Phase-space overlap measures. II. Design and implementation of staging methods for free-energy calculations

Di Wu and David A. Kofke[a)]
Department of Chemical and Biological Engineering, University at Buffalo, The State University of New York, Buffalo, New York 14260-4200

(Received 27 May 2005; accepted 8 July 2005; published online 1 September 2005)

We consider staged free-energy calculation methods in the context of phase-space overlap relations, and argue that the selection of work-based methods should be guided by consideration of the phase-space overlap of the systems of interest. Stages should always be constructed such that work is performed only into a system that has a phase-space subset relation with the starting system. Thus multiple stages are required if the systems of interest are not such that one forms a phase-space subset with the other. Three two-stage methods are possible, termed *umbrella sampling, overlap sampling*, and *funnel sampling*. The last is appropriate for cases in which the subset relation holds, but only in the extreme, meaning that one system's important phase space constitutes a very small portion of the others. Umbrella sampling is most suitable for nonoverlap systems, and overlap sampling is appropriate for systems exhibiting partial phase-space overlap. We review recently introduced metrics that characterize phase-space overlap, showing that the performance of the single- and two-stage methods is consistent with the phase-space picture. We also demonstrate that a recently introduced bias-detection measure is effective in identifying inaccuracy in single- and multistage calculations. The examples used are the chemical-potential calculation for a Lennard-Jones liquid at moderate and at high densities, the same for model water at ambient conditions, and a process of charging a neutral ion in water. © *2005 American Institute of Physics*. [DOI: 10.1063/1.2011391]

I. INTRODUCTION

The calculation of free energies is an important capability required of molecular simulation, because knowledge of the free energy is necessary to understand a wide range of thermophysical behaviors. A broad range of popular free-energy techniques can be derived from the nonequilibrium work (NEW) formalism proposed by Jarzynski.[1] A significant problem with these work-based methods is their tendency to provide results that are statistically biased or inaccurate. Often the remedies used to address this widely recognized problem are either ineffective or inefficient. In recent work[2] we examined how phase-space overlap concepts could be used to understand the performance of work-based free-energy methods, and we proposed a heuristic that could be applied to detect the presence of bias in a calculated result. The heuristic is valuable because it can help practitioners to sort out the quality of conflicting results, and to apply the right amount of computational effort to get a good result.

The aim of a free-energy calculation is to determine the free-energy difference between two systems, labeled A and B. The Jarzynski formula for the free-energy difference $\Delta F = F_B - F_A$ is

$$\exp(-\beta\Delta F) = \overline{\exp(-\beta W_{A\to B})}, \tag{1}$$

where $\beta = 1/kT$ is the reciprocal temperature in energy units. Here $W_{A\to B}$ is the work involved in a nonequilibrium process that transforms the system from A to B, and the overbar indicates an average over many such transformations, each beginning from an equilibrated point in system A. The bias observed in a free-energy calculation for a given pair of systems is in general asymmetric. Equation (1) may instead be written for a transformation $B\to A$, and in principle this formula is no more or less valid than the one already written for $A\to B$. In practice, however, the bias seen in one direction will be very different than that seen in the other, for an equal amount of sampling. Sometimes the free energy calculated in one direction will be very good, while that calculated in the other is wildly (but reproducibly) inaccurate. The heuristic developed in our previous work can be especially helpful in this circumstance.

One of the conclusions of the phase-space analysis is that for some system pairs it is not possible to obtain an accurate free energy with any feasible amount of computational effort (at least not without approaching a reversible-work limit), regardless of the direction of the transformation. In such situations it is necessary to apply staging methods. Staging methods introduce an addition system (or systems) labeled C, and attempt to evaluate the free-energy difference as the sum of differences with respect to system C;

$$\Delta F = (F_B - F_C) + (F_C - F_A). \tag{2}$$

Phase-space overlap ideas can be used to guide the selection of system C, and it is the aim of the present work to demonstrate how this can be done.

In Sec. II we review our previous work involving phase-space overlap, its quantification, and its use in detecting bias

[a)]Electronic mail: kofke@buffalo.edu

in work-based free-energy calculations. Then in Sec. III we examine staging methods in light of these phase-space overlap ideas. In Sec. IV, we demonstrate the application of the ideas with several examples, and in Sec. V we summarize and conclude.

II. BACKGROUND

The *important phase space* (by which we often mean just the configuration space, as momentum coordinates are usually not relevant) of a system A describes the set of all configurations of the system that contribute significantly to its properties. It is necessary to get a representative sample of the important phase space of a system in order to assess its free energy. Free-energy calculations necessarily involve two such phase-space subspaces, the *A-important* subspace and the *B-important* subspace. The relation between these subspaces determines the ability of a single-stage free-energy calculation (that is, one involving only the two systems A and B) to yield a bias-free result.

There are three qualitative categories of phase-space overlap. These are *subset, partial overlap*, and *nonoverlap*. Our previous work has emphasized that only for systems obeying a subset relation can single-stage NEW calculations yield an accurate result. Moreover, such calculations must be performed in the direction $A \to B$, where we define system A to be the one with the larger important phase space (thus implying that the B-important space is a subset of it).

We have introduced two metrics to quantify phase-space overlap.[2] The first metric is based on four total-energy distributions. A given configuration (i.e., point in phase space) will have an energy U_A defined for system A, and energy U_B for system B. Of interest is the distribution of A-defined energies observed when sampling configurations important to system A, denoted $\langle A \rangle_A$, and the corresponding A-energy distribution observed when sampling configurations important to system B, denoted $\langle A \rangle_B$. If the A energies observed when sampling system B are significantly greater than those observed when sampling system A, then system B is sampling configurations outside the A-important phase space. To determine whether system A samples configurations outside of the B-important phase space, we examine the corresponding B-energy distributions, denoted $\langle B \rangle_B$ and $\langle B \rangle_A$. We defined an overlap integral for each of these pairs, and we take these as a convenient quantification of the phase-space overlap determined with these metrics. Thus K_{AB} can be loosely understood as the "amount of A in B," and ranges from 0 (no configurations important to A are important to B) through unity (all configurations important to A are important to B) to a maximum of 2 (configurations important to A are among the lowest-energy configurations of system B). Likewise, K_{BA} is interpreted as the "amount of B in A." Together the values of K_{AB} and K_{BA} can be used to distinguish the three general cases of subset, partial overlap, and nonoverlap.

The second metric is based on the ideas from information theory, and, in particular, the notion of relative entropy. We again have two measures, denoted s_A and s_B, which are interpreted as the "distance" between the phase-space distributions. These quantities do not connect as clearly as K_{AB} and K_{BA} do to the qualitative overlap categories, but they have two very appealing features. First, they are directly related to the dissipated work associated with the NEW calculation

$$s_A = \langle \beta W_{A\to B} \rangle_A - \beta \Delta F,$$
$$s_B = \langle \beta W_{B\to A} \rangle_B - \beta \Delta F. \tag{3}$$

Second, the relative entropies can be used to define measures that are indicative of whether sufficient sampling has been conducted to produce an accurate free energy. These measures are

$$\Pi_{A\to B} = \sqrt{\frac{s_A}{s_B} \mathbf{W}_L\left(\frac{1}{2\pi}(M-1)^2\right)} - \sqrt{2s_A},$$
$$\Pi_{B\to A} = \sqrt{\frac{s_B}{s_A} \mathbf{W}_L\left(\frac{1}{2\pi}(M-1)^2\right)} - \sqrt{2s_B}, \tag{4}$$

where M is the number of work samples taken, and $\mathbf{W}_L(x)$ is the Lambert W function, defined as the solution for w in $x = we^w$. We proposed the heuristic that sufficient sampling is performed if the corresponding value of Π is greater than zero (e.g., $\Pi_{A\to B} > 0$ if work is performed in the direction $A \to B$). As shown in Eq. (3), s_A and s_B both depend on the free energy that is being measured. We have argued that any bias in this measured free energy will *not* be such that it causes the Π values to be greater than zero and thereby give a false indication of an accurate result. In this sense the heuristic provides a fail-safe method for bias detection.

It should be noted then that only when applied in the free-energy perturbation (FEP) limiting case of the NEW method (for which the change from A to B is instantaneous) will this expressions given in Eq. (3) connect to the definitions of s_A and s_B made in terms of phase-space distributions. Nevertheless we anticipate that the heuristic based on the measures defined by Eq. (4) will apply to general NEW calculations, and not just the FEP limit for which they were originally developed and tested. We note, for example, that in the opposite limit of a reversible process, the dissipated work is zero and both Πs become positive with a very small amount of sampling.

III. STAGING METHODS

In the case where the A and B phase spaces exhibit only partial- or nonoverlap relations, no (irreversible) single-stage calculation will be able to provide an accurate value of their free-energy difference. A different problem can arise when the subset relation is satisfied, but the size of the phase spaces differ by many orders of magnitude. Staging methods are needed in all such circumstances, and this opens up a wide array of design decisions. First, one must consider the number of additional stages to use. Then one must select for each intermediate staging system a Hamiltonian, or energy function, which determines the weight given to each phase-space point and thereby defines the important phase space for that system. Finally, there is the question of which direction to perform the perturbation for each stage. To simplify the analysis we will focus consideration to the use of a single

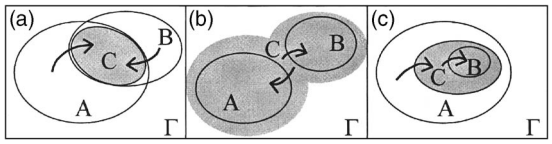

FIG. 1. Schematic descriptions of the relations between systems A, B, and C involved in free energy in staging methods. Large square represents all phase space, and unfilled ovals represent important phase space of system A and system B, respectively. Filled regions represent the important phase space of the intermediate system C. (a) Overlap sampling in the context of A and B exhibiting partial overlap relation. C system lies inside the partial overlap region of A and B systems, and sampling is conducted in the directions $A \rightarrow C \leftarrow B$. (b) Umbrella sampling in the context of an A-B nonoverlap relation. C system encompasses both A and B systems, and sampling is conducted in the directions $A \leftarrow C \rightarrow B$. (c) Funnel sampling in the context of A-B extreme subset relation. C system's important phase space is larger than B system but smaller than A system and lies in between them. Sampling is conducted in the directions $A \rightarrow C \rightarrow B$.

additional system, corresponding to a two-stage approach. The concepts entering in this case are generalized in a straightforward way to higher-order staging methods, and we briefly provide an example of such an extension in Sec. V.

The key requirement of the added system C is that it exhibits a phase-space subset relation with both the A and B systems. The subset relation can apply in either direction, so, for example, C may be selected as a subset of A, or such that A is a subset of it. The only restriction is in the logic imposed by the existing overlap relations for A and B. Then during each perturbation stage, the calculation is performed in the direction from the larger phase-space system to the smaller subset system. For such a procedure, each perturbation stage should produce an accurate free-energy calculation, and consequently the overall accuracy is improved.

Each of the three A-B phase-space overlap relations suggests its own strategy in designing and using the added system. The staging methods are termed *overlap sampling, umbrella sampling,* and *funnel sampling.* We will now consider each in turn.

A. Overlap sampling and overlap relation

The important parts of phase space for a system pair exhibiting a partial-overlap relation are depicted in Fig. 1(a). The appropriate staging approach to this case is what we call overlap sampling, and its formulation is illustrated in the figure as well. The intermediate stage C is designed to have an important phase space that is inside the region of overlap of the A and B phase spaces, and the two perturbation stages are performed in the direction $A \rightarrow C$ and $B \rightarrow C$, respectively, so that each follows the rule to perturb only into a subset region. Now we can rewrite the free-energy difference as a two-stage ensemble average

$$\exp(-\beta\Delta F) = \frac{\overline{\exp(-\beta W_{A\rightarrow C})}}{\overline{\exp(-\beta W_{B\rightarrow C})}}. \tag{5}$$

Bennett's method[3] as applied for FEP is a type of overlap sampling, and the optimization process followed by Bennett leads to the specification of the best choice for the system C. However, Bennett's method is not usually viewed as a multistage method, but instead as an approach to properly combine perturbation averages in the two directions $A \rightarrow B$ and $B \rightarrow A$; here it is known as the "acceptance ratio" method. When considered in the context of simple FEP, there is no practical difference in these perspectives, but the two approaches lead to different methods when generalized to non-FEP NEW calculations, because in the general case the path between A and B is relevant. Crooks[4] has proposed an extension of Bennett's acceptance-ratio formulation to NEW, while Lu *et al.*[5] have shown how to generalize Bennett's method to NEW when viewed as a staging method.

Bennett's optimum choice for system C is derived through a variance-minimization procedure, and it is explained well in the book by Frenkel and Smit;[6] Bennett also showed that this choice is good from the standpoint of producing an accurate (bias-free) result. It is important to note that, unlike the other staging cases treated below, the optimization for overlap sampling is complete. The selection of the system C has no effect on the sampling of configurations, so issues of convergence and ergodicity of sampling cannot enter into the optimization process. The result of the optimization can be expressed in terms of the system-C energy U_C,

$$\exp(-\beta U_C(\gamma)) = [(1-\gamma)\exp(+\beta U_A) + \gamma\exp(+\beta U_B)]^{-1}. \tag{6}$$

Note that $\gamma=0$ corresponds to system A, and $\gamma=1$ is system B. The optimum value for $\gamma(\in[0,1])$ is given by

$$\gamma = [1+\exp(+\beta\Delta F)]^{-1}, \tag{7}$$

where ΔF is the free-energy difference being measured by the procedure. An appealing aspect of overlap sampling is that it can be applied noniteratively even though the system C is defined in terms of ΔF. One must only tabulate the averages in (5) for a range of values of γ, and then determine the optimum γ self-consistently after the simulation averaging is complete. For NEW calculations, γ provides an appropriate path parameter, for which the optimum can be ascertained postsimulation with knowledge of the work averages (in both directions) tabulated for a range of γ.[5]

B. Umbrella sampling and nonoverlap relation

The nonoverlap relation [illustrated in Fig. 1(b)] is the most unfavorable case for free-energy calculations because there is no overlap region between the two systems at all. However, many real systems exhibit this relation and a well-known solution to it is umbrella sampling,[7] whereby an intermediate stage is constructed to be a superset of both A and B systems, like an umbrella encompassing both of them. Figure 1(b) provides a schematic representation of a two-stage umbrella sampling scheme. The two stages are performed from $C \rightarrow A$ and $C \rightarrow B$, respectively, and the ensemble average can be written as

$$\exp(-\beta\Delta F) = \frac{\overline{\exp(-\beta W_{C\rightarrow B})}}{\overline{\exp(-\beta W_{C\rightarrow A})}}. \tag{8}$$

By superficially following Bennett's approach for optimization of overlap sampling, we can derive a two-stage intermediate potential U_C that apparently minimizes the statistical error.[8] The result can be expressed as

$$\exp(-\beta U_C) = [(1-\gamma)\exp(-2\beta U_A) + \gamma \exp(-2\beta U_B)]^{1/2}, \tag{9}$$

with

$$\gamma = [1+\exp(-2\beta\Delta F)]^{-1}. \tag{10}$$

This is a reasonable choice, but inasmuch as sampling is performed according to the system-C potential, and the effect of the potential on sampling does not enter into the optimization, the "optimal" result given here may in practice be somewhat less than optimal. It is optimal only if it can be assumed that system C is well sampled. In practice it may occur that sampling gets stuck in one of the two systems, and transition from the A-important space to the B-important space occurs infrequently. Multiple-staging methods (i.e., involving more than two stages) might help in these situations.

As with the overlap-sampling optimum, the optimal umbrella potential is expressed in terms of the free energy that is the aim of the calculation. Unlike overlap sampling, this feature presents a significant inconvenience to the application of the method. In umbrella sampling, the sampling is performed according to the system-C potential, and consequently it is not possible in a single simulation to collect averages for a range of γ and determine the optimum after the simulation is complete. If a suitable estimate of ΔF is not available at the outset, then some iteration is needed to arrive at an approximate value before conducting the calculation in earnest. Nevertheless, this inconvenience is not considered to be a serious obstacle to the practical application of the method. Still, one might note that the partial-overlap case considered previously could also be addressed through the application of umbrella sampling, but we consider overlap sampling to be preferable for the reasons outlined here.

C. Funnel sampling and subset relation

If systems A and B have a subset relation (with B the subset), then NEW calculations conducted in the $A \to B$ direction can be reliable. However, a staging method can sometimes improve the outcome, especially for the extreme subset case. We have argued previously[9] that an advantage is found if the dimensionless entropy difference $(S_A - S_B)/k$ is greater than 2. Figure 1(c) shows a two-stage funnel sampling scheme. Here the intermediate stage C constitutes a subset region of A but a superset region of B. The perturbation is conducted in the direction $A \to C \to B$. Each time the perturbation moves into a smaller subset region, like going through a funnel, thus we call this sampling scheme funnel sampling. The total free energy is given as the sum in Eq. (2), which leads to the following expression in terms of the NEW averages for the two stages:

$$\exp(-\beta\Delta F) = \overline{\exp(-\beta W_{A\to C})} \times \overline{\exp(-\beta W_{C\to B})}. \tag{11}$$

An optimization process can be performed as in the case of overlap and umbrella samplings, to specify a system C that minimizes the variance of the calculated result. This system influences sampling for only one of the stages, so some of the associated problems discussed with umbrella sampling would be attenuated in this case. However, the optimization procedure does not lead to a clean result such as that given in Eqs. (6) and (9). Instead it gives U_C as the solution of a cubic polynomial that includes a Lagrange multiplier that must also be determined. It does not seem to provide a practical specification for the system C. Lacking this, we formulate it by applying whatever system features that seem physically appropriate (such as an atom diameter).

Jarzynski's NEW method[1] as it is applied in its basic formulation could be viewed in the context of the multiple-stage funnel sampling scheme. It evolves a sequence of stages perturbing from A to B, and averages the total work done through this process as prescribed by Eq. (1), wherein $W_{A\to B} = \Sigma_{i=1}^{n} W_{C_{i-1}\to C_i}$ is the cumulative work combined from each stage with $C_0 = A$ and $C_n = B$. Of course the primary difference between Eqs. (1) and (11) is that the NEW method does not fully equilibrate each intermediate stage before perturbing to the next. Still, a properly implemented NEW method must advance at each step into a subset region, because each step involves a nonequilibrium perturbation from one staging system to another. This process repeats down through system B. Thus the NEW method, to the extent it is not implemented reversibly, requires for accuracy a path that follows a funnel-sampling scheme from A to B. If the phase-space relations of A and B are such that this cannot be done, analogs of overlap or umbrella staging methods must be employed.

For the extreme-subset case, the B system phase space is much too small to be sampled sufficiently in a single stage, and two or more stages are needed for an accurate calculation. It may be difficult to define the stages in a way that they follow a smooth progression from A to B. For a NEW calculation, a partial remedy to this problem might be found through the recently developed Rosenbluth-sampled NEW methods,[10] which are formulated to enable the NEW process to avoid the adverse effects of large work values typical in the extreme-subset case.

IV. EXAMPLES AND APPLICATIONS

Now we study three cases to demonstrate the applicability of phase-space overlap relations in realistic problems. We measure each system's phase-space overlap relation and apply different work-based staging methods to them. We compare the performance of these staging methods in the context of each system's overlap relations and show how these overlap relations can help us understand the calculations and provide useful solutions. Finally, we examine the bias in each FEP stage and consider it in comparison to the bias-detection measure Π.

A. Lennard-Jones system

We performed canonical-ensemble simulations of a Lennard-Jones (LJ) system (with size and energy parameters σ and ε) at constant temperature $kT/\varepsilon \equiv T^* = 1$, density $\rho\sigma^3 \equiv \rho^* = 0.7$, and number of molecules $N = 108$ or 109. We define a system A that contains 108 atoms and a system B that has 109 atoms, so that the calculated free-energy difference is the system's residual chemical potential (i.e., in excess of an ideal gas at the same density). We use a truncation radius

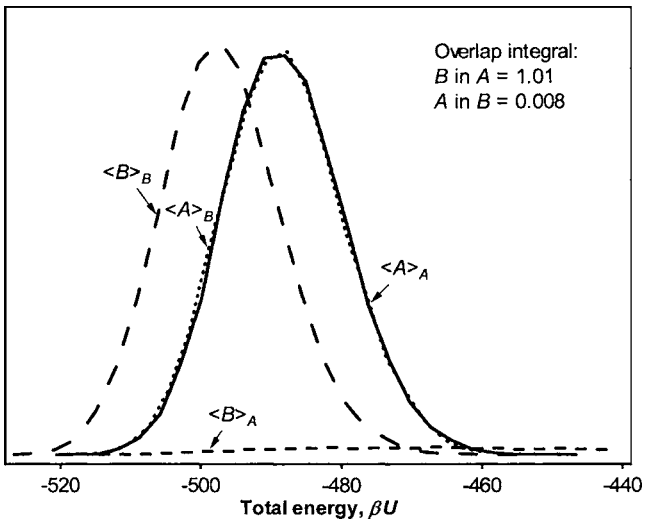

FIG. 2. Total-energy pair distributions for particle insertion in a Lennard-Jones system at a density $\rho^*=0.7$. All curves are normalized to unity. Curve labeled $\langle A\rangle_A$ is the distribution of system-A energies observed when sampling system A, while $\langle A\rangle_B$ is the distribution of system-A energies observed when sampling system B. $\langle B\rangle_B$ and $\langle B\rangle_A$ are defined correspondingly. Also indicated in this figure are the overlap integral measurements which are listed in Table I.

2.5σ and apply standard long-range corrections. In order to know the system's phase-space overlap relations, we performed total-energy distribution measurements and calculated overlap integrals K_{AB} and K_{BA} as discussed in Sec. II. The energy distributions are shown in Fig. 2 and overlap-integral values are included in Table I. The pair distributions $\langle A\rangle_A$ and $\langle A\rangle_B$ are almost identical, which means that system B's important phase space is inside that of system A; while for the other pair distribution $\langle B\rangle_A$ is very much on the right of $\langle B\rangle_B$ indicating that system A's phase space is outside the system B's phase space. This situation is summarized by the K values: $K_{AB}=0.008$ (A in B) and $K_{BA}=1.01$ (B in A). Therefore the Lennard-Jones system under this condition exhibits a simple phase-space subset relation [Fig. 1(c)], and we anticipate that the FEP insertion calculation can provide accurate results but deletion calculation will not.

The expected behavior is clearly demonstrated in Fig. 3, where we plot the free-energy difference as a function of the amount of sampling, and compare the results with the LJ equation of state value of Johnson *et al.*[11] Each FEP measurement is taken every 50 simulation cycles, where one cycle is defined as N attempted Monte Carlo trial moves. Since it is not guaranteed that an atom can be successfully inserted into the phase at each attempt, the calculation from insertion direction ($A\rightarrow B$) usually yields large biased results at the beginning, sometimes infinity. However, as the sampling increases, the results from the insertion direction become more and more accurate; in contrast, results from the deletion direction are always inaccurate, and remain almost unchanged over the whole process. We also performed NEW calculations, using $n=10$ steps during the traversal from $A\rightarrow B$ as well as $B\rightarrow A$ direction, and 50 cycles between finishing one step and starting the next (these measurement settings are the same for the calculations for the following systems as well). We see that the NEW calculation can improve the results of deletion-direction calculation but will not be accurate unless the process is reversible. This is consistent with the system's subset relation, where the B system's phase space is smaller and inside the A system. There is always an unreachable space which cannot practically be sampled in B but is important to A.

TABLE I. Properties of the application systems studied in Sec. IV. K_{AB} and K_{BA} are the total-energy overlap integrals of the corresponding systems, and $\beta\Delta F$ is the true free-energy difference. Data shown in the column Π_{app} are the apparent Π values calculated using the biased free-energy results obtained from the simulations. Data with the asterisks are the apparent Π values with the appropriate term $\sqrt{s}$ factored out.

Systems	Samplings		Π_{app}	Bias in $\beta\Delta F$
LJ $\rho^*=0.7$ $\beta\Delta F=-3.48$ $K_{AB}=0.008$ $K_{BA}=1.01$	Direct	$A\rightarrow B$	0.33^*	−0.02
		$B\rightarrow A$	−2.02	−4.35
	Overlap	$A\rightarrow C$	0.57^*	−0.03
		$C\leftarrow B$	2.23	
	Umbrella	$A\leftarrow C$	3.87^*	−0.04
		$C\rightarrow B$	3.97^*	
LJ $\rho^*=0.9$ $\beta\Delta F=0.24$ $K_{AB}=0.0002$ $K_{BA}=0.84$	Direct	$A\rightarrow B$	-0.24^*	1.46
		$B\rightarrow A$	−2.66	−8.98
	Overlap	$A\rightarrow C$	0.02^*	0.25
		$C\leftarrow B$	0.89	
	Umbrella	$A\leftarrow C$	4.03^*	0.09
		$C\rightarrow B$	3.69^*	
Water $\beta\Delta F=-9.39$ $K_{AB}=0.015$ $K_{BA}=0.8$	Direct	$A\rightarrow B$	-0.69^*	6.3
		$B\rightarrow A$	−5.65	−8.83
	Overlap	$A\rightarrow C$	-0.06^*	−0.03
		$C\leftarrow B$	0.27	
	Umbrella	$A\leftarrow C$	4.08^*	−0.16
		$C\rightarrow B$	2.36^*	
Ion-charging $\beta\Delta F=-169.85$ $K_{AB}=3\times10^{-5}$ $K_{BA}=0.005$	Direct	$A\rightarrow B$	−9.22	94.82
		$B\rightarrow A$	−9.61	−117.45
	Overlap	$A\rightarrow C$	−7.76	−22.48
		$C\leftarrow B$	−8.53	
	Umbrella	$A\leftarrow C$	−4.61	−7.56
		$C\rightarrow B$	−4.79	

From the arguments above, we do not expect much improvement in the calculations when we apply umbrella-sampling or overlap-sampling staging methods. Accordingly the results in Fig. 3 show that overlap sampling does little better than direct insertion calculations, but that umbrella sampling does in this case provide accurate results with less sampling. We note that these and all of the following umbrella-sampling calculations were performed using our preexisting knowledge of the free-energy difference, as required to specify the system-C weight function. In practice, of course, the free-energy difference is not known *a priori*. Therefore an iteration process is required for the determination of the weighting parameter C, and this process may reduce the overall efficiency of the algorithm correspondingly; one approach sometimes advocated is the use of short

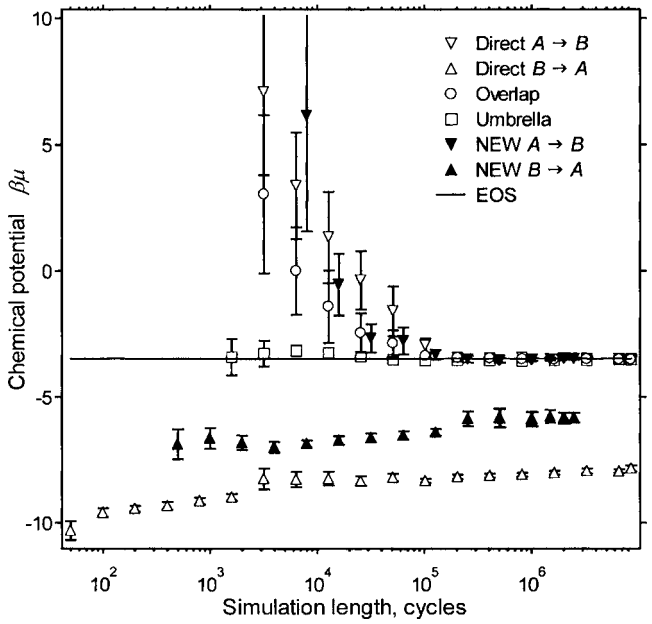

FIG. 3. Chemical-potential calculations for the Lennard-Jones system at a density $\rho^*=0.7$. Calculation method descriptions are given in Sec. IV. Chemical potential $\beta\mu$ is plotted against simulation length, cycles. Each symbol, attached with error bars, represents the average of ten independent simulation results at the specified amount of sampling (symbol descriptions are same for Figs. 5, 7, and 9).

simulations, perhaps on a smaller system, to estimate the free-energy difference to obtain C for a "production" calculation. Calculations performed without knowledge of C (e.g., $C=0$) can still yield a more accurate free-energy difference compared to FEP calculation, but they may well still be prone to inaccuracy depending on the nature of the phase-space overlap.

We consider now the same model but under different conditions, and see that its phase-space overlap relation can change. Keeping all the other parameters fixed we take the density to $\rho^*=0.9$ and use a truncation radius 2.4σ. This system's total-energy distributions are plotted in Fig. 4. Now

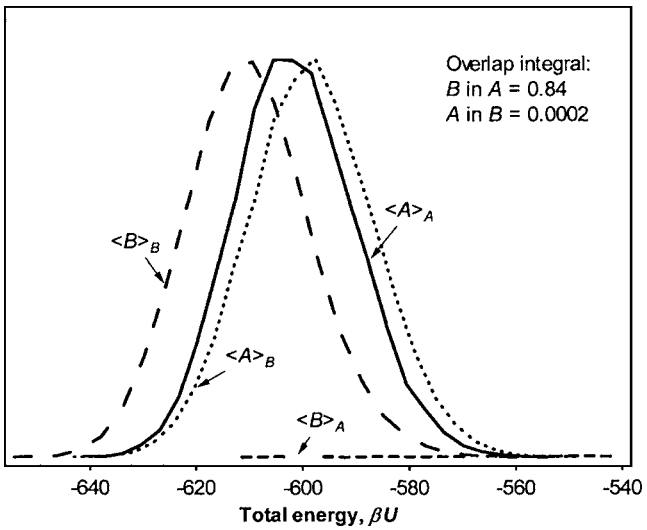

FIG. 4. Particle-insertion total-energy pair distributions and the overlap integral values for the Lennard-Jones system at a density $\rho^*=0.9$. Description of the curves is as in Fig. 2.

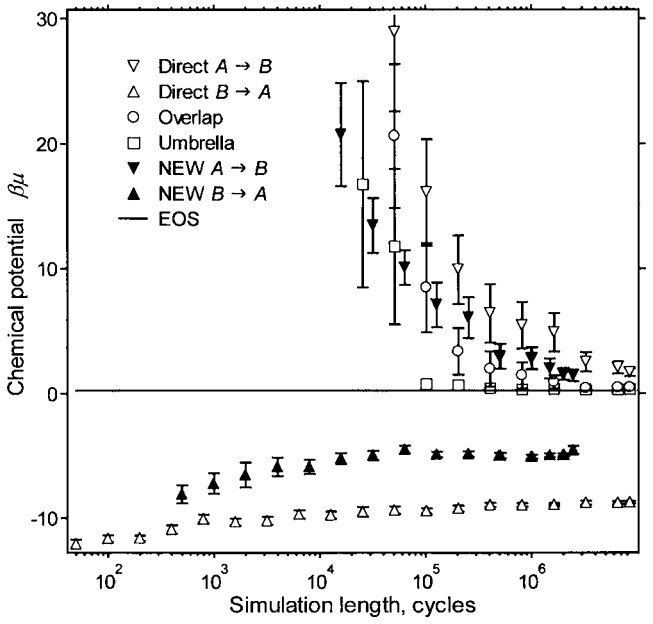

FIG. 5. Chemical-potential calculations of Lennard-Jones system, same as Fig. 3, except $\rho^*=0.9$.

the distribution $\langle A\rangle_B$ has shifted to the right of $\langle A\rangle_A$, and this means that part of the B system's phase space is outside A. We have $K_{AB}=0.0002$ and $K_{BA}=0.84$. The system's phase-space overlap relation has evolved from subset to overlap [Fig. 1(a)]. Correspondingly the difficulties of the free-energy calculation will increase, and the FEP insertion calculation will no longer be reliable. Results from the application of various FEP methods are presented in Fig. 5. Compared to the previous case, we see that all the single-stage results are less accurate at the same sampling length. Even the insertion calculations retain some bias at the end of the sampling, consistent with the system's overlap relation as understood through the total-energy distributions. On the other hand, both umbrella- and overlap-sampling approaches introduce improvement to the calculations, and are capable of yielding accurate results for the amount of sampling used. The NEW approach again gives a modest improvement over simple single-stage FEP in each direction.

B. Water system

Next, we study a model water system in canonical ensemble, using 256 SPC water molecules at temperature $T=298$ K and density $\rho=1$ g/cm^3. We truncate the intermolecular interactions at 6.124 Å, and have not used any correction for long-range electrostatic interactions; we apply only a simple long-range LJ correction and a self-polarization energy to the results. We define system A as one having 256 water molecules and system B having 257 water molecules, so the free-energy difference is again the residual chemical potential. The total-energy distribution pairs are plotted in Fig. 6, where we see that the relations between the two pairs look like that for the LJ system at density 0.9, where distributions $\langle A\rangle_B$ is on the right of $\langle A\rangle_A$ and $\langle B\rangle_A$ is on the right of $\langle B\rangle_B$. With $K_{AB}=0.015$ and $K_{BA}=0.8$, this again indicates an overlap phase-space relation.

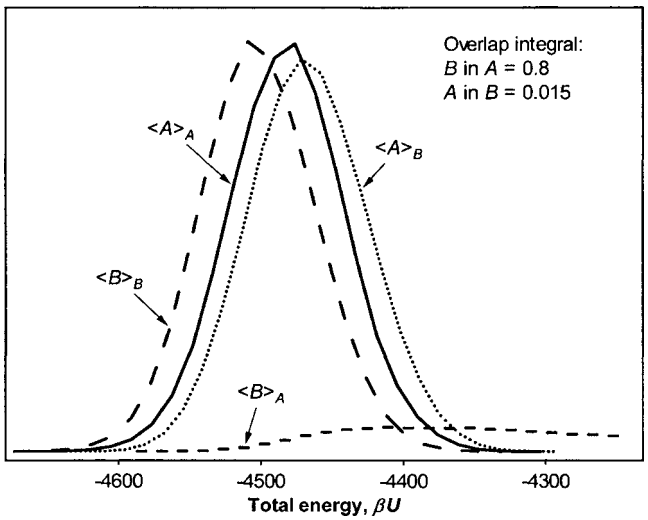

FIG. 6. Molecule-insertion total-energy pair distributions of water system and overlap integral measurements. Descriptions of the curves are same as in Fig. 2.

Results for the calculation of the chemical potential are given in Fig. 7, where they are compared to literature data for the same model.[12] The results are qualitatively similar to that seen for the 0.9-density LJ system, although the bias is larger (note the different scales in Figs. 5 and 7) and the insertion and deletion biases are a bit more symmetric. Overlap-sampling and umbrella-sampling methods both yield accurate results, converging to the correct chemical potential at about the same rate.

C. Ion-charging system

As the last example we examine a system demonstrating ionic hydration. We define system *A* having 216 SPC water molecules and a neutral LJ ion (neon), while system *B* has the same number of water molecules but with a charged ion

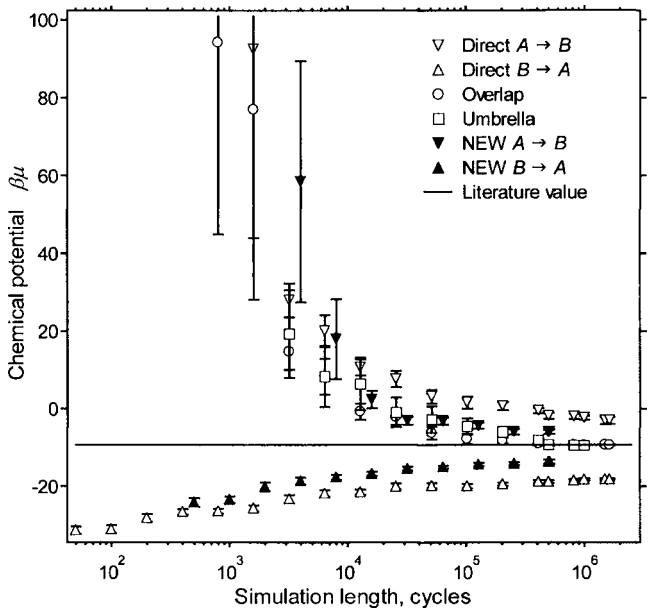

FIG. 7. Chemical-potential calculation for the water system. Descriptions of the plot is the same as in Fig. 3.

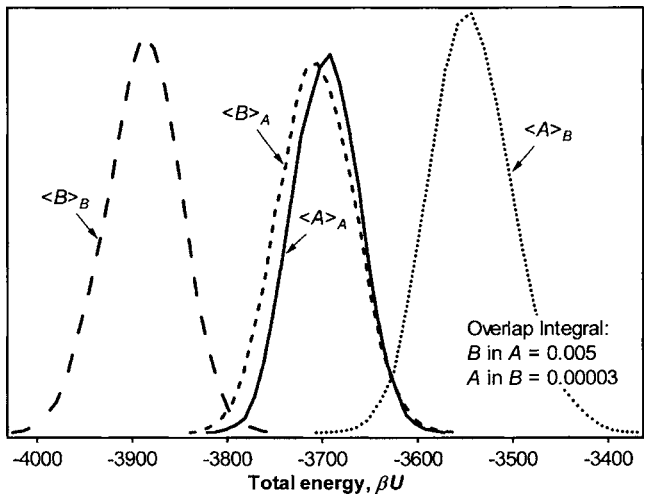

FIG. 8. Total-energy pair distributions and overlap integral measurements of the ion charging system. Descriptions of the curves are the same as in Fig. 2.

(sodium ion). The model parameters are as given in Ref. 13. Again we study the system in the canonical ensemble at temperature $T=298$ K and density $\rho=1$ g/cm^3. The system's total-energy distribution pairs are plotted in Fig. 8, for which the overlap integrals are $K_{AB}=3\times10^{-5}$ and $K_{BA}=0.005$. Each pair of distributions is widely separated, indicating that this system has a nonoverlap relation. This means clearly that FEP calculations in both insertion and deletion directions will be inaccurate. The lack of phase-space overlap indicates that the two-stage overlap method will not work well. Umbrella sampling is the only two-stage method with potential to yield an accurate result.

Results from the free-energy calculations are presented in Fig. 9, where they are compared with the established literature value.[13] Both single-stage methods fail completely, albeit symmetrically, so in this case (and this case only) the

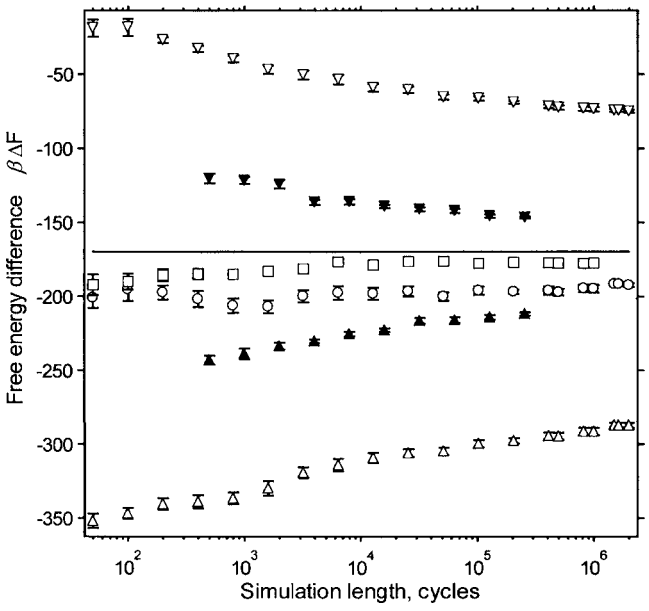

FIG. 9. Free-energy calculation of ion-charging system. System specifications are described in Sec. IV. Bias in the free energy is plotted against the simulation length, in cycles. Symbol descriptions are same as those for Lennard-Jones systems and water system (Figs. 3, 5, and 7).

average of the two yields a result better than either by themselves. Application of the NEW method gives considerable improvement, but is still very inaccurate. Overlap sampling is better still, but also shows significant bias by the end of the sampling. Simple umbrella sampling using the intermediate defined by Eq. (9) failed (not shown) because the system did not move well between the two important phase-space regions. As an *ad hoc* remedy, we considered an intermediate defined by Eq. (9), but at a temperature 1000 times higher (this scheme was employed in a previous study[5]). The results of this calculation are best of all the methods, but some bias is still present in the average. We believe that higher-order staging methods, or a combination of staging methods with more reversible NEW calculations, would be required to obtain an accurate result for this system.

D. Bias-detection measures

The foregoing discussion has dealt with the use of the phase-space overlap measures to understand qualitatively the performance of different staging approaches. Our previous work[2,14] has identified a bias-detection metric, and we consider now how well it works in identifying inaccuracy in the results discussed above. A bias-free result requires sufficient sampling for the scaled sampling measures defined in Eq. (4) to be positive. For single-stage FEP calculations, evaluation of the dissipated work is required for perturbations in both directions in order to evaluate the Π value in either direction. For multistage methods, this procedure must be applied to each stage used for the evaluation of the overall free-energy difference. This means that we must perform perturbations $A \to C$, $C \to A$, $B \to C$, and $C \to B$, all of them, when using either umbrella sampling or overlap sampling, if we wish to have a clear determination of whether sampling has converged for each of the two stages. Fortunately, the results needed for the "extra" perturbations (e.g., $C \to A$ and $C \to B$ for overlap sampling) require only the quickly convergent work averages (e.g., $\langle W_{C \to A} \rangle_C$) to obtain the sampling metrics. The free-energy averages required for computing Π values are already available from the staging calculations.

Table I lists values for the Π sampling metrics evaluated at the end of the sampling (as indicated in Figs. 3, 5, 7, and 9) for each FEP stage used in the results presented above. The table also shows the bias in the overall free energy that is given by the single- or two-stage calculation, as the case may be. The tabulated values are the "apparent" values of Π, calculated using only the simulation data (i.e., the Π values are not calculated using knowledge of the correct free-energy difference), so these values represent the type of information one would have when measuring the free-energy difference for a new system. In some of the cases, the value of one of the relative entropies (say, s_A) is measured to be infinity, and the Π values are then reported with the corresponding term factored out (e.g., $\sqrt{s_A}$ is factored out of $\Pi_{A \to B}$). Values calculated this way are indicated by an asterisk. The modification of course does not change the sign of Π, but it does make it more difficult to interpret its magnitude because this modified definition does not exhibit the universal behavior observed in the multiharmonic test cases used to aid their formulation.[2]

The tabled values demonstrate a clear connection between the sign of Π and the presence of bias in the calculated free-energy difference. The low-density LJ insertion calculations find that Π is positive for insertion, and for both stages of umbrella and overlap sampling, while it is negative for the deletion calculation, all in accord with the observed biases. The presence of bias in the high-density LJ insertion calculation is flagged by the negative value of Π. Here also the value of Π for one of the overlap-sampling stages is nearly zero but still positive; correspondingly there is a more significant bias in the free energy calculated using the method (although on the larger scale of Fig. 5 this bias looks less significant). The lesson here is that the crossover from biased to unbiased with increasing Π is not perfectly sharp. If possible one should aim for Π to be at least +0.5 or 1.0 before considering that the result is safely bias-free (again, however, this determination is problematic if dealing with Π^* instead of Π). For umbrella sampling, the values of Π are well into the positive region, and consistent with this we find that a bias-free result is obtained well before the end of the sampling used here.

For the water chemical potential, both insertion and deletion are flagged as inaccurate by negative values of Π, and the free-energy results are indeed biased. The overlap-sampling result has a slightly negative Π value for one of the stages, yet the free-energy difference does not show much bias. This again demonstrates the need for caution in interpreting the "gray area" in the vicinity of $\Pi = 0$. Umbrella sampling shows strongly positive values of Π for both stages, and the calculated free energy does not show much bias. Finally, in the ion-charging cases, all of the Π values are clearly negative, and significant bias is observed in all free-energy results.

To this point we have demonstrated that the bias-detection measures are effective in identifying the biases from different staging sampling strategies as indicated by the Π_{app} values in Table I. Another feature of this bias-detection heuristic is that we can project the continuous bias behavior (as a function of simulation length) observed during a simulation onto the picture of a single universal-bias curve. Figure 10 shows such a picture, where symbols represent the continuous biases from different sampling stages (which are also shown in Figs. 3, 5, 7, and 9) plotted against Π values defined by Eq. (4). Curves are reproductions of the multiharmonic cases studied previously (Fig. 7 in Ref. 2). It is clear that the biases from different cases approximately all collapse onto a single universal-bias curve. However, the calculation of the Π values requires the knowledge of a free-energy difference, which is apparently unknown at the moment. To test the fail-safe nature of the heuristic, we plot again all the biases shown in Fig. 10 as the function of apparent Π values (calculated using the biased free energy) in Fig. 11. We obtain the similar bias behaviors as examined previously,[2,14] such that all the biases are less than 0.1 when Π_{app} values are larger than zero. This means that the heuristic

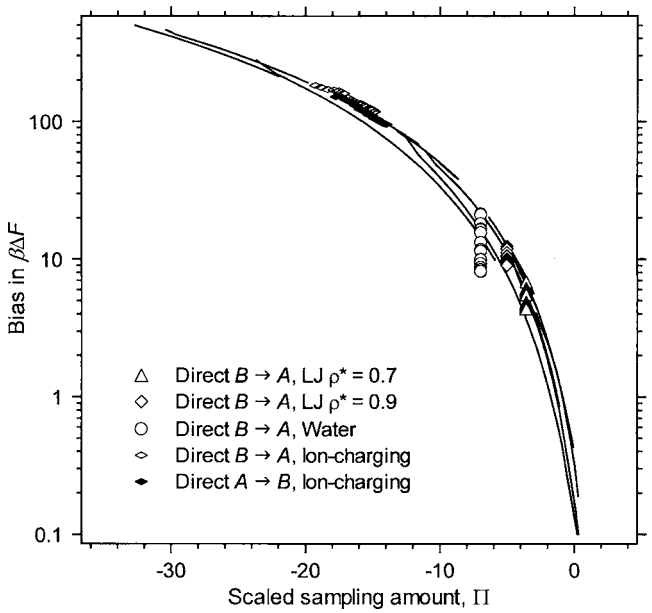

FIG. 10. Biases (in units of kT) of free-energy calculations from different cases are plotted against the scaled sampling amount Π defined in Eq. (4). Symbols represent the biases from the selected sampling cases for the three application systems studied in Sec. IV which are already plotted in Figs. 3, 5, 7, and 9. Solid curves represent the nine different overlap cases for the multiharmonic model studied previously (Fig. 7 in Ref. 2). This plot gives a clear picture that different bias behaviors presented this way tend toward a single universal bias curve.

will not give a false indication of a bias-free result even under the influence of the biased input. Thus it is a fail-safe method in detecting the bias.

V. CONCLUSIONS

In this and previous work we have argued that phase-space overlap is an important determinant of the performance of FEP calculations, and more generally NEW calculations (to the extent that they are performed irreversibly). The phase-space overlap picture appeals to one's intuition, as it explains the failures of the FEP methods in terms of simple concepts relating to the overlap of objects in space. Even though phase space is multidimensional and the important regions for a system can be quite complex, the basic idea is conveyed effectively using simple two-dimensional illustrations such as those given in Fig. 1.

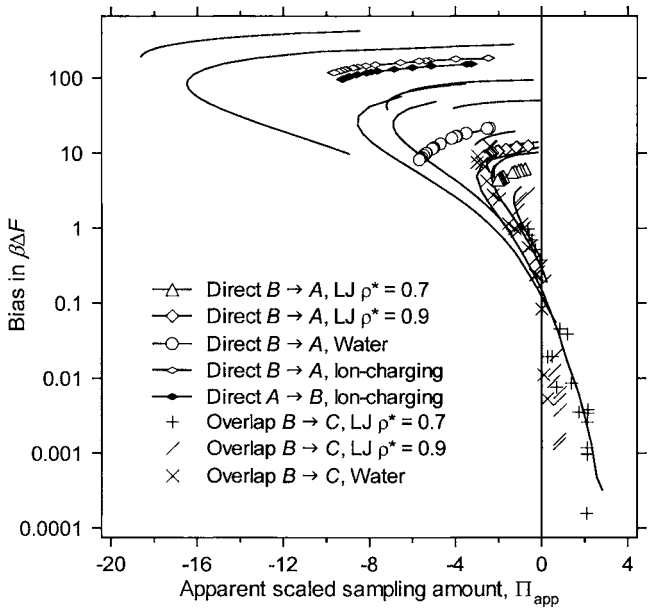

FIG. 11. The same biases studied in Fig. 10 (including three more cases) are plotted against the apparent scaled sampling amount Π_{app} defined previously (Ref. 2) and the curves represent cases of multiharmonic model previously plotted in Fig. 8 of Ref. 2.

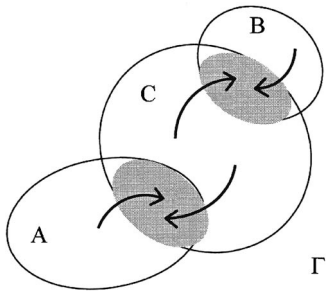

FIG. 12. An appropriate design of multistage sampling strategy for a system exhibiting a nonoverlap phase-space relation. The intermediate C system is formulated to connect A and B systems' important phase space with two overlap areas (shaded ovals). Perturbations are conducted in the directions from A, B, and C systems into the shaded overlap areas, thereby giving the A-B free-energy system via four free-energy calculation stages.

The value of this picture is manyfold. First, it directs us to the key issue involved in the success or failure of FEP methods, namely, the overlap of two distinct subspaces of phase space. In this way we are steered to formulate metrics that (unlike the free energy itself) quantify the geometric relation between the phase spaces. This has led us to consider total-energy distribution pairs, their overlap integrals, and the relative entropy. The latter has proven useful in formulating a metric for detecting bias in a FEP calculation. Second, the phase-space overlap picture helps us to understand the bias asymmetry often observed in FEP calculations. The asymmetric relation of the subspaces (superset versus subset) connects directly to the asymmetry of the bias. From this we know that we cannot rely on forward-reverse averaging to cancel errors in the hope of obtaining an accurate result. Third, the overlap picture gives us a conceptual framework for developing and applying staging methods. We know that in some circumstances single-stage calculations can never work, and that staging must be used. Then we can use the phase-space picture to select an appropriate staging method, and construct the stages. The choices are not too difficult for the staging methods presented here, but in other cases it is necessary to use more than two stages. It then becomes difficult to propose a general approach, but the ideas presented here can be used to guide the formulation of a multistage method for a given situation. As an example, in Fig. 12 we show how one might aim to formulate stages for a system (such as the ion-charging example) where the phase-space overlap is negligible for the systems of interest.

Our work introducing the scaled sampling metric Π showed that it is effective in identifying bias in free-energy calculations applied to a simple model system, the multiharmonic model.[2] In the present work we have further demonstrated the effectiveness of these quantities as an indicator of

ias in a FEP calculation. It seems clear that if this quantity is well below zero, the calculated free energy is inaccurate, while if it is well above zero, the result is reliable. Interpretation of values near zero will improve with application of the metric to different systems, and experience is developed regarding its use. Future work might also consider how the phase-space overlap picture can be used to design free-energy transition pathways, aid the application and interpretation of other free-energy methods (e.g., histogram methods), and—moving beyond computational applications—improve understanding of physical processes involving two systems defined on a common phase space.

ACKNOWLEDGMENTS

This work is supported by the Division of Chemical Sciences, Office of Basic Energy Sciences of the U.S. Department of Energy (Contract No. DE-FG02-96ER14677). Computer resources have been provided by the Center for Computational Research at the University at Buffalo, the State University of New York.

[1] C. Jarzynski, Phys. Rev. Lett. **78**, 2690 (1997); Phys. Rev. E **56**, 5018 (1997).
[2] D. Wu and D. A. Kofke, J. Chem. Phys. **123**, 054103 (2005).
[3] C. H. Bennett, J. Comput. Phys. **22**, 245 (1976).
[4] G. E. Crooks, Phys. Rev. E **61**, 2361 (2000).
[5] N. D. Lu, D. Wu, T. B. Woolf, and D. A. Kofke, Phys. Rev. E **69**, 057702 (2004).
[6] D. Frenkel and B. Smit, *Understanding Molecular Simulation: From Algorithms to Applications*, 2nd ed. (Academic, San Diego, 2002).
[7] G. M. Torrie and J. P. Valleau, J. Comput. Phys. **23**, 187 (1977).
[8] K. K. Han, Phys. Lett. A **165**, 28 (1992).
[9] N. D. Lu and D. A. Kofke, J. Chem. Phys. **115**, 6866 (2001).
[10] D. Wu and D. A. Kofke, J. Chem. Phys. **122**, 204104 (2005).
[11] J. K. Johnson, J. A. Zollweg, and K. E. Gubbins, Mol. Phys. **78**, 591 (1993).
[12] J. Hermans, A. Pathiaseril, and A. Anderson, J. Am. Chem. Soc. **110**, 5982 (1988).
[13] T. P. Straatsma and H. J. C. Berendsen, J. Chem. Phys. **89**, 5876 (1988).
[14] D. Wu and D. A. Kofke, J. Chem. Phys. **121**, 8742 (2004).