Potential

To better characterize the complete free energy surface $F(x)$, we need to calculate the free energies of a large sample $\{x^i\}$. We can divide the large sample into several sub-samples, and independently (*i.e* parallelly) calculate the free energies in each sub-sample. The arbitrary constant from each sub-sample can be determined by an additional calculation that includes

at least one sample point from each sub-sample. Thus, the final free energies in the large sample can be determined with only one arbitrary constant. This subdivision is very useful when the size of the sample is very large.

Obviously, the choice of sample will affect the effective potential. In general, a good sample should consist of conformations that effectively cover all the important regions of the conformational space. When our initial sample is from a finite-length simulation of the original system, it is often not good enough because the simulation might only cover part of the stable regions. Some recent techniques for enhanced conformational sampling, such as the parallel tempering method [31], can be used to generate better samples. In addition, because the sample is not required to satisfy any particular distribution, we can collect conformations from many simulations under different conditions in order to better cover interesting conformational space.

Once we have the free energy of the sample, we can derive the effective potential of the CG model. In a CG model with identical superatoms, the effective potential is typically assumed to be a pair-wise additive interaction,

$$V_{eff}(x) = 2\pi\rho n \int z^2 g(z; x) u(z) dz, \quad (7)$$

where ρ is the number density of superatoms, n is the number of superatoms, and $u(z)$ is the pair-wise interaction at the inter-superatomic distance z ; $g(z; x)$ is the radial distribution function (RDF) of the CG conformation x . Thus, it is straight forward to fit the interaction potential function $u(z)$ by minimize the cost functional,

$$f[u(z)] = \sum_i [F(x^i) - F_0 - 2\pi\rho n \int z^2 g(z; x^i) u(z) dz]^2, \quad (8)$$

where the constant F_0 is also a fitting parameter. The residual variance of the fitting can be used to evaluate the assumption of the pair interaction. if the variance is too large, we know it is necessary to add multi-body interactions.

C. Refinement of CG models

Canonically, thermodynamic properties are calculated from the ensemble average of the corresponding microscopic variables. A standard MC or MD simulation generates an equi-

librium sample, thus the ensemble average can be replaced by the sample average,

$$\begin{aligned}\langle A \rangle &= \frac{\int dr A(r) \exp[-\beta V(r)]}{\int dr \exp[-\beta V(r)]} \\ &= \frac{1}{M} \sum_i A(r^i),\end{aligned}\tag{9}$$

where M is the size of sample $\{r^i\}$, $A(r)$ is the macroscopic variable A as a function of conformation r . If we have an equilibrium sample $\{r_0^i\}$ of another potential surface $V_0(r)$, the ensemble average of the original system should be calculated by a reweighting process,

$$\langle A \rangle = \frac{\sum_i A(r_0^i) w(r_0^i)}{\sum_i w(r_0^i)},\tag{10}$$

where the weight factor, $w(r_0^i) \propto \exp\{\beta[V_0(r_0^i) - V(r_0^i)]\}$.

This means that we can use a cheap potential $V_0(r)$ to replace the original expensive potential $V(r)$ and gain significant speedup. For example, if we run simulation M steps in $V(r)$, we need to calculate $V(r)$ at each step. However, these M calculations only generate M/m sample points, because we take one sample point every $m = 100 - 1000$ steps to ensure that the conformations are uncorrelated. If we simulate in V_0 for M steps, we will generate M/m sample points that can be used to calculate M/m potential $V(r)$, from which we will calculate the weight factors. Then using Eq. 10, we reproduce the ensemble average. If the computed cost of V_0 is negligible in comparison with that of $V(r)$, we achieve almost m -times speedup.

This result can be expanded to reweight CG sample. We only need to change the weight factor to

$$w(x^i) \propto \exp\{\beta[V_{eff}(x^i) - F(x^i)]\},\tag{11}$$

where $F(x^i)$ and $V_{eff}(x^i)$ are the free energy and the effective potential in the equilibrium sample $\{x^i\}$, respectively. We can use this weight factor to evaluate the effective potential. For example, if we are given a CG model with the effective potential V_{eff} , we can generate a CG sample to calculate the ensemble average of any function $A(x)$ to compare with that from the original system. Alternatively, we can also compute the free energy using the CG sample, then evaluate the effective potential by looking at the weight factor. The larger is the fluctuation of the weight factor, the worse is the CG effective potential in representing the original system. Thus, we can correct the results of any CG model by calculating the

weighting factors. The fluctuation of the weight factors measures the quality of the CG model.

Besides the direct reweighting correction, it is also possible to refine the CG sample by a refined JIS simulation. The refined simulation is similarly organized into two types of MC random walk: local y displacement and the jump among the CG sample $\{x^i\}$. The acceptance probability of the jump is,

$$acc((x^i, y) \rightarrow (x^j, y)) = \min(1, \exp\{-\beta[V(x^j, y) - V(x^i, y) - V_{eff}(x^j) + V_{eff}(x^i)]\}). \quad (12)$$

An important difference between the current refined simulation and the calculation of free energies is that $V_{eff}(x)$ is fixed to be the effective potential of the CG model, rather than the arbitrary adjust parameters U_i in Eq.(3). In other word, the refined simulation generates an equilibrium sample of the original system $V(x, y)$ from an CG sample of the special effective potential $V_{eff}(x)$. This equilibrium sample can be used to measure the difference between the CG model and the original system. However, the refined simulation does not generate new CG conformations, so its results are usually similar to the direct reweighting correction.

We can further improve the refined simulation by relaxing x to the neighborhood of the CG sample $\{x^i\}$. The generated sample in the x -relaxed refined simulation satisfies the equilibrium distribution of the original system $V(x, y)$. An advantage of the x -relaxed refined simulation is the correlation between the successive generated conformations is usually small due to the jumping. Two types MC steps are (1) locally displace both x and y , and (2) jump x from the neighborhood of one sample point to that of another. The accepted probability of the two types MC steps can be written in a single formula,

$$acc((x, y) \rightarrow (x', y'); i, j) = \min(1, \exp\{-\beta[V(x', y') - V_0(x^j) - V(x, y) + V_0(x^i)]\}), \quad (13)$$

where a jumping move is related to a randomly chosen index pair (i, j) with $i \neq j$, the new CG coordinates is calculated from $x' = x + x^j - x^i$. The local displacement corresponds to $i = j$. More details of the x -relaxed simulations will appear elsewhere.

III. RESULTS

A. Single Diatomic Molecule in 1D Double-Well Potential

To illustrate our JIS method, we consider a simplest model: one two-atom molecule in a 1D double-well potential. The bond between the atoms follows the Hooks Law with an elastic coefficient k and an equilibrium length a . We have

$$V(x_1, x_2) = \frac{\Delta E}{2}[(x_1^2 - 1)^2 + (x_2^2 - 1)^2] + \frac{1}{2}k(|x_1 - x_2| - a)^2, \quad (14)$$

where x_1 and x_2 are the position of the two atoms, respectively. We can coarse-grain this system by considering only the position of the center of molecule, $x = (x_1 + x_2)/2$. The remaining variable is $y = x_1 - x_2$.

When we choose a large value for ΔE , we establish a high barrier separating the two potential wells. Now we run a finite-length traditional MD (or MC) simulation in the original system, and project the conformations to the CG variable space, we have generated a CG sample. The sample will only cover part of these stable regions, from which we can not get the correct, complete effective potential. As we mentioned eariler, the sample which are used to calculate their free energies is not required to satisfy any special distribution, we can generate the sample under any potential. We generate an equilibrium sample under a potential $V_0(x) = x^2$ and calculate the free energies in the sample. In the simple example, the equilibrium sample of $V_0(x) = x^2$ is good enough since it cover all the whole interesting conformational space. By iteratively adjusting U_i through the JIS method as in Eq.(3) we can visit all conformations in the sample to estimate their free energies.

Fig.(1) shows the results of a JIS simulation without using any intermediate potential ($\alpha = 1$). The energy barrier, $\Delta E = 156.25 \text{ k}_B\text{T}$, is high. the bond parameters are $k = 20 \text{ k}_B\text{T}$, and $a = 0.1$. The free energies computed from the CG model faithfully reproduce that of the orginal system over a large range ($0 \sim 300 \text{ k}_B\text{T}$) with high accuracy (the error is smaller than $1 \text{ k}_B\text{T}$), except for the region near $z = 0$. In the original double well potential, $z = 0$ is a unstable extremum. The measured free energy surface shows a stabe state around $z = 0$, which corresponds to a bond-expanded state due to the finite-strength bond connected the two atoms. This discrepancy is expected because in the CG model we neglect the intramolecular degree of freedom completely.

For comparison, in the inset of Fig.(1), we also show the free energies by using an intermediate potential defined in Eq.(4) with $\alpha = 0.2$. The difference between these two methods is negligible, which means that the statistic errors from the intermediate potential is ignored.

B. Liquids of Tetrahedral Molecules

We now consider a more complex system of n tetrahedral molecules consisting of $q = 4$ atoms of the same mass m_0 [32, 33]. The interaction between all non-bonded atoms follows a truncated and shifted 12 – 6 Lennard-Jones potential with a cutoff r_c ,

$$u(z) = 4\epsilon\left[\left(\frac{\sigma}{z}\right)^{12} - \left(\frac{\sigma}{z}\right)^6\right] - E_c, \quad (15)$$

for $z \leq r_c$, and $u(z) = 0$ otherwise. Here z is the interatomis distance, $E_c = 4\epsilon\left[\left(\frac{\sigma}{r_c}\right)^{12} - \left(\frac{\sigma}{r_c}\right)^6\right]$. Within each molecules, the atoms are bonded via an attractive finite extensible nonlinear elastic (FENE) potential

$$U_{FENE}(z) = -\frac{1}{2}kR_0^2 \ln[1 - (z/R_0)^2], \quad (16)$$

for $z \leq R_0$, and $U_{FENE}(z) = \infty$ otherwise. Here z is the bond length, the divergence length $R_0 = 1.5\sigma$, and stiffness $k = 30\epsilon/\sigma^2$. We use the reduced units by seting σ , ϵ and m_0 as units of length, energy and mass, respectively.

A natural CG model is to treat each molecule as a super-atom, neglecting the atomic details, and use an intermolecular effective potential. Praprotnik *et al.* [32] simulated such a system with the temperature $T = 1$, the number density $\rho = 0.1$ and $r_c = 2^{1/6}$. They required that the CG model reproduce the intermolecular RDF, and fitted the effective potential as the Morse potential,

$$U(z) = \gamma\{1 - \exp[-\kappa(z - r_0)]\}^2, \quad (17)$$

with parameters $\gamma = 0.105, \kappa = 2.4, r_0 = 2.31$ and cutoff at r_0 .

To test our JIS method, we simulate the same system with $n = 864$ molecules. At $T = 1$ and $\rho = 0.1$, we first generate 1000 CG conformations under the Morse potential as described by eq.(17), then calculate the free energies of these conformations. In this low-density case, the distances between the non-bonded atoms are almost always larger the cutoff $r_c = 2^{1/6}$. Therefore almost all trial jumps are accepted and the histogram p_i is nearly flat. The free

energies are found to be in a very small range (about $1k_B T$). However, the potential energy of these conformations are in the range of a few tens of $k_B T$. Thus even though the Morse potential, eq.(17), can reproduce the center-of-mass RDF of atomistic simulations, it cannot match the free energies, and hence is not a good effective potential.

Furthermore, we can use the free energies and Morse potential energies of the conformations to compute the weight factor according to Eq.(11). We then generate a new ensemble-averaged RDF using the weighting factor [Eq.(10)]. This new RDF is nearly the same as the RDF from the original system (see Fig.(2)). It indicates that the two very different effective potential can generate the same RDFs, suggesting that the traditional method to form effective potential is not sufficient. Fig.(2) also shows that the fluctuation of the reweighting factors (or the fluctuation of $F(x^i) - V_{morse}(x^i)$) is huge.

In the current system, the correct effective potential should be, $u(z) = 0$ for $z > 1.3$, rather than the previous Morse potential $u(z) = 0$ until $z > 2.31$. The correct $u(z)$ at $z \leq 1.3$ can be obtained if we calculate the free energies of other conformations where the shortest intermolecular distance is smaller. The current 1000 conformations sampled from the repulsive Morse potential is not sufficient to fit $u(z)$ for $z < 1.3$. We can generate these conformations with methods such as parallel tempering simulations, then fit $u(z)$ over the complete range of z .

To illustrate how the JIS method calculates free energies of a high-dimensional sample, we still use the same 1000 CG conformations, but set the cutoff of interatomic LJ potential $r_c = 2$. In this case, the free energies in the CG sample cover a very large range. Since the exact free energies of the sample are unknown, we run independent simulations by using different control parameters (such as the seed of random number generator, initial conformation, termination condition, and the parameter α of the intermediate potential) to compare the free energies. Each simulation needs about $10^8 \sim 10^9$ MC steps (1 ~ 3 days CPU time on a 2.8 GHz single processor). The results of two independent simulations are shown in Fig.(3). Over a large range (about $150 k_B T$), the free energies from two different simulations differ only by an additive constant; the error is less $1k_B T$. By fitting to eq.(7), we obtain the pair interaction (see Fig.(4)). Because the applied CG conformations come from a simulation of the repulsive Morse potential, the minimal intermolecular distance in all these CG conformations is larger (about 1.3) than the equilibrium intermolecular distance in the current original system with the attractive interatomic LJ interaction ($r_c = 2$), thus

the CG sample does not allow the pair interaction in the small r region. From the fitted pair interaction at large r region, we can approximate the intermolecular LJ interaction with $\epsilon \sim 8.5$ and $\sigma \sim 1.1$.

thus we can better fit the pair interaction over the whole range of r .

IV. DISCUSSION

We present the JIS method that can calculate the free energies of high-dimensional conformations. The method can be used to fit the effective potentials of CG models or to evaluate and improve the CG models. The JIS method can be viewed as a generalization of the expanded ensemble method [34], which calculate the free energies of a system at different temperatures. In the expanded ensemble method, the system at each temperature is a sub-ensemble and these sub-ensembles form a expanded ensemble. The simulation uses jump walks among the sub-ensembles and use normal random walks inside each sub-ensemble to estimate the free energies of these sub-ensembles. In our method, the sub-ensembles are indexed by the (high-dimensional) CG coordinates rather than the temperature.

The selection of CG sample, which is independent of the JIS method, is important to characterize the whole effective potential energy surface. How to generate a good sample is outside the topic of this paper. Usually, it is sufficient to reproduce thermodynamics of original systems if the CG sample covers almost all the important conformational regions. Adding some interesting transition conformations can increase the ability of the formed CG model in getting kinetics. The samples generated from the parallel tempering simulations or other enhanced conformational techniques in the original system are usually sufficient.

In the jumping walk of the JIS method, conformations of all molecules are changed at the same time, the acceptance possibility of the jumping is usually low. We can use the intermediate potential with small parameter α to overcome the difficulty. The statistic errors of the obtained free energies in the small- α simulation might be significant. However, the estimates are usually sufficient to fit the effective potential. In comparison with other methods for calculating free energies, *e.g.* thermodynamics integration [35], the JIS method can estimate the free energies more efficiently. The JIS method can also be combined with another method to refine our results if necessary. For example, we can calculate the free

energy of any CG conformations from the thermodynamics integral,

$$F(x^i; \alpha = 1) = F(x^i; \alpha_0) + \int_{\alpha_0}^1 d\alpha \langle V_{inter}(x^i, y) \rangle_{\alpha}, \quad (18)$$

where $\langle V_{inter}(x^i, y) \rangle_{\alpha}$ is the ensemble average under the intermediate potential $V(x, y; \alpha)$ described by eq.(4) with the constraint $x = x^i$. For each x^i , we can independently run the constraint simulation with $x = x^i$ and calculate the average intermolecular energy at many α points, then integrate the free energy from $\alpha = 0$ to $\alpha = 1$. Here $F(x^i; \alpha = 0)$ is independent of x^i and $F(x^i; \alpha = 1)$ is the desired free energies. The main errors of the method come from the very large fluctuation of the intermolecular energy at small α . Therefore, we can first calculate $F(x^i; \alpha_0)$ from the JIS method, then obtain good estimates of $F(x^i; \alpha = 1)$ by integrating $\langle V_{inter} \rangle_{\alpha}$ from α_0 to 1, to obtain $F(x^i)$ with high accuracy. Otherwise, we can directly estimate the free energies from the JIS simulation with α_0 .

Acknowledgments

This work was supported by the US DOE under contract No. DE-AC52-06NA25396. We are grateful to H. Chen for stimulating discussions, comments and suggestions.

-
- [1] B. A. Berg and T. Neuhaus, *Phys. Rev. Lett.* **68**, 9 (1992).
- [2] J. Lee, *Phys. Rev. Lett.* **71**, 211 (1993).
- [3] K.-K. Han, *Phys. Rev. E* **54**, 6906 (1996).
- [4] X. Wu and S. Wang, *J. Phys. Chem. B* **102**, 7238 (1998).
- [5] Z. Zhu, M. E. Tuckerman, S. O. Samuelson, and G. J. Martyna, *Phys. Rev. Lett.* **88**, 100201 (2002).
- [6] A. Laio and M. Parrinello, *Proc. Natl. Acad. Sci. U.S.A.* **99**, 12562 (2002).
- [7] D. Passerone, M. Ceccarelli, and M. Parrinello, *J. Chem. Phys.* **118**, 2025 (2003).
- [8] P. Minary, M. E. Tuckerman, and G. J. Martyna, *Phys. Rev. Lett.* **93**, 150201 (2004).
- [9] D. Frenkel, *Proc. Natl. Acad. Sci. U.S.A.* **101**, 17571 (2004).
- [10] A. B. Bortz, M. H. Kalos, and J. L. Lebowitz, *J. Comp. Phys.* **17**, 10 (1975).
- [11] R. Olender and R. Elber, *J. Chem. Phys.* **105**, 9299 (1996).
- [12] P. G. Bolhuis, D. Chandler, C. Dellago, and P. L. Geissler, *Ann. Rev. Phys. Chem.* **53**, 291 (2002).
- [13] A. F. Voter, M. F., and G. T. C., *Annu. Rev. Mater. Res.* **32**, 321 (2002).
- [14] X. Zhou, Y. Jiang, K. Kremer, H. Ziock, and S. Rasmussen, *Phys. Rev. E*, **74**, R035701 (2006).
- [15] D. J. Wales, *Int. Rev. Phys. Chem.* **25**, 237 (2006).
- [16] P. J. Hoogerbrugge and J. M. V. A. Koelman, *Europhys. Lett.* **19**, 155 (1992).
- [17] R. Benzi, S. Succi, and M. Vergassola, *Phys. Rep.* **222**, 145 (1992).
- [18] F. Müller-Plathe, *Chem. Phys. Chem.* **2**, 754 (2002).
- [19] M. Neri, C. Anselmi, M. Cascella, A. Maritan, and P. Carloni, *Phys. Rev. Lett.* **95**, 218102 (2005).
- [20] V. Tozzini, *Curr. Opin. Struct. Bio.* **15**, 144 (2005).
- [21] D. I. Kopelevich, A. Z. Panagiotopoulos, and I. G. Kevrekidis, *J. Chem. Phys.* **122**, 044908 (2005).
- [22] C. F. Abrams and K. Kremer, *Macromolecules* **36**, 260 (2003).
- [23] C. F. Abrams, L. Delle Site, and K. Kremer, *Phys. Rev. E* **67**, 021807 (2003).
- [24] L. Delle Site, C. F. Abrams, A. Alavi, and K. Kremer, *Phys. Rev. Lett.* **89**, 156103 (2002).
- [25] L. Delle Site, S. Leon, and K. Kremer, *J. Am. Chem. Soc.* **126**, 2944 (2004).

- [26] X. Zhou, D. Andrienko, L. Delle Site, and K. Kremer, *Europhys. Lett.* **70**, 264 (2005).
- [27] X. Zhou, D. Andrienko, L. Delle Site, and K. Kremer, *J. Chem. Phys.* **123**, 104904 (2005).
- [28] F. Wang and D. P. Landau, *Phys. Rev. E* **64**, 056101 (2001).
- [29] F. Wang and D. P. Landau, *Phys. Rev. Lett.* **86**, 2050 (2001).
- [30] X. Zhou, unpublished, (2007).
- [31] D. J. Earl and M. W. Deem, *Phys. Chem. Chem. Phys.* **7**, 3910 (2005).
- [32] M. Praprotnik, L. Delle Site, and K. Kremer, *J. Chem. Phys.* **123**, 224106 (2005).
- [33] M. Praprotnik, L. Delle Site, and K. Kremer, *Phys. Rev. E* **73**, 066701 (2006).
- [34] A. P. Lyubartsev, A. A. Martsinovski, S. V. Shevkunov, and P. N. Vorontsov-Velyaminov, *J. Chem. Phys.* **96**, 1776 (1992).
- [35] M. J. Mitchell and J. A. MaCammon, *J. Comp. Chem.* **12**, 271 (1991).

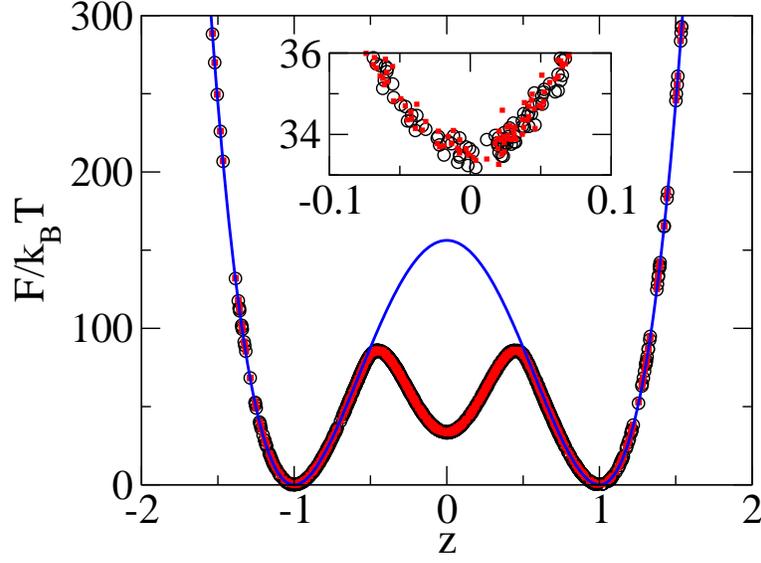


FIG. 1: (Online color). The free energies of a sample in the single diatom molecule. The calculated free energies in the equilibrium sample of $V_0(z) = z^2$ with $\alpha = 0.2$ (opened circles) and $\alpha = 1.0$ (filled squares). If fixing the intramolecular bond length, a direct expected effective potential is the double-well external potential with $\Delta E = 156.25$ (line). The real effective potential (*i.e* free energy) is different from it near $z = 0$. Inset: the details of the calculated free energies near $z = 0$.

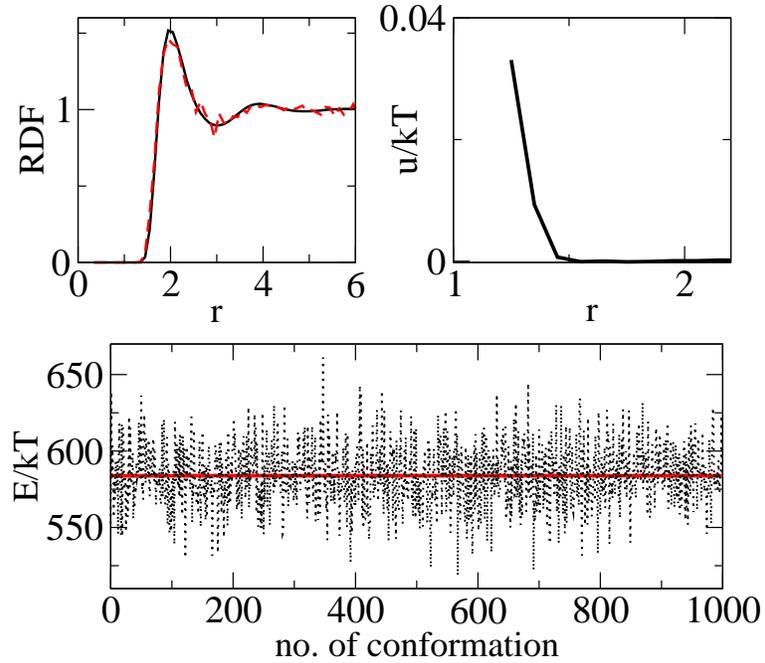


FIG. 2: (Online color). Coarse-graining the tetrahedral molecular liquid. Top left: the radial distribution function (RDF) from direct averaging (solid line) and from reweighting averaging (dashed line) are shown; Top right: the fitted pair interaction u vs the intermolecular distance r . Bottom: the calculated free energies (thick line) and the Morse energies (thin dotted line) of the 1000 CG conformations.

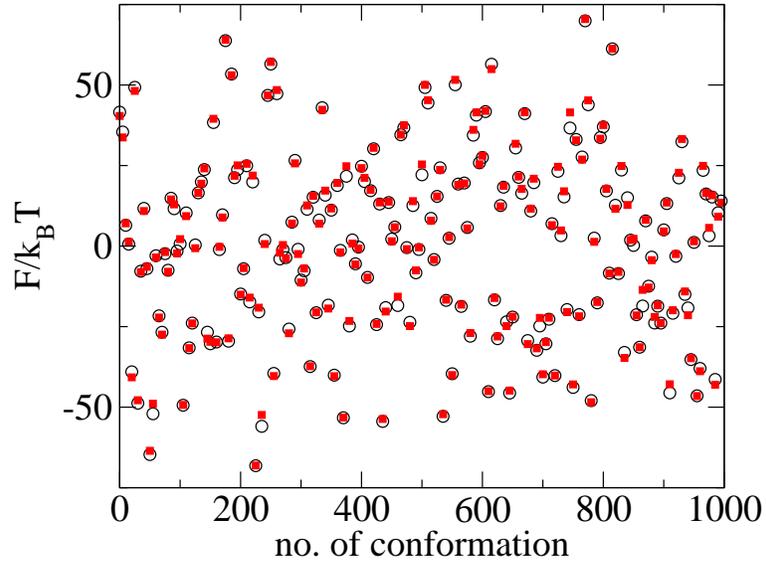


FIG. 3: (Online color). The free energies of some coarse-grained conformations in the tetrahedral molecular liquids are shown. The results of two independent simulations with different control parameters are in consistent each other. Here the free energies from one simulation have been shifted by a constant.

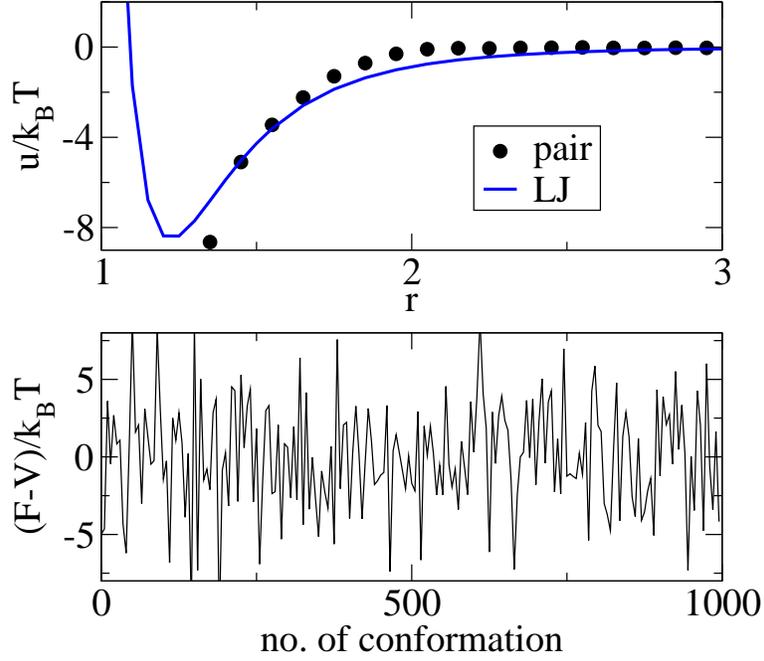


FIG. 4: The fitted intermolecular interaction of the tetrahedral molecular liquids. Top: the fitted pair-additive interactive potential u vs the intermolecular distance r . Since the minimal intermolecular distance in the applied CG sample is larger than 1.3, no $u(r)$ for $r < 1.3$ is available. Solid line is a Lennard-Jones potential with $\epsilon \sim 8.5$ and $\sigma \sim 1.1$ as a guide for the eye. Bottom: the difference between the free energies and the fitted pair potential in all the CG conformations.