ARTICLES

Accuracy of free-energy perturbation calculations in molecular simulation. I. Modeling

Nandou Lu and David A. Kofke[a)]
Department of Chemical Engineering, State University of New York at Buffalo, Buffalo, New York 14260-4200

(Received 6 December 2000; accepted 5 February 2001)

We examine issues involved in applying and interpreting free-energy perturbation (FEP) calculations in molecular simulation. We focus in particular on the accuracy of these calculations, and how the accuracy differs when the FEP is performed in one or the other direction between two systems. We argue that the commonly applied heuristic, indicating a simple average of results taken for the two directions, is poorly conceived. Instead, we argue that the best way to proceed is to conduct the FEP calculation in one direction, namely that in which the entropy of the target is less than the entropy of the reference. We analyze the behavior of FEP calculations in terms of the perturbation-energy distribution functions, and present several routes to characterize the calculations in terms of these distributions. We also provide prescriptions for the selection of an appropriate multistage FEP scheme based on how the important phase-space regions of the target and reference systems overlap one another. © *2001 American Institute of Physics.* [DOI: 10.1063/1.1359181]

I. INTRODUCTION

Free-energy perturbation (FEP) is one of the most widely used methods to compute free energies via molecular simulation.[1–4] A single-stage FEP calculation[5] yields the free-energy difference between two systems as an ensemble average performed on only one of them. The working equation can usually be put in the form

$$e^{-\beta(A_1-A_0)}=\langle e^{-\beta(U_1-U_0)}\rangle_0. \tag{1}$$

The ''0'' and ''1'' subscripts indicate properties for the two systems of interest; A is the Helmholtz free energy, U is the configurational energy, and $\beta=1/kT$ with k the Boltzmann's constant and T the absolute temperature. The angle brackets indicate an ensemble average performed on the 0 system, which we call the *reference*; the 1 system we call the *target*.

Either system of interest can be selected to be the reference, and thus the FEP calculation for a given pair of systems can be performed in either of two directions. If the systems differ in the presence of a single molecule, the FEP calculation yields the chemical potential, and in this case the two implementations are commonly known as the Widom insertion[6,7] and Widom deletion[8] methods. In what follows we will present the idea that in all FEP calculations one of the directions can be considered uniquely as an ''insertion'' calculation while the other represents a ''deletion,'' so we will adopt this terminology here to refer to the two directions of an FEP calculation.

In principle, insertion and deletion methods yield equivalent results for the free-energy difference, but in practice the calculated results differ systematically.[9–16] It is well known that the calculation in one direction typically overestimates the free-energy difference, while calculation in the other direction underestimates it. What is not well appreciated is the asymmetry in the *magnitude* of the over- and underestimations. Many researchers treat the errors in the two methods as being of the same magnitude and opposite sign.[17–26] Therefore, a heuristic has emerged that the best estimate of the free-energy difference should be taken as the average of the insertion and deletion (forward and reverse) free energies. This practice is wrong. It is the aim of this work to highlight this misconception and present a means to analyze the magnitude asymmetry.

The reliability of the FEP calculations has two facets, representing the *precision* and *accuracy*, respectively. *Precision* describes reproducibility of the measurements, and can be characterized by a variance statistic. In a previous study[27] we set up quantitative models to describe the precision of FEP calculations, and showed that precision is closely related to the entropy difference between the reference and target systems. We applied our analysis to develop a prescription for selecting optimal intermediates in staged FEP calculations.

The *accuracy* of a calculation is distinct from its precision. It indicates the correctness of the estimate, or how the measurement result deviates from the *true* or *exact* value in repeated measurements. Arguably, the accuracy of a result is much more important than its precision, yet it usually gets little attention. One reason is that in many situations the inaccuracy of an ensemble average is much less than its imprecision. However, this circumstance often does not hold for FEP calculations, even in cases where other ensemble averages are of good accuracy. Indeed, FEP results could be in good precision but terrible accuracy. A simple (and ex-

a)Electronic mail: kofke@eng.buffalo.edu

treme) example is the deletion calculation of the chemical potential for the hard-sphere (HS) system. According to this method the chemical potential is always zero whatever the system condition. Obviously, this is a wrong answer (except for zero density), but it has perfect reproducibility. Fortunately the result is so obviously wrong that no one would make the mistake of interpreting its good precision as good accuracy. But this situation is anomalous, and comes about because the hard-sphere system has only two possible energy levels (zero and infinity). Inaccuracy is much better disguised in systems only slightly more realistic than the hard-sphere model, even one having only three energy levels.[28] Poor understanding of the accuracy in realistic systems leads to misinterpretations of simulation results and misapplications of simulation methodologies.

In contrast to precision, where the block-average variance provides a good measure, there is no simple way to assess the accuracy of an FEP calculation. In this study we address this problem, and examine the inaccuracy in the FEP calculations as a function of the simulation length and basic characteristics of the target and reference systems. We analyze the source of the simulation error and develop the methodology and models to characterize it using energy distribution functions. We also further study the asymmetry behavior of the FEP calculations and correct some related misconceptions. In the next paper of this series we apply this understanding to develop an easy-to-use heuristic that gauges the inaccuracy of insertion FEP calculations.

We present in Sec. II the analysis and modeling of the inaccuracy. In Sec. III we describe simulations used to verify the inaccuracy model and heuristic. Simulation results and discussions are given in Sec. IV, and concluding remarks in Sec. V.

II. BACKGROUND AND THEORY

A. Generalized insertion and deletion

Our previous studies[13,14,27,28] show that the entropy difference between the reference and target systems is an important quantity in determining both the accuracy and precision of FEP calculations. Therefore, it is convenient to identify the FEP systems according to their entropy. In our study, the higher-entropy system is denoted H and the lower-entropy one is L. We define the difference of a quantity between two systems as $(L)-(H)$. For example, the free-energy, potential-energy, and entropy differences are defined as $\Delta A=A_L-A_H$, $\Delta U=U_L-U_H$, and $\Delta S=S_L-S_H$, respectively; by definition, $\Delta S<0$.

In a Widom insertion calculation for the chemical potential, the reference system can be viewed as one in which the inserted molecule moves around without interference from any of the other molecules in the system. The extra ''freedom'' available to this molecule accords the reference with a higher entropy than the target. Correspondingly, we define generalized insertion and deletion directions for any FEP calculation based on the sign of entropy difference along the perturbation direction. We call an FEP calculation a generalized insertion if the H system is used as the reference, where the entropy difference along the perturbation direction is less than zero, $S_1-S_0<0$ (where the 1 and 0 subscripts indicate the target and reference systems, respectively, as discussed in the Introduction). Similarly, a generalized deletion calculation uses the L system as the reference, and proceeds in a direction of increasing entropy: $S_1-S_0>0$.

In a generalized insertion calculation we have

$$\exp(-\beta\Delta A)=\langle\exp(-\beta u)\rangle_H, \tag{2}$$

where u is the difference of configurational potential energy $U(\mathbf{r}^N)$ between the L and H systems, i.e., $u=U_L(\mathbf{r}^N)-U_H(\mathbf{r}^N)$, the energy change encountered in a trial sampling. For convenience, we simply call u ''potential energy'' or ''energy'' in the following, but one should keep in mind that it really means the potential energy change during the perturbation. The momentum contributions to the Hamiltonian are ignored in this study, assuming they are the same in both systems. The subscripted angle brackets $\langle\ldots\rangle_H$ indicate the canonical-ensemble average sampled according to the Hamiltonian for the higher-entropy system. Similarly, the free-energy difference in a generalized deletion calculation is given by

$$\exp(+\beta\Delta A)=\langle\exp(+\beta u)\rangle_L. \tag{3}$$

Note that Eqs. (2) and (3) are equivalent, as shown by Shing and Gubbins[8,29] in the context of the chemical potential calculation.

B. Entropy and configuration space

The reference and target systems in an FEP calculation share the same phase space, but the important regions of phase space—those that contribute most to their respective partition functions—may differ between the two systems. The entropy of a system is closely related to the size of the important region of phase space. The ''important region'' could be defined, for example, by considering the energy-distribution function $P(U)$

$$P(U)=\frac{1}{Q}\Omega(U)e^{-\beta U}, \tag{4}$$

where Q and Ω are the canonical and microcanonical partition functions, respectively. We could define a threshold value P^* for $P(U)$, such that an integral over only those energies U that have $P(U)$ above the threshold $[P(U)>P^*]$ would yield some substantial fraction (e.g., 99%) of the integral taken instead over all energies. Then, any configuration $\mathbf{r}^N$ having an energy $U(\mathbf{r}^N)$ such that $P[U(\mathbf{r}^N)]>P^*$ would be considered an ''important configuration.'' Note that a configuration may be important if it has a low (large negative) energy, or if it is one of many configurations (large Ω) having a nonprohibitive energy.

In the following analysis, we will consider FEP calculations between systems in which the set of important configurations for the low-entropy system (designated Γ_L) forms a wholly contained subset of the important configurations (Γ_H) of the high-entropy system: $\Gamma_L\subset\Gamma_H$, as shown in Fig. 1. This is the case, for example, for systems of hard spheres in which one sphere is (L system) or is not (H system) interacting with the others. It is very likely that this situation (i.e., $\Gamma_L\subset\Gamma_H$) holds for other molecule-insertion system pairs in-

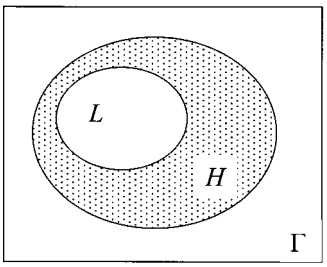

FIG. 1. Schematic diagram of phase space and important configurations in an FEP calculation. The shaded region represents the important configurations for the high entropy system (Γ_H). The white region inside (and part of) Γ_H is the important configurations for the low-entropy system, Γ_L. The significant feature of this diagram is Γ_L is completely inside Γ_H. In real systems the shapes of the *H* and *L* regions are vastly more complex than the simple two-dimensional representations presented here.

volving more realistic and complex potentials. In the event that the important regions do not relate this way, it is unlikely that FEP can be made effective without setting up stages that break the calculation into pieces involving pairs of systems that do follow the relation $\Gamma_L \subset \Gamma_H$. We discuss this matter further in the Discussion section below.

The configurations in Γ_H that do not lie in Γ_L fall outside because their energy is greater than the energies of the configurations in Γ_L. Thus, we expect that configurations in Γ_H will exhibit energy differences u that are typically positive and large. The primary exceptions are those configurations that lie also in Γ_L, where the configurations will tend to have a value of u that is zero or negative. Note that this characterization of the value of u in Γ_H and Γ_L is more of an expectation than a rule; deviations from this description, perhaps significant ones, might be found in particular systems. Much of what follows does not rely on this characterization, but it can provide guidance to the intuition.

C. Energy distributions

We can rewrite the ensemble averages in Eqs. (2) and (3) in terms of a one-dimensional integral weighted by the distribution of energy changes encountered in a simulation, as proposed by Shing and Gubbins.[8] Following their notation, we define $f(u)$ and $g(u)$ as the distribution functions of the potential energy change u encountered in a simulation using insertion method (*H* system as the reference) and deletion method (*L* system reference), respectively. Then, Eq. (2) becomes

$$\exp(-\beta\Delta A) = \int du \exp(-\beta u) f(u), \tag{5}$$

where the integral goes from $-\infty$ to $+\infty$. Likewise, the deletion free-energy formula is given by

$$\exp(+\beta\Delta A) = \int du \exp(+\beta u) g(u). \tag{6}$$

Note that both $f(u)$ and $g(u)$ are normalized; also, physical considerations place a lower bound on the energy for which $g(u)$ is nonzero. A useful relation between the f and g distributions is[8,30]

$$f(u)\exp(\beta\Delta A) = g(u)\exp(\beta u). \tag{7}$$

The difference in the ensemble-averaged energies between the *L* and *H* systems is given as a derivative of the free-energy difference, $\Delta U = \partial(\beta\Delta A)/\partial\beta$. With Eqs. (5)–(7) the result can be expressed in terms of the energy distributions and their temperature derivatives

$$\Delta U = \int du \left[u - \frac{\partial \ln f(u)}{\partial \beta} \right] g(u)$$
$$= \int du \left[u + \frac{\partial \ln g(u)}{\partial \beta} \right] f(u). \tag{8}$$

These temperature derivatives can in turn be interpreted via fluctuation formulas

$$\frac{\partial \ln f(u)}{\partial \beta} = -\left[\frac{\langle U_H \delta \rangle_H}{\langle \delta \rangle_H} - \langle U_H \rangle_H \right], \tag{9}$$

$$\frac{\partial \ln g(u)}{\partial \beta} = -\left[\frac{\langle U_L \delta \rangle_L}{\langle \delta \rangle_L} - \langle U_L \rangle_L \right], \tag{10}$$

where the angle brackets represent an ensemble average in the *H* or *L* system as indicated, and U is the energy in the corresponding system; also, δ is the Dirac delta function applied to the energy change u, for which the ensemble average then gives the distribution functions: $f(u) = \langle \delta \rangle_H$ and $g(u) = \langle \delta \rangle_L$.

These results show that the energy change cannot be obtained from knowledge of the distribution functions at only one temperature. But, it is interesting to consider the degree of correlation between the energy change u and the configurational energy U in the *H* and *L* systems. According to the formulas above, correlations in the *H* system are relevant for ranges of u where $g(u)$ is non-negligible, i.e., for configurations in Γ_L. While these configurations have a favorable u, it is not likely that the energy U_H will differ markedly or systematically from the average $\langle U_H \rangle_H$ in this region. For particle insertion, these are configurations in which a hole has opened up to permit the insertion of a molecule. Except for very high-density systems, opening such a hole need not be associated with a large change in the energy U_H. In contrast, the correlations between u and U_L in the *L* system, over the region Γ_H where $f(u)$ is non-negligible, will be significant. The energy-change u is large in these regions, which indicates that U_L will also be systematically large. From these arguments we anticipate that the β derivative of $\ln f(u)$ will be small, while the derivative of $\ln g(u)$ will be non-negligible. Thus, from Eq. (8) we can approximate the energy change as a simple g integral, but cannot express it in terms of the f integral

$$\Delta U \approx \int u g(u) du. \tag{11}$$

Assuming this result is valid, then the entropy difference can also be expressed in terms of the f and g distributions; thus

$$\Delta S \approx \int g(u) \ln[f(u)/g(u)] du, \tag{12}$$

which, again, is by definition less than zero.

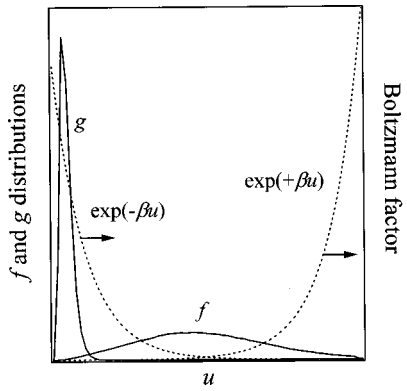

FIG. 2. A plot of typical f and g distributions as functions of potential energy change u. Both distributions are normalized. The associated Boltzmann factor integrands for the f and g distributions are also schematically shown in the plot, as the dashed curves.

Figure 2 shows the typical shape of the $f(u)$ and $g(u)$ distributions, together with the Boltzmann factor that is averaged to obtain the free-energy difference. By looking at the plot it is clear that the important region for the insertion calculation is the low-energy region, while the high-energy region is critical to the deletion calculation. These regions correspond to the tails of the f or g distributions. They are prone to poor sampling in a finite length simulation since they correspond to configurations to be sampled with low probability. So, in terms of the f and g distribution functions, the error in the free-energy calculation comes from the poor sampling of the distribution tails, more explicitly, the low energy tail for insertion calculation and the high energy tail for deletion calculation.

D. Modeling inaccuracy

Let us generalize the problem in the following way. Suppose a quantity X is an average of function $F(u)$ according to weighting distribution $p(u)$, i.e.,

$$X=\int_{-\infty}^{\infty}p(u)F(u)du. \tag{13}$$

In connecting Eq. (13) to FEP calculations, the average quantity X is either $\exp(-\beta\Delta A)$ for the insertion calculation or $\exp(+\beta\Delta A)$ for the deletion calculation; $p(u)$ is the weight function, either $f(u)$ or $g(u)$; and $F(u)$ is $\exp(-\beta u)$ or $\exp(+\beta u)$ for the insertion and deletion calculations, respectively. During the simulation we sample u according to $p(u)$ and finally measure the average X as a sample average of $F(u)$. It is clear that the inaccuracy or error in the average X depends on how well $p(u)$ is sampled. For a long but finite length simulation, the high-probability region of $p(u)$ will be well sampled, while the low-probability regions, or the wings of the distribution, are generally not well sampled. This situation is shown schematically in Fig. 3(a), where the shaded regions indicate energies that are not well sampled. For any given sample, one can identify energy values, u_1 and u_2, such that u is never sampled for the regions $(-\infty,u_1)$ and $(u_2,+\infty)$. Each such energy is referred to as a *limit energy* in the following. The exact X is given by the averaging over whole range of u

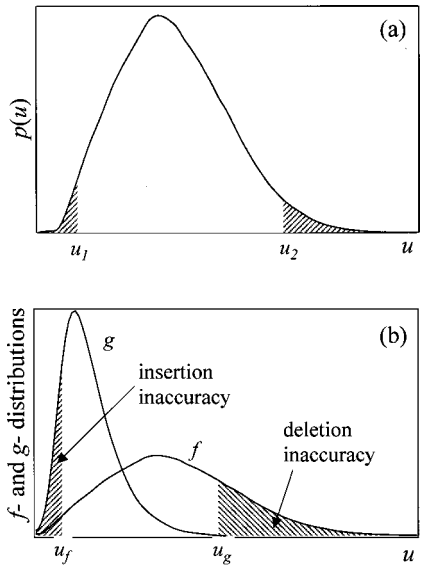

FIG. 3. Depiction of the probability distribution (representing f or g), sampled region, and the relative inaccuracy of the FEP calculations. (a) Shaded regions [the tails of distribution $p(u)$] are those that have not been well sampled during a finite length simulation, while the region in between is perfectly sampled; (b) the relative inaccuracy of the insertion (f) calculation is given by the area with energy less than u_f under the g distribution, while the relative inaccuracy of the deletion (g) calculation is given by the area with energy larger than u_g under the f distribution.

$$X^{\text{exact}}=\int_{-\infty}^{u_1}p(u)F(u)du+\int_{u_1}^{u_2}p(u)F(u)du+\int_{u_2}^{+\infty}p(u)F(u)du. \tag{14}$$

As part of our model development, we now assume that in a simulation u is sampled perfectly according to $p(u)$ for u between u_1 and u_2. Then, the X measured in the simulation will be

$$X^{\text{sim}}=\int_{u_1}^{u_2}p(u)F(u)du, \tag{15}$$

and the simulation error in X is given by the difference between X^{sim} and X^{exact}

$$\delta X=X^{\text{sim}}-X^{\text{exact}}=-\int_{-\infty}^{u_1}p(u)F(u)du-\int_{u_2}^{+\infty}p(u)F(u)du. \tag{16}$$

Now we return specifically to the FEP calculations. For the insertion calculation, the high-energy term in Eq. (16) has an extremely small value and therefore can be dropped. Similarly, the low-energy term in Eq. (16) for a deletion calculation can be ignored. We then can write the inaccuracy of insertion and deletion calculations as follows:

$$\delta E_{\text{ins}}=\exp(-\beta\Delta A_{\text{ins}}^{\text{sim}})-\exp(-\beta\Delta A^{\text{exact}})=-\int_{-\infty}^{u_f}f(u)\exp(-\beta u)du \tag{17}$$

and

$$\delta E_{\mathrm{del}} \equiv \exp(+\beta \Delta A_{\mathrm{del}}^{\mathrm{sim}}) - \exp(+\beta \Delta A^{\mathrm{exact}})$$
$$= -\int_{u_g}^{+\infty} g(u)\exp(+\beta u)\,du, \tag{18}$$

where u_f is the lowest (limit) energy sampled in an insertion calculation and u_g is the highest (limit) energy sampled in a deletion calculation. We can see in Eqs. (17) and (18) that the inaccuracy of FEP calculations is given by the contributions of the important configurations that are never sampled.

It is helpful to look at the relative error (δe) of the FEP calculations. Knowing Eq. (7), the relationship between f and g distributions, we have

$$\delta e_{\mathrm{ins}} \equiv \frac{\exp(-\beta \Delta A_{\mathrm{ins}}^{\mathrm{sim}}) - \exp(-\beta \Delta A^{\mathrm{exact}})}{\exp(-\beta \Delta A^{\mathrm{exact}})}$$
$$= -\int_{-\infty}^{u_f} g(u)\,du \tag{19}$$

and

$$\delta e_{\mathrm{del}} \equiv \frac{\exp(+\beta \Delta A_{\mathrm{del}}^{\mathrm{sim}}) - \exp(+\beta \Delta A^{\mathrm{exact}})}{\exp(+\beta \Delta A^{\mathrm{exact}})}$$
$$= -\int_{u_g}^{+\infty} f(u)\,du. \tag{20}$$

Note that for small errors, the magnitude of the relative error in $\exp(\pm \Delta A)$ is roughly equal to the magnitude of the absolute error in ΔA itself.

Equations (19) and (20) tell us that the relative inaccuracy of the insertion (f) calculation is given by the area with energy less than u_f under the g distribution, while the relative inaccuracy of the deletion (g) calculation is given by the area with energy larger than u_g under the f distribution. This relation is depicted in Fig. 3(b).

One may note that the relative errors of free-energy calculation, as shown by Eqs. (19) and (20), are always less than or equal to zero since the area under an f or g curve is non-negative. This implies that $\Delta A_{\mathrm{ins}}^{\mathrm{sim}} > \Delta A^{\mathrm{exact}}$ and $\Delta A_{\mathrm{del}}^{\mathrm{sim}} < \Delta A^{\mathrm{exact}}$, or in other words, for an insertion calculation the free-energy difference is overestimated, and for a deletion calculation it is underestimated. This analysis agrees with the common understanding and observation of FEP calculations. However, this common understanding often includes a misconception in the magnitude of the accuracy of the FEP calculations. The prevalent view is that both calculations are equally wrong, and the best result is obtained by splitting the difference.[17–26] Equations (19) and (20) do not explicitly provide information as to the magnitude of inaccuracy. To further look at the asymmetry problem, we need to study some properties of the f and g distribution functions.

E. Asymmetry of FEP calculations

Asymmetry in the FEP calculation is easily understood in the context of the preceding analysis by considering the overlap of the f and g distributions. If the upper limit energy of the g distribution lies well above the bulk of the f distribution, then the error in deletion is small; likewise, if the lower limit energy of f lies below the bulk of g, the insertion error is small. In the extreme cases of no overlap [Fig. 4(a)], both calculations give poor results, while for complete overlap both give equally good results. It is in the intermediate case that the asymmetry of the methods emerges. For some intermediate degree of overlap, the better result is obtained by sampling the distribution that is broader. As the amount of sampling is increased, and the limit energies move further into the wings of the distributions, the broader distribution will encompass the narrower distribution first, meaning that sampling on the broader distribution will provide a result of greater accuracy [Figs. 4(b) and 4(c)].

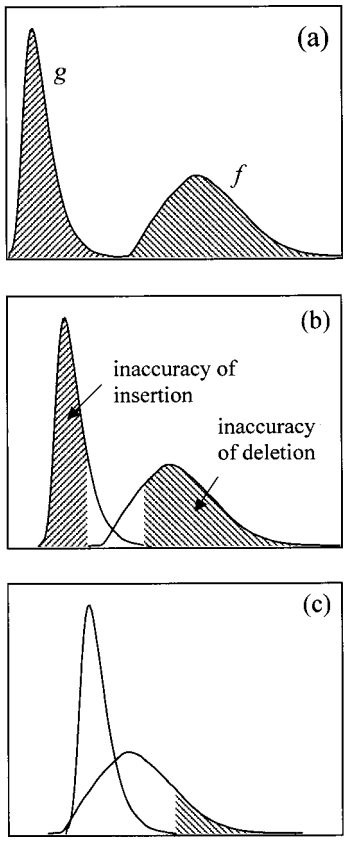

FIG. 4. Asymmetry in accuracy of FEP calculations. The inaccuracy of the insertion (deletion) calculation is given by the shadow area under the $g(f)$ distribution. The size of the area is related to the length of simulation, the overlap level of the energy distributions, and the breadth of each distribution. This series of plots show how the inaccuracy of both calculations changes as increasing the simulation length increases the degree of overlap of distributions. The insertion calculation, which typically samples the broader energy distribution (f), will give less simulation error than the deletion calculation.

This analysis raises an obvious question: Which distribution is wider, the f (insertion) distribution or the g (deletion) distribution? The f distribution describes sampling of the system having the larger entropy, so it is tempting to conclude that it is always the wider distribution. For the Widom calculation of the chemical potential, this is indeed the case. However, we are unable to prove this as a general result, so at present we leave it as a reasonable speculation indicating that insertion FEP calculations are inherently more accurate than corresponding deletion calculations.

F. Most-likely analysis

The next step in our model development is the formulation of a means for estimating the limit energies that govern the inaccuracy of the FEP calculations. What inaccuracy can be expected for a simulation with given length M? Assuming prior knowledge of the f and g distributions, it is possible to have a general answer to this question. It is clear from the foregoing inaccuracy model that the magnitude of the error depends on the value of limit energy u_g (for insertion calculation) or u_f (for deletion calculation). Once we have the limit energy, we know the inaccuracy of the calculation. So, the question becomes what is the limit energy for a given simulation?

The limit energy u_f or u_g can be found easily in a simulation. Of course, for a given simulation length M we would find different values of u_f or u_g in different independent simulation runs. However, a distribution of limit energies $P(u)$ can be defined to describe the limit energies observed in repeated runs of the same length M. This distribution will exhibit some maximum corresponding to the most likely limit energy u^* to be observed in any simulation. Then, we identify the most likely free energy to be that obtained by integrating Eq. (17) or (18) to this most likely limit energy. Accordingly, we define the inaccuracy as the difference between the most-likely free-energy outcome and the true value. Now, the key quantities needed are the probability function $P(u)$ and, from this, the most-likely limit energy, u^*.

First, let us look at these quantities for an insertion calculation. The probability density that u_f is the lowest energy sampled in a simulation with M insertion trials is given by

$$P(u_f)=f(u_f)\left[\int_{u_f}^{+\infty} f(u)du\right]^{M-1}. \tag{21}$$

The right-hand side of Eq. (21) is the product of the probability density that $u=u_f$ is sampled exactly once, and the probability that for all other $(M-1)$ samplings the energy value is between u_f and infinity ($u<u_f$ is never sampled).

The most-likely limit energy u_f^* is obtained by maximizing $P(u)$, and thus u_f^* should satisfy $[\partial P(u)]/\partial u|_{u=u_f^*}=0$, or (more conveniently) $[\partial \ln P(u)]/\partial u|_{u=u_f^*}=0$. After straightforward derivation starting with Eq. (21), we get (for large M)

$$\left.\frac{\partial \ln f(u)}{\partial u}\right|_{u=u_f^*} = M f(u_f^*). \tag{22}$$

Similarly, for the deletion calculation we have

$$P(u_g)=g(u_g)\left[\int_{-\infty}^{u_g} g(u)du\right]^{M-1}. \tag{23}$$

Then the most-likely limit energy u_g^* is

$$\left.\frac{\partial \ln g(u)}{\partial u}\right|_{u=u_g^*} = -M g(u_g^*). \tag{24}$$

Equations (22) and (24) indicate the most-likely limit energies for the insertion and deletion FEP calculations with simulation length M, knowing the energy distribution functions f and g. Insertion of these limit-energy values into Eqs. (17) [or (19)] and (18) [or (20)] yields the most-likely inaccuracy of the corresponding FEP calculations.

III. VERIFICATION AND SIMULATION TESTS

As described in Sec. II D, our inaccuracy model agrees with some common qualitative observations and understanding of the FEP error. Here, we conduct simulation tests to assess quantitatively the ability of the model to describe the inaccuracy of FEP calculations. We performed Monte Carlo simulation of simple Lennard-Jones (LJ) fluids in the NVT ensemble. We consider a reduced density of 0.9 [all units are made dimensionless by the LJ size (σ) and energy (ε) parameters], and two reduced temperatures: 0.9 and 2.0. The perturbation systems are defined as follows: ''system H'' contains 107 particles interacting according to the LJ potential (i.e., it has 107 LJ particles), and one particle that interacts with the others according to the hard-sphere potential with diameter 0.8. The LJ potential is truncated at a separation of 2.5, and no long-range correction is included. ''System L'' has 108 LJ particles. The insertion calculation ($\Delta S<0$) corresponds to the perturbation direction from system H to L, i.e., the reference is system H. The trial insertion process involves the identity change of the HS to a LJ particle and the growth of the particle size. The deletion calculation uses system L as the reference and involves the change of a LJ particle to a hard sphere. Both insertion and deletion FEP calculations were performed during the simulation. This introduces a small approximation to the calculation, in that the deletion calculation should be performed for 108 LJ spheres, but separate tests show that the error introduced by this consideration is negligible.

A test of the inaccuracy model requires as input the exact energy distribution functions. These we obtained by making histograms of the perturbation energy encountered in a long simulation with 5M cycles (where each cycle comprises 108 attempts of MC particle displacement followed by one trial of insertion or deletion measurement). The true free-energy difference is also needed to gauge the inaccuracy of the FEP calculations. We obtained this value using the relationship described in Eq. (7), or more explicitly, $\exp(\beta\Delta A)=\exp(\beta u)g(u)/f(u)$, applied to the measured f and g distributions over their region of overlap. Particularly important to the analysis are the low-energy tail of f distribution and the high-energy tail of g distribution, which typically have poor statistics. To improve them, we correct the low-energy tail of $f(u)$ using the information of the ''true'' $\exp(\beta\Delta A)$ just obtained, along with $g(u)$ which has very good statistics in the low-energy range. Similarly, the high-energy tail of $g(u)$ is corrected using the ''true'' free-energy difference and the f distribution.

We examined the effect of the simulation length M since it is a very important quantity affecting the error of the simulation. Simulations with $M=1000$, 2000, 5000, and 10 000 cycles were performed independently. After each simulation, we have free-energy difference by both insertion and deletion calculations. Simulations of each length were repeated 200 times independently. After all repeating simulations, the

probability histogram of free-energy results was made and the free-energy difference corresponding to the peak of the histogram was obtained and used as the ''most-likely'' value. Based on the most-likely free-energy difference and the true value, the most-likely inaccuracy is computed. In accord with the development, we report our results as the fractional error in the exponential of ΔA instead of ΔA directly. The most-likely inaccuracy observed from the simulation is given as follows. For insertion calculation

$$\delta E^{s}_{\mathrm{ins}} = \exp(-\beta \Delta A_{\mathrm{sim}}) - \exp(-\beta \Delta A_{\mathrm{true}}), \tag{25}$$

for deletion calculation

$$\Delta E^{s}_{\mathrm{del}} = \exp(+\beta \Delta A_{\mathrm{del}}) - \exp(+\beta \Delta A_{\mathrm{true}}), \tag{26}$$

where the superscript s indicates the most-likely value observed from simulations.

For the comparison we also computed the most-likely inaccuracy according to the model. We first computed the most-likely limit energies with the f and g distribution functions obtained from the long simulation. The most-likely lowest energy sampled in an insertion calculation, u_f^*, and the highest energy sampled in a deletion calculation, u_g^*, were calculated using Eqs. (22) and (24), respectively. Then, using the discrete form of our model [Eqs. (17) and (18)], the most-likely inaccuracies are

$$\delta E^{*}_{\mathrm{ins}} = -\sum_{u_i < u_f^*} f(u_i) \exp(-\beta u_i) \tag{27}$$

for insertion calculations and

$$\delta E^{*}_{\mathrm{del}} = -\sum_{u_i > u_g^*} g(u_i) \exp(+\beta u_i) \tag{28}$$

for deletion calculations. Note the discrete distribution functions $f(u_i)$ and $g(u_i)$ are normalized.

We compare the inaccuracies observed from the simulations [Eqs. (25) and (26)] and those from our most-likely model [Eqs. (27) and (28)] to verify if the inaccuracy analysis and model are valid.

IV. RESULTS AND DISCUSSION

The most-likely inaccuracies computed directly by the simulation data and by the inaccuracy model for both insertion and deletion calculations are presented in Fig. 5 as a function of simulation length M. We report results for two system temperatures: Fig. 5(a) for $T=0.9$ and Fig. 5(b) for $T=2.0$. The state condition $T=0.9$ and $\rho=0.9$ corresponds to a LJ fluid located slightly above the liquid–solid coexistence temperature.[31]

From Figs. 5(a) and 5(b), we can see that the model results agree with the simulation inaccuracies very well in both cases. This gives us the confidence that the error analysis conducted is appropriate. The difficulties of sampling the complex important configuration space can be simply studied by investigating the one-dimensional energy distribution functions. The major error of the FEP calculations is due to the poor sampling of the energy distribution wings.

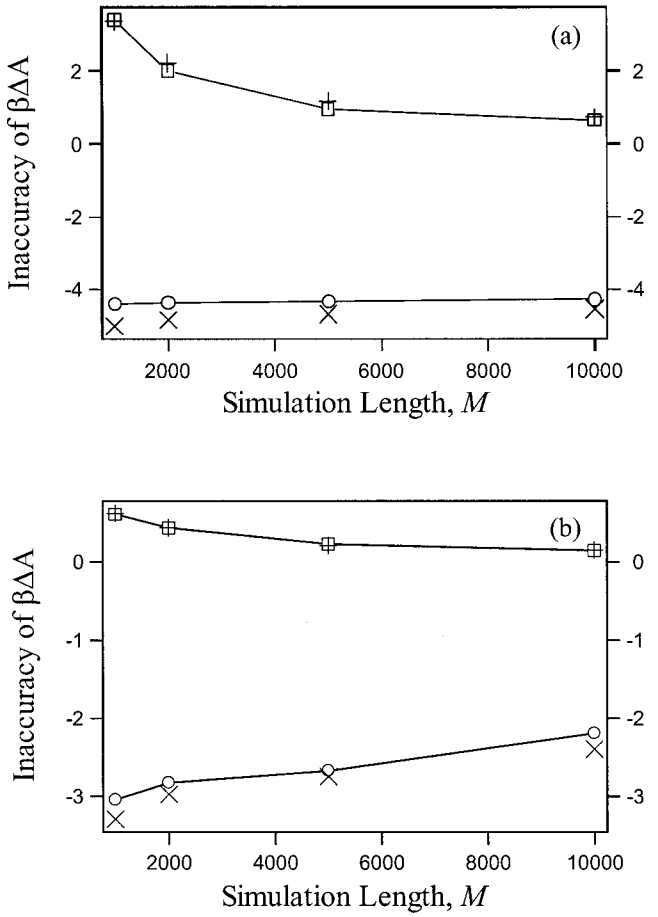

FIG. 5. Comparison of the inaccuracy in the free-energy change, evaluated by repeated simulations and by the model. Both the inaccuracy of insertion (square) and deletion (circle) calculations are plotted as a function of simulation length M, together with the model predictions ($+$ for insertion and $\times$ for deletion). $T=0.9$ for (a) and $T=2.0$ for (b). In both cases the system density is 0.9.

In Fig. 5 there are more points worth noting. As we expected, the free-energy errors for the insertion and deletion calculations have opposite sign, and both decrease with the increasing simulation length M. A longer simulation is able to sample the configuration space better than a shorter simulation, and therefore gives the estimation of the free-energy difference with better statistics. From the model point of view, in a longer simulation the limit energy is pushed further away toward the end of the energy tail. For an infinite simulation, the whole range of the f and g distributions will be well sampled and the result will be exact. However, the speed of error reduction over the simulation length M is not the same for the insertion and deletion calculations. From Fig. 5 we can see the error of the insertion calculation decreases faster than its deletion counterpart, with increasing simulation length. It is more efficient to reduce the simulation error by increasing the simulation length for insertion calculations than deletion ones. We observed similar behavior for the HS or simple three-state models[28] and believe it is common for many simulation situations. This behavior also reveals the asymmetric character of the insertion and deletion calculations. The magnitudes of the simulation errors by insertion and deletion calculations are not the same, a point

which is clearly shown in Fig. 5. We find the insertion calculation results in less error, or more reliable free-energy estimation. One should not assume these two FEP calculations are symmetric only because the errors have different sign.

It is worthwhile to consider how the superiority of insertion impacts the design of multistage FEP methods. Multistage methods[13,21,32] calculate the free energy between a reference and target by breaking the difference into smaller parts. Thus

$$A_1 - A_0 = (A_1 - A_W) + (A_W - A_0), \tag{29}$$

where the W subscript defines a system that in some way is intermediate between the target and reference system. Given the 0, 1, and W systems, there are four ways in which the overall FEP calculation can be completed, differing in the selection of the target and reference systems for the two subcalculations involved in computing the total difference. Two are commonly known as umbrella sampling[33] and Bennett's method,[34] while the other two we have called[13] staged insertion and staged deletion, although these names are appropriate only if $S_1 < S_W < S_0$.

In previous work[27] we showed that the intermediate should be selected such that the entropy differences $S_1 - S_W$ and $S_W - S_0$ are equal. This result anticipated the present work in that it assumed that both FEP calculations would be computed in the insertion direction, and that in particular $S_1 < S_W < S_0$. Such a selection of the intermediate system is possible only if the important configurations of the target and reference systems are related as described in Fig. 6(a). In our limited experience, examining Widom-type FEP calculations involving the insertion (or partial insertion) of a single molecule, this has always been the case.

For other perturbations this relationship may not hold, and it is important then to be especially careful in the formulation of the intermediate system. In particular Bennett's method or umbrella sampling may be more appropriate staging schemes. The cases in which they are appropriate for use are depicted in Figs. 6(b) and 6(c). Umbrella sampling is needed if there is virtually no overlap of the important configurations of the target and reference systems. Then, the W system must be selected to encompass both of them. Bennett's method can be applied if the systems have some overlap in their important regions of phase space. Bennett's method is special in that the W system is never used as a reference in one of the substage FEP calculations. As shown by Bennett, this feature permits some optimization to be applied in selecting the intermediate. Interestingly, if the target and reference are related as shown in Fig. 1, with the target forming a complete subset of the reference, then (at least for hard spheres) the optimization procedure yields an intermediate that is identical to the target, i.e., the deletion stage of the calculation is eliminated and the method reduces to a single-stage FEP calculation. Staged insertion is then the method of choice. Higher-order staging algorithms could be developed to extend these ideas.

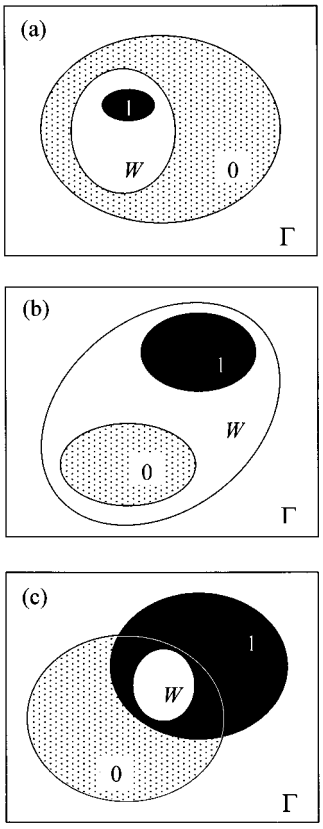

FIG. 6. Appropriate staging strategy in different situations according to the overlap of the important phase space. A two-stage FEP calculation is used as an example, where three systems, the reference 0, the target 1, and the intermediate W are involved. In the phase-space diagrams, these three systems are represented as the partial shaded, black, and white regions, respectively. For FEP calculations containing more stages, the same idea can be applied. The relationship of the important configurations of three systems is schematically shown. Information about choosing an appropriate intermediate system is described in the text for each case. (a) The important configurations of the target system sits inside that of the reference system completely. A staged insertion is the best choice for an accurate result. (b) No overlap region of the important configurations of the reference and target. Umbrella sampling is the most suitable technique. (c) The important configurations of the target and reference are partially overlapped. Bennett's method is appropriate for FEP calculations.

V. CONCLUSIONS

We are strongly of the opinion that FEP calculations should be performed only in the insertion direction, i.e., the simulation should sample the system of larger entropy, and compute the free-energy difference for a perturbation to a target system of smaller entropy. This outcome follows as a consequence of a more important and more restrictive criterion for obtaining accurate measurements in FEP calculations. It is necessary that the configurations important to the target system form a subset of the configurations important to the reference. This situation is depicted in Fig. 1. For small, localized perturbations such as those employed in chemical-potential calculations, the more restrictive condition is usually satisfied by the less restrictive condition that looks only at the entropy difference. In more complex cases, where the target cannot be selected such that its important configurations are entirely a subset of the important configu-

rations of the reference, it is necessary to apply a staging approach to the free-energy calculation. However, the design of the staging method should still be guided by the criteria outlined here.

ACKNOWLEDGMENT

Acknowledgment is made to the Donors of the Petroleum Research Fund, administered by the American Chemical Society, for support of this research.

[1] D. Frenkel and B. Smit, *Understanding Molecular Simulation: From Algorithms to Applications* (Academic, New York, 1996).
[2] M. P. Allen and D. J. Tildesley, *Computer Simulation of Liquids* (Clarendon Press, Oxford, 1987).
[3] K. E. Gubbins, in *Foundations of Molecular Modeling and Simulation, AIChE Symposium Series*, edited by P. Cummings and P. Westmoreland, Vol. 97, No. 325 (2001).
[4] P. Kollman, Chem. Rev. **32**, 2395 (1993).
[5] R. W. Zwanzig, J. Chem. Phys. **22**, 1420 (1954).
[6] J. L. Jackson and L. S. Klein, Phys. Fluids **7**(2), 228 (1964).
[7] B. Widom, J. Chem. Phys. **39**, 2808 (1963).
[8] K. S. Shing and K. E. Gubbins, Mol. Phys. **46**(5), 1109 (1982).
[9] B. Guillot and Y. Guissani, Mol. Phys. **54**(2), 455 (1985).
[10] J. G. Powles, W. A. B. Evans, and N. Quirke, Mol. Phys. **46**(6), 1347 (1982).
[11] J. G. Powles, Chem. Phys. Lett. **86**(4), 335 (1982).
[12] K.-K. Han, J. H. Cushman, and D. J. Diestler, J. Chem. Phys. **93**, 5167 (1990).
[13] D. A. Kofke and P. T. Cummings, Mol. Phys. **92**(6), 973 (1997).
[14] D. A. Kofke and P. T. Cummings, Fluid Phase Equilibria **150**, 41 (1998).
[15] N. G. Parsonage, Mol. Phys. **89**(4), 1133 (1996).
[16] N. Parsonage, J. Chem. Soc., Faraday Trans. **92**(7), 1129 (1996).
[17] R. H. Henchman and J. W. Essex, J. Comput. Chem. **20**(5), 499 (1999).
[18] W. L. Jorgensen and C. Ravimohan, J. Chem. Phys. **83**, 3050 (1985).
[19] D. A. Pearlman and P. A. Kollman, J. Chem. Phys. **90**, 2460 (1989).
[20] R. H. Wood, W. C. F. Mühlbauer, and P. T. Thompson, J. Phys. Chem. **95**, 6670 (1991).
[21] R. J. Radmer and P. A. Kollman, J. Comput. Chem. **18**, 902 (1997).
[22] C. Chipot, L. G. Gorb, and J.-L. Rivail, J. Phys. Chem. **98**, 1601 (1994).
[23] W. L. Jorgensen, J. K. Buckner, S. Boudon, and J. Tirado-Rives, J. Chem. Phys. **89**, 3742 (1988).
[24] D. A. Pearlman, J. Phys. Chem. **98**(5), 1487 (1994).
[25] R. H. Wood, J. Phys. Chem. **95**(12), 4838 (1991).
[26] C. Chipot, C. Millot, B. Maigret, and P. A. Kollman, J. Phys. Chem. **98**(44), 11362 (1994).
[27] N. Lu and D. A. Kofke, J. Chem. Phys. **111**, 4414 (1999).
[28] N. Lu and D. A. Kofke, in *Foundations of Molecular Modeling and Simulation, AIChE Symposium Series*, edited by P. Cummings and P. Westmoreland, Vol. 97, No. 325 (2001).
[29] K. S. Shing and K. E. Gubbins, Mol. Phys. **43**(3), 717 (1981).
[30] M. P. Allen, in *Proceedings of the Euroconference on Computer Simulation in Condensed Matter Physics and Chemistry*, edited by K. Binder and G. Ciccotti (Como, Italy, 1996), Vol. 49, pp. 255–284.
[31] R. Agrawal and D. A. Kofke, Mol. Phys. **85**(1), 23 (1995).
[32] J. P. Valleau and D. N. Card, J. Chem. Phys. **57**, 5457 (1972).
[33] G. M. Torrie and J. P. Valleau, J. Comput. Chem. **23**, 187 (1977).
[34] C. H. Bennett, J. Comput. Chem. **22**, 245 (1976).