Equilibrium and nonequilibrium foundations of free energy computational methods

C. Jarzynski
Theoretical Division, Los Alamos National Laboratory, Los Alamos, NM
chrisj@lanl.gov

Abstract

Statistical mechanics provides a rigorous framework for the numerical estimation of free energy differences in complex systems such as biomolecules. This paper presents a brief review of the statistical mechanical identities underlying a number of techniques for computing free energy differences. Both equilibrium and nonequilibrium methods are covered.

LAUR-01-1412

Typeset using REVTEX

The development of *ab initio* methods of computing free energy differences represents an essential component of progress in computational biology and chemistry. Protein-ligand binding affinities, hydrophobic forces, potentials of mean force, chemical potentials, reaction times, thermodynamic stability ... all these are quantities either expressed as, or determined by, free energy differences. However, for all but the simplest of systems, computing a free energy difference ΔF can be notoriously time-consuming. Despite decades of effort, the first-principles calculation of ΔF for many problems of practical importance remains too slow for satisfaction, and the need for improved efficiency of free energy computations remains high.

The aim of this contribution is to present a brief review of several identities of statistical mechanics which provide the theoretical foundation for a number of free energy computational methods. Traditionally, these methods have relied on *equilibrium sampling*: estimates of ΔF are constructed from a number of randomly generated microstates of the system under consideration, and those microstates are assumed to be statistically representative of specified thermal equilibrium states of the system. It is convenient to think of the numerical generation of a sequence of such microstates as a *dynamical simulation*, representing the evolution of the system in thermal contact with a heat reservoir. Often, finite relaxation times mean that true equilibrium sampling is unattainable within a realistic amount of computer time, particularly when the aim is to sample from *numerous* equilibrium distributions. Roughly speaking, the system inevitably gets forced out of equilibrium during a numerical simulation whose aim is to generate microstates sampled from a sequence of equilibrium distributions. In the context of traditional methods, this is a nuisance, introducing systematic errors into the estimate of ΔF. In an effort to deal with this problem, methods have been developed which explicitly account for the fact that the sampling which occurs in practice does not coincide perfectly with the targeted equilibrium distributions. Indeed, in recent years it has been realized that even if the system is driven *far* from equilibrium, the value of ΔF can still be constructed, in principle, from a number of such simulations.

In Section I, the basic problem of computing a free energy difference ΔF is stated, followed by a brief discussion of equilibrium (canonical) sampling. Section II summarizes a number of free energy computational methods based on equilibrium sampling. (For more comprehensive reviews, see Refs.[1,2].) In Section III, nonequilibrium methods are discussed.

I. PRELIMINARIES

A. Statement of the problem

Consider some system with a finite number of degrees of freedom, and let a *microstate* of the system be represented by a point in phase space, $\mathbf{z} = (\mathbf{q}, \mathbf{p})$. If the system under consideration is a biomolecule, for instance, then $\mathbf{q}$ might denote the degrees of freedom specifying the location of each atom of the molecule, along with a number of solvent molecules, and $\mathbf{p}$ would be the collection of associated momenta. Next, let $H_\lambda(\mathbf{z})$ denote a parameter-dependent Hamiltonian, which gives the internal energy of the system as a function of microstate for a fixed setting of an external *work parameter*, λ. This work parameter might specify the strength of an externally applied field, or perhaps a volume of configuration space within which the system is confined (in which case $H_\lambda(\mathbf{z})$ is formally infinite for points falling outside of this region of space). The parameter need not be physically realizable. For instance, in *computational alchemy*[3], λ parametrizes the atom-atom interaction forces, so that by changing λ one type of atom or molecule is effectively transformed into another.

We will be interested in the situation when the system is in thermal contact with a heat reservoir at a temperature T. In this case an *equilibrium state* of the system, for a given value of λ, is represented by a canonical distribution in phase space,

$$\rho_\lambda = \frac{1}{Z_\lambda} e^{-H_\lambda(\mathbf{z})/T}, \tag{1}$$

where

$$Z_\lambda = \int d\mathbf{z}\, e^{-H_\lambda(\mathbf{z})/T} \tag{2}$$

is the corresponding partition function. (Throughout this paper, the dependence of Z_λ on temperature T will be suppressed, and Boltzmann's constant will be set to unity.) The *free energy* of this equilibrium state is then given by

$$F_\lambda = -T \, \ln Z_\lambda, \tag{3}$$

and, following convention, we will be interested in computing the free energy difference between the $\lambda = 0$ and $\lambda = 1$ equilibrium states:

$$\Delta F \equiv F_1 - F_0 = -T \, \ln \frac{Z_1}{Z_0}. \tag{4}$$

The problem of computing ΔF is thus one of computing a ratio of partition functions, and this is what renders it numerically challenging. The direct computation of Z_λ would require the evaluation of a multi-dimensional integral, and is typically out of the question: the effort grows exponentially with the dimensionality of phase space, becoming impractical for non-ideal systems with more than a modest number of degrees of freedom. Thus what is needed is a way to estimate Z_1/Z_0 (hence ΔF) without separately computing Z_0 and Z_1. In Secs.II and III, we review a number of statistical mechanical identities which, in principle, allow one to do just that.

We mention in passing that if the parametrization of the Hamiltonian takes the form $H_\lambda = H_0 + \lambda \Delta H$, where $\Delta H = H_1 - H_0$, then some of the identities presented below are modestly simplified.

B. Canonical sampling

The canonical distribution (Eq.1) is central to any discussion of first-principles estimation of free energy differences; ultimately, ΔF can be viewed as just a particular measure of the difference between two such distributions. It is not surprising, then, that methods of estimating ΔF rely on *sampling* from canonical distributions.

Generally speaking, canonical sampling algorithms – that is, algorithms for generating a sequence of microstates which can be viewed as having been randomly drawn from $\rho_\lambda(\mathbf{z})$ – fall into two classes: *molecular dynamics* (continuous) and *Monte Carlo* (discrete) algorithms.

In the former case, one numerically integrates equations of motion for the evolution of the microstate, $\mathbf{z}(t)$, meant to mock up the evolution of a system in contact with a heat reservoir. A sequence of microstates is then obtained by taking "snapshots" of the system at regularly spaced intervals, which might be as small as the time steps in the integration algorithm. In the discrete case, by contrast, the system evolves by finite *Monte Carlo steps* from one microstate to the next, resulting in a chain of microstates:

$$\mathbf{z}_0 \to \mathbf{z}_1 \to \cdots. \tag{5}$$

For remarkably simple prescriptions for generating these moves – most famously the Metropolis algorithm[4] – it can be shown that the sequence of microstates samples the canonical distribution. Throughout this paper, the discussion of free energy methods will be framed in terms of Monte Carlo algorithms, although the results themselves apply to many molecular dynamics schemes as well.

Eq.1 defines a family of canonical distributions, parametrized by the value of λ. We assume therefore that we have a parametrized family of sampling algorithms as well: when we implement the sampling corresponding to a particular value of λ, we get a chain of microstates drawn from the associated distribution $\rho_\lambda(\mathbf{z})$.

Associated with any canonical sampling algorithm is an inherent *relaxation time.* In the context of discrete, Monte Carlo algorithms, this is the number of steps which must be taken, starting from any initial microstate $\mathbf{z}_0$, before the current microstate becomes statistically representative of the targeted canonical distribution. To be more precise, imagine using some arbitrary prescription to randomly choose infinitely many initial microstates, $\mathbf{z}_0$. This *ensemble* will be described by a distribution $f_0(\mathbf{z}_0)$. Now imagine evolving each member of this ensemble by a single Monte Carlo step: $\mathbf{z}_0 \to \mathbf{z}_1$; the distribution of the new microstates will be given by some $f_1(\mathbf{z}_1)$. Iterating the process, we get a progression of phase space distributions:

$$f_0(\mathbf{z}_0) \to f_1(\mathbf{z}_1) \to \cdots \to f_n(\mathbf{z}_n) \to \cdots, \tag{6}$$

where f_n describes the ensemble after n Monte Carlo steps. When we say that the Monte Carlo algorithm *samples the canonical distribution* $\rho_\lambda(\mathbf{z})$, we mean that

$$\lim_{n\to\infty} f_n = \rho_\lambda. \tag{7}$$

The relaxation time is the number of steps characterizing this "relaxation" to ρ_λ.

Finite relaxation times are particularly relevant in the context of certain of the computational methods discussed below, where the sampling takes place as the value of λ itself evolves. That is, consecutive Monte Carlo steps are generated using an ever-changing algorithm, corresponding to small increments in λ from one step to the next:

$$\lambda_0 \to \lambda_1 \to \cdots \to \lambda_t \to \cdots. \tag{8}$$

(Here we use the subscript t rather than n, anticipating later sections in which we explicitly view this as a time-dependent, dynamical process.) If such a process is carried out *quasi-statically* – i.e., if we take infinitely many Monte Carlo steps, changing λ infinitesimally between steps – then in effect we "sweep through" a quasi-continuous sequence of canonical distributions. Each $\mathbf{z}_t$ is then statistically representative of the canonical distribution corresponding to the current parameter value: $f_t = \rho_{\lambda_t}$. In practice, however, if we are to change λ we must do so in finite increments, and then a *lag* develops[5,6]: the instantaneous canonical distribution ρ_{λ_t} becomes in effect a moving target, and f_t (the distribution from which $\mathbf{z}_t$ is sampled) is unable to keep up with this target, as a result of the finite relaxation time. The more rapidly we change λ, the more significant the lag.

II. EQUILIBRIUM METHODS

In this section we review four identities (Eqs.9, 13, 16, and 21 below) for ΔF, the free energy difference between two equilibrium states of a system. Each represents the theoretical justification for a particular method of estimating ΔF from a number of sampled microstates. In each case these microstates are, ideally, drawn from canonical distributions, hence these methods are explicitly based on *equilibrium* sampling. Moreover, these methods can be interpreted as limiting cases of a single formula (Eq.26), which gives an estimate of ΔF in terms of a long chain of microstates.

A. Free Energy Perturbation

We begin with perhaps the most widely used identity for free energy differences[7]:

$$e^{-\Delta F/T} = \left\langle e^{-\Delta H/T} \right\rangle_0. \tag{9}$$

Here, $\Delta H(\mathbf{z}) \equiv H_1(\mathbf{z}) - H_0(\mathbf{z})$ is the energy difference associated with changing the work parameter from one value ($\lambda = 0$) to another ($\lambda = 1$), while holding fixed the microstate $\mathbf{z}$. The angular brackets $\langle \cdots \rangle_0$ denote an average over microstates sampled from the canonical distribution $\rho_0(\mathbf{z})$. The derivation of Eq.9 could hardly be simpler:

$$\left\langle e^{-\Delta H/T} \right\rangle_0 = \int d\mathbf{z}\, \rho_0(\mathbf{z}) e^{-\Delta H(\mathbf{z})/T} \tag{10}$$

$$= \frac{1}{Z_0} \int d\mathbf{z}\, e^{-H_1(\mathbf{z})/T} = \frac{Z_1}{Z_0}, \tag{11}$$

using Eqs.1,2.

Eq.9 is the basis of the *free energy perturbation* method of estimating ΔF, which amounts to averaging $e^{-\Delta H/T}$ over microstates sampled from the canonical distribution ρ_0:

$$e^{-\Delta F/T} \approx \frac{1}{N} \sum_{n=1}^{N} e^{-\Delta H(\mathbf{z}_n)/T}, \tag{12}$$

where $\mathbf{z}_1, \cdots, \mathbf{z}_N$ denote the N sampled microstates. By Eq.9, this approximation becomes an equality in the limit of infinitely many samples, $N \to \infty$.

The perturbation method runs into practical difficulties if the $\lambda = 0$ and $\lambda = 1$ equilibrium states are significantly dissimilar. More precisely, if the canonical distributions $\rho_0(\mathbf{z})$ and $\rho_1(\mathbf{z})$ overlap very little in phase space, then the convergence of the perturbation estimate (the right side of Eq.12) will be slow. Heuristically, this makes sense: ΔF quantifies a difference between two canonical distributions; if we sample microstates typical of one distribution (ρ_0) but atypical of the other (ρ_1), then we will very slowly accumulate information about the latter. At the level of implementation, we will find in this situation that a small fraction of the sampled microstates produce relatively huge values of $e^{-\Delta H/T}$, so that the average is dominated by these few samples. Hence, most of the computational effort is devoted to generating microstates that have little impact on the average being computed, and consequently the estimate converges slowly.

A number of refinements of the perturbation method have been developed over the years. Perhaps most notable are Bennett's *overlapping distributions* method[8], and the *umbrella sampling* scheme proposed by Torrie and Valleau[9]. Interestingly, these can be viewed as complementary techniques, involving an intermediate phase space distribution which enjoys overlap with both ρ_0 and ρ_1.[10] (See also Refs.[1,2] for discussions of these and related methods.)

B. Window Sampling and Thermodynamic Integration

Given that poor convergence results from little overlap between the distributions ρ_0 and ρ_1, the following strategy naturally suggests itself: divide the λ interval $[0,1]$ into M "windows" $[\lambda_m, \lambda_{m+1}]$, where for instance $\lambda_m = m/M$, then use the perturbation method to compute the free energy difference associated with each window:

$$\Delta F = \sum_{m=0}^{M-1} \delta F_m \tag{13a}$$

$$\delta F_m = F_{\lambda_{m+1}} - F_{\lambda_m} = -T \, \ln\left\langle e^{-\delta H_m/T} \right\rangle_{\lambda_m}, \tag{13b}$$

where $\delta H_m \equiv H_{\lambda_{m+1}} - H_{\lambda_m}$. By choosing M sufficiently large, the overlap between any ρ_{λ_m} and $\rho_{\lambda_{m+1}}$ can be improved to the point where δF_m is computed easily using the perturbation method. This is known as *window sampling.*

It is interesting to consider window sampling in the limit $M \to \infty$. Applying Eq.9 to a particular window, and expanding the exponentials to first order in the window width, $\delta\lambda = M^{-1}$, we get

$$\delta F_m = \left\langle \delta H_m \right\rangle_{\lambda_m} + \mathcal{O}(\delta\lambda^2). \tag{14}$$

Dividing both sides by $\delta\lambda$ and taking the limit $M \to \infty$ then gives:

$$\frac{\partial F_\lambda}{\partial \lambda} = \left\langle \frac{\partial H_\lambda}{\partial \lambda} \right\rangle_\lambda . \tag{15}$$

This identity, due to Kirkwood[11], is the basis of the *thermodynamic integration* (TI) method of computing ΔF. The implementation is again straightforward: $\partial F_\lambda/\partial\lambda$ is estimated at a number of λ values, by averaging $\partial H_\lambda/\partial\lambda$ over microstates sampled from the corresponding canonical distributions, and the integral

$$\Delta F = \int_0^1 d\lambda \, \frac{\partial F_\lambda}{\partial \lambda} = \int_0^1 d\lambda \left\langle \frac{\partial H_\lambda}{\partial \lambda} \right\rangle_\lambda \tag{16}$$

is in turn estimated from these values.

Window sampling and thermodynamic integration rely on generating microstates from *numerous* canonical distributions, corresponding to parameter values $\lambda_0, \lambda_1, \cdots, \lambda_{M-1}$. The sampling from each of these distributions is usually preceded by a number of relaxation steps, during which the system adjusts to the value of λ. It is often convenient to use the final microstate sampled at λ_m as the seed for the relaxation sequence preceding the sampling at λ_{m+1} (rather than starting with a new microstate), since if the λ's are closely spaced, then a typical microstate sampled from ρ_{λ_m} will be "nearly typical" of $\rho_{\lambda_{m+1}}$, and therefore relatively few relaxation steps will be required. Thus, implementation of window sampling or thermodynamic integration might proceed as follows. Following generation of an initial microstate $\mathbf{z}_0$ sampled from the canonical distribution ρ_0, n_s *sampling steps* are taken with the work parameter held at λ_0, and the value of δF_0 is estimated – using either Eq.13b or Eq.14 – from the n_s microstates thus generated. The parameter λ is then changed from λ_0 to λ_1, and n_r *relaxation steps* are taken to allow the system to adjust to the new parameter value. The cycle is iterated, ultimately resulting in a long chain of microstates, with sampling intervals of length n_s alternating with relaxation intervals of length n_r as the value of λ marches through the sequence $\lambda_0, \lambda_1, \cdots, \lambda_{M-1}$. (Here the "length" of a relaxation or sampling interval just refers to the number of Monte Carlo steps in that interval.) The final estimate of ΔF is obtained by adding together the M estimates of δF_m, each computed from the values of δH_m measured during the corresponding sampling interval. The number of contributing values of δH_m is thus $M n_s$.

C. Slow Growth

In the above scheme, the total number of microstates in the chain is given by

$$\tau = M n_s + (M-1) n_r, \tag{17}$$

corresponding to M sampling intervals and $M-1$ relaxation intervals. (This count does not include the relaxation steps used to generate the initial microstate $\mathbf{z}_0$.) For a given amount of computer time – effectively, a given total number of steps, τ – one must strike a compromise between the number of λ intervals (M), and the number of relaxation steps (n_r) and sampling steps (n_s) taken at each λ. While there is no simple prescription for optimizing these quantities, given a fixed τ, a common implementation involves sampling only a single microstate at each parameter value ($n_s = 1$), and by dropping the relaxation steps altogether ($n_r = 0$), thus allowing for a huge number of tiny λ intervals.^{5} This is the *slow growth* method, which can be viewed as follows. After generation of the initial microstate $\mathbf{z}_0$, the value of λ is instantaneously "switched" from $\lambda_0 = 0$ to $\lambda_1 = 1/\tau$, resulting in a small change in the energy of the system, $\delta W_0 = H_{\lambda_1}(\mathbf{z}_0) - H_{\lambda_0}(\mathbf{z}_0)$. A new microstate is then generated, $\mathbf{z}_0 \to \mathbf{z}_1$, using a single Monte Carlo step taken at the parameter value λ_1. The process is then iterated, with evaluations of

$$\delta W_t = H_{\lambda_{t+1}}(\mathbf{z}_t) - H_{\lambda_t}(\mathbf{z}_t) = \delta H_t(\mathbf{z}_t) \tag{18}$$

alternating with Monte Carlo steps $\mathbf{z}_t \to \mathbf{z}_{t+1}$ generated at λ_{t+1}, where $\lambda_t = t/\tau$. This ultimately produces a chain of microstates

$$\mathbf{z}_0 \to \mathbf{z}_1 \to \cdots \to \mathbf{z}_\tau, \tag{19}$$

as λ progresses in small increments from 0 to 1, and at the end ΔF is estimated as the sum of the small energy changes accumulated by the sequence of changes in λ:

$$\Delta F \approx W \equiv \sum_{t=0}^{\tau-1} \delta W_t. \tag{20}$$

(The final microstate $\mathbf{z}_\tau$ does not actually contribute to this estimate, hence is unnecessary. However, for purpose of presentation, it is convenient to assume that the last Monte Carlo step is the one from $\mathbf{z}_{\tau-1}$ to $\mathbf{z}_\tau$, generated at $\lambda = 1$.) Note the subtle shift in interpretation: we now view each value of $\delta H_t(\mathbf{z}_t)$ as *a small change in the energy of the system* due to a sudden change in λ, rather than simply the value of a function δH_t at a sampled microstate $\mathbf{z}_t$. This shift represents a somewhat more dynamical point of view: we think of the chain in Eq.19 as a *trajectory*, depicting the evolution of the system (in discrete time steps t) as λ is switched incrementally from 0 to 1.

The slow growth approximation, Eq.20, becomes an equality in the *quasi-static limit* of infinitely many, infinitesimal increments in λ:

$$\Delta F = W_\infty \equiv \lim_{\tau\to\infty} \sum_{t=0}^{\tau-1} \delta W_t. \tag{21}$$

As this point is not immediately obvious – and perhaps not universally appreciated – it merits a brief, semi-quantitative discussion.

In the quasi-static limit, each microstate $\mathbf{z}_t$ is sampled from the instantaneous canonical distribution $\rho_{\lambda_t}(\mathbf{z})$; see Section I B. Now, as the work parameter advances across a tiny but fixed interval $[\lambda, \lambda + \Delta\lambda]$, the system takes $n_{\Delta\lambda} = \tau\Delta\lambda \gg 1$ Monte Carlo steps. The contribution from this interval to the slow growth estimate of ΔF is thus a sum of $n_{\Delta\lambda}$ values of δW_t:

$$F_{\lambda+\Delta\lambda} - F_\lambda \approx {\sum_t}' \delta W_t, \tag{22}$$

where $\sum_t'$ denotes a sum over $\lambda\tau \le t < (\lambda + \Delta\lambda)\tau$. But $\delta W_t = (\partial H_\lambda/\partial\lambda) \cdot \tau^{-1} + \mathcal{O}(\tau^{-2})$, hence

$${\sum_t}' \delta W_t \to \tau^{-1} {\sum_t}' \frac{\partial H_\lambda}{\partial\lambda}(\mathbf{z}_t) = \Delta\lambda \cdot \left[\frac{1}{n_{\Delta\lambda}} {\sum_t}' \frac{\partial H_\lambda}{\partial\lambda}(\mathbf{z}_t) \right], \tag{23}$$

to leading order. In the limit $\tau \to \infty$ (hence $n_{\Delta\lambda} \to \infty$), the term in square brackets converges to a unique value:

$$\lim_{\tau\to\infty} \frac{1}{n_{\Delta\lambda}} {\sum_t}' \frac{\partial H_\lambda}{\partial\lambda}(\mathbf{z}_t) = \left\langle \frac{\partial H_\lambda}{\partial\lambda} \right\rangle_\lambda + \mathcal{O}(\Delta\lambda). \tag{24}$$

Now summing up over adjacent intervals of width $\Delta\lambda$ spanning $[0, 1]$, and finally taking the limit $\Delta\lambda \to 0$ (*after* having taken $\tau \to \infty$), we get:

$$\lim_{\tau\to\infty} \sum_{t=0}^{\tau-1} \delta W_t = \lim_{\Delta\lambda\to 0} \sum_{\Delta\lambda} \Delta\lambda \left[\left\langle \frac{\partial H_\lambda}{\partial\lambda} \right\rangle_\lambda + \mathcal{O}(\Delta\lambda) \right] = \int_0^1 d\lambda \left\langle \frac{\partial H_\lambda}{\partial\lambda} \right\rangle_\lambda = \Delta F. \tag{25}$$

D. Synthesis

The methods discussed so far – free energy perturbation, window sampling, thermodynamic integration, and slow growth – can all be viewed as special cases of a general formula for the numerical estimation of ΔF, namely:

$$\Delta F^{\mathrm{est}}(M, n_s, n_r) = \sum_{m=0}^{M-1} -T \ln \left[\frac{1}{n_s} \sum_{n=1}^{n_s} e^{-\delta H_m(\mathbf{z}_{m,n})/T} \right]. \tag{26}$$

Here, we are assuming the procedure discussed above: sampling intervals of length n_s alternate with relaxation intervals of length n_r, at M discrete values of the work parameter; $\mathbf{z}_{m,n}$ represents the n'th microstate sampled at λ_m. $\Delta F^{\mathrm{est}}(M, n_s, n_r)$ is the numerical estimate of ΔF, for given values of M, n_s, and n_r. Note that the dependence of the right side of Eq.26 on n_r is implicit rather than explicit: the microstates generated during a given relaxation intervals do not contribute directly to the estimate of ΔF, but rather "set the stage" for the subsequent sampling interval.

The identities on which the various methods are based (Eqs.9, 13, 16, 21) are recovered as limiting cases of Eq.26:

free energy perturbation	$M = 1$	$n_s \to \infty$	
window sampling	$M > 1$	$n_s \to \infty$	
thermodynamic integration	$M \to \infty$	$n_s \to \infty$	
slow growth	$M \to \infty$	$n_s = 1$	$n_r = 0$

Note that $M = \tau$ for slow growth, i.e. one Monte Carlo step per λ interval. Furthermore, when $M \gg 1$ (as in slow growth and thermodynamic integration), it is convenient to include only the leading-order contribution to the term summed on the right side of Eq.26:

$$\Delta F^{\mathrm{est}}(M, n_s, n_r) = \sum_{t=0}^{M-1} \frac{1}{n_s} \sum_{n=1}^{n_s} \delta H_m(\mathbf{z}_{m,n}) + \mathcal{O}(M^{-1}). \tag{27}$$

The various limiting cases listed above can be summarized by the identity

$$\Delta F = \lim_{M n_s \to \infty} \Delta F^{\mathrm{est}}(M, n_s, n_r). \tag{28}$$

This tells us that the estimate, ΔF^{est}, converges to the exact value of ΔF *as the total number of contributing values of δH_m goes to infinity.*

III. NONEQUILIBRIUM METHODS

The free energy methods discussed to this point rely on the assumption of *equilibrium* sampling: each microstate which actually contributes to the estimate of ΔF is assumed to have been drawn from a canonical distribution $\rho_\lambda(\mathbf{z})$. This is usually an idealization, as most sampling algorithms converge only asymptotically to the targeted distribution. Sometimes this idealization is a good one. For instance, in the basic implementation of the perturbation method, it is often feasible to take sufficiently many relaxation steps prior to the commencement of sampling, that the subsequent microstates are to a very good approximation drawn from $\rho_0(\mathbf{z})$. In other cases, the equilibrium assumption is noticeably violated. This is particularly evident in slow growth, where no relaxation steps are taken once the value of λ begins to change.

This section discusses the use of *nonequilibrium* methods of estimating ΔF. As in the equilibrium case, these methods are based on the sampling of microstates, but here it is explicitly *not* assumed that these are drawn from canonical distributions. The motivation for developing such methods is to some extent a desire to face reality, especially in the context of slow growth: if the system is going to be driven away from equilibrium by the finite rate of switching the work parameter, then we ought to develop ways to cope with this inevitability. However, nonequilibrium methods can also be useful in their own right. Even when we have available the computer time to perform a nearly quasi-static slow growth computation, there may be advantages to using a nonequilibrium method instead.

A. Dynamical interpretation

In Section II C we mentioned the dynamical interpretation of the chain of microstates $\mathbf{z}_0 \to \mathbf{z}_1 \to \cdots \to \mathbf{z}_\tau$ generated during a slow growth estimation of ΔF. Namely, we view this chain as a trajectory depicting the evolution (in discretized time) of our system, as the work parameter λ is changed in small increments from 0 to 1. We now elaborate on this interpretation, which plays a central role in the free energy methods discussed below.

We interpret our trajectory specifically as representing the evolution of a system *in contact with a heat reservoir at temperature T*. A slow growth calculation then represents the numerical simulation of the following *switching process*: an initially equilibrated system evolves with time, in contact with a heat reservoir, as an external work parameter is switched from 0 to 1. The total number of steps, τ, represents the *switching time*, i.e. the duration of the switching process; and $1/\tau$ is the *rate of switching*. If the process is not carried out sufficiently slowly, then the system gets driven away from equilibrium as a result of the finite rate of variation of the work parameter. That is, the lag mentioned in Section I B develops.

Recall that the quantity W defined by Eq.20 is the sum of energy changes resulting from increments in λ (see also Eq.18). This is *not* equal to the net change in the internal energy of the system, since it does not include energy changes due to the Monte Carlo steps, the sum of which we will denote by:

$$Q = \sum_{t=0}^{\tau-1} \delta Q_t \equiv \sum_{t=0}^{\tau-1} \Big[H_{\lambda_{t+1}}(\mathbf{z}_{t+1}) - H_{\lambda_{t+1}}(\mathbf{z}_t) \Big]. \tag{29}$$

As easily verified, the net change in internal energy of the system, $\Delta E \equiv H_1(\mathbf{z}_\tau) - H_0(\mathbf{z}_0)$, is given by:

$$\Delta E = W + Q, \tag{30}$$

with W and Q defined above.

The use of the symbols W and Q is meant to be suggestive: we interpret W as the external *work* performed on the system over the course of the switching process, by whatever agent drives the work parameter from 0 to 1.[5,12,13] Then Q is the net *heat* absorbed by the system, and Eq.30 is simply a statement of the first law of thermodynamics. This point of view allows us to interpret the foundation of the slow growth method, Eq.21, in terms of another basic law of thermodynamics, which states that *the external work performed on a system over the course of a reversible, isothermal process is equal to the free energy difference between the initial and final states of the system.*[14] In effect, the slow growth method represents an attempt to compute ΔF by simulating such a process.

To be truly reversible, a switching process must be carried out infinitely slowly ($\tau \to \infty$). For switching processes carried out at a finite rate, the second law of thermodynamics tells us that the work performed actually represents an *upper bound* on the free energy difference[12,13]:

$$W > \Delta F. \tag{31}$$

In other words, to the extent that the system gets driven out of equilibrium, additional work is required to change λ at the specified rate. Let us now consider this inequality in greater detail, as this will lead naturally to consideration of the use of *repeated*, nonequilibrium switching simulations to estimate ΔF.

B. Statistical and systematic errors

For a given switching process, the value of W which emerges from a simulation depends on a string of random numbers: those used during relaxation to the initial microstate $\mathbf{z}_0$, and those used to generate the subsequent Monte Carlo steps $\mathbf{z}_t \to \mathbf{z}_{t+1}$. If we were to carry out the same switching process repeatedly, keeping all things the same except the string of random numbers, then we would obtain a collection of different trajectories, and correspondingly different values of W. These represent different microscopic *realizations* – or *histories* – of the same switching process, with values of W which differ from one realization to the next as a result of microscopic fluctuations. Now, Eq.31 is not necessarily true for every realization of a given process, but is true *on average*:

$$\overline{W} > \Delta F, \tag{32}$$

where the overbar now indicates an average over the ensemble of possible realizations of the given switching process.[15] Thus, by performing N independent switching simulations, we obtain N independent work values, scattered around an average greater than ΔF.

Imagine that we have indeed performed N such simulations, perhaps using N different computers or processors, and let $W_1, \cdots, W_N$ be the values of work obtained from these simulations. How do we construct an estimate of ΔF from these values? Perhaps the first estimate that comes to mind is simply the ordinary (linear) average of these values:

$$\Delta F \approx \frac{1}{N} \sum_{n=1}^{N} W_n. \tag{33}$$

This average is of course subject to *statistical error*, which is easily estimated as $\sigma_W / \sqrt{N-1}$, where σ_W^2 is the variance of the N work values. More problematic is the *systematic error* – due to the fact that on average the work W will over-estimate the free energy difference ΔF (Eq.32) – which does *not* vanish in the limit $N \to \infty$. How do we cope with this bias?

Reinhardt *et al*[12,13] have suggested using Eq.32 to place upper and lower bounds on ΔF. A number of *forward* switching simulations (with λ switched from 0 to 1) are performed, and the average work is taken as an upper bound on ΔF. Then a number of *reverse* simulations ($\lambda : 1 \to 0$) are carried out, and the average of these work values represents an upper bound on $-\Delta F = F_0 - F_1$; hence a lower bound on ΔF. Combining the two sets of simulations, we get

$$-\overline{W}_{1\to 0} < \Delta F < \overline{W}_{0\to 1}. \tag{34}$$

As discussed in Refs.[12,13], minimizing the difference between the upper and lower bounds is an objective criterion for optimizing the parametrization of H_λ, given fixed "end points" H_0 and H_1. This is the *variational path optimization* scheme. Very recently, this method has been used in conjunction with a *metric scaling* strategy[16]; the combination shows promise of dramatically improving the efficiency of certain free energy calculations.

Taking a different approach, Hermans[17] has related the systematic bias ($\overline{W} - \Delta F$) to the *variance* of the work values, σ_W^2:

$$\Delta F = \overline{W} - \frac{\sigma_W^2}{2T} + \mathcal{O}(\tau^{-2}). \tag{35}$$

Thus, by adjusting the estimate of ΔF downward by an amount $\sigma_W^2/2T$, we remove the leading-order systematic error. Here, both $\overline{W}$ and σ_W^2 are defined with respect to infinitely many independent realizations of the same switching process. In practice, one estimates these quantities from a finite number of realizations. Eq.35 is a *near*-equilibrium result. Therefore, if the parameter λ is switched rapidly enough to drive the system significantly away from equilibrium over the course of a typical simulation, then the $\mathcal{O}(\tau^{-2})$ corrections to Eq.35 may be large.

C. Fast Growth

In recent years, the following *non-equilibrium work relation* has been derived:[20,21]

$$\overline{e^{-W/T}} = e^{-\Delta F/T}. \tag{36}$$

(This result was subsequently shown to follow from a finite-time extension of detailed balance[22], and more recently from the well-known Feynman-Kac theorem of stochastic processes[23].) Again, the overbar denotes an average over an ensemble of realizations. Eq.36 suggests the following *fast growth* method of computing free energy differences: N independent switching simulations are performed, and then the *exponential average* of the work values, W^x, is taken as the estimate of the desired free energy difference:

$$\Delta F \approx W^x \equiv -T \ln\left(\frac{1}{N}\sum_n e^{-W_n/T}\right). \tag{37}$$

By Eq.36, this approximation becomes an equality in the limit of infinitely many simulations, $N \to \infty$, *for any value of* τ:

$$\Delta F = \lim_{N\to\infty} W^x \qquad , \qquad \text{arbitrary } \tau. \tag{38}$$

Thus, no matter how slowly or rapidly each simulation is carried out, the value of ΔF can be estimated to arbitrary accuracy, given sufficiently many simulations. This remains true even if the system is driven far from equilibrium as λ is varied from 0 to 1.

We can understand Eqs.37 and 38 as follows. For a finite number N of independent switching simulations, the fast growth estimate of ΔF, Eq.37, is subject to both statistical and systematic error.[21,24] As we perform more and more simulations, however, *both the statistical and the systematic errors vanish.* Thus, the exponential average W^x converges to ΔF as $N \to \infty$, in contrast with ordinary average which converges to a value $\overline{W} > \Delta F$.

It is interesting to consider the relationship between fast growth and some of the previously discussed free energy methods. First, consider the extreme case in which the value of λ is switched from 0 to 1 in a single step ($\tau = 1$). In this situation, $W = \Delta H(\mathbf{z}_0)$ (see Eqs.18 and 20), and the average over "trajectories" is simply an average over microstates $\mathbf{z}_0$ sampled from the $\lambda = 0$ canonical distribution. Hence, fast growth reduces to the free energy perturbation method in this limit of sudden switching. By contrast, when $\tau \to \infty$ fast growth becomes equivalent to slow growth: $W = \Delta F$ for every realization (Eq.21), and so the average of $\exp(-W/T)$ is trivially $\exp(-\Delta F/T)$. Thus, at the two ends of the spectrum – namely, instantaneous switching ($\tau = 1$) and quasi-static switching ($\tau \to \infty$) – fast growth reduces to two tried and true methods of computing free energy differences. The real novelty of Eq.36 resides in its validity for all *intermediate* values of the switching time ($1 < \tau < \infty$), corresponding to simulations during which the system is genuinely driven out of equilibrium.

Combining Eq.36 with Jensen's inequality, $\overline{\exp x} \geq \exp \overline{x}$ (see Ref.[1], p. 137), we immediately obtain

$$\overline{W} \geq \Delta F. \tag{39}$$

The equality holds only in the reversible limit $\tau \to \infty$, hence Eq.36 implies $\overline{W} > \Delta F$ for irreversible processes. Recall that this is the theoretical basis of the variational path optimization method.

Finally, taking the logarithm of both sides of Eq.36, then expanding $\ln \overline{\exp -W/T}$ in terms of cumulants of W[20] and keeping only the first two cumulants, we get Hermans' result

$$\Delta F \approx \overline{W} - \frac{\sigma_W^2}{2T}. \tag{40}$$

As discussed elsewhere[24], truncation after the second term in the cumulant expansion ought to be valid precisely when the switching is sufficiently slow to maintain the system near equilibrium, in agreement with the discussion following Eq.35.

Eq.36 thus offers a common point of contact for a number of earlier free energy identities and methods.

Because fast growth drops the requirement of reversibility, it allows us to estimate ΔF using switching simulations of considerably shorter duration than with slow growth; there is no explicit need to maintain the system near equilibrium. The price paid, however, is the need for numerous simulations, as the convergence of W^x to ΔF is guaranteed only in the limit $N \to \infty$. Thus, as with methods based on equilibrium sampling, fast growth only recovers the exact value of ΔF if we devote an infinite amount of computational time to the problem. A question of practical importance is therefore: given a fixed amount of computer time, which method is likely to produce the best estimate of ΔF? In other words, is it better to devote all the computer time to a single, long simulation, or to perform a number of shorter one and compute the exponential average of the corresponding work values? In Ref.[24], this question was addressed in the context of computing the excess chemical potential for a (modified) Lennard-Jones argon fluid. (The excess chemical potential is the free energy difference associated with "turning on" the interactions between a tagged particle and the rest of the fluid.) For various values of the switching time τ, fast growth was compared with slow growth. It was found that, except for the smallest value of τ, fast and slow growth yielded comparably accurate estimates of ΔF, for the same amount of computational effort. As discussed in greater detail in Ref.[24], this suggests two possible advantages of fast growth. The first is the easy estimation of statistical errors, as ΔF is obtained from a number of independent values of W, in contrast with slow growth which produces only a single value. The second is the parallelizability of fast growth: it is much simpler to let N copies of a simulation code run independently on N processors, than to efficiently distribute a single simulation code over those processors. It should be stressed that these conclusions have been reached in the context of the particular system studied in Ref.[24]. Whether they are more generally valid remains to be seen.

The nonequilibrium work relation on which fast growth is based, Eq.36, is similar in structure to the free energy perturbation identity, Eq.9. This means that it is subject to the same potential problem of poor convergence: if the distribution of work values is very wide, then W^x will be dominated by the small fraction of simulations which happen to produce the lowest values of W. On the other hand, it may be possible to take advantage of the similarity between Eqs.9 and 36: a number of the refinements developed over the years for improving the efficiency of the perturbation method might easily carry over to fast growth. Frenkel[25] has suggested a version of fast growth analogous to Bennett's overlapping distributions method. Hummer[26] has shown that higher-order cumulant expansions derived in the context of the perturbation identity, Eq.9, are readily extended to Eq.36. Hu, Yun, and Hermans[27] have found empirically that taking the ordinary average of two exponential averages – $W^x_{0\to 1}$ obtained from a set of forward switching simulations, and $W^x_{1\to 0}$ from a set of reverse switching simulations – can yield an estimate of ΔF in which the systematic errors inherent in both $W^x_{0\to 1}$ and $W^x_{1\to 0}$ cancel.

IV. CONCLUSION

The computation of a free energy difference ΔF is ultimately a problem in statistical mechanics. The purpose of this paper has been a review of a number of rigorous results – specifically, Eqs.9, 16, 21, 34, 35, and 36 — which provide the theoretical basis for various methods of computing free energy differences. As these results are not completely independent of one another, an effort has been made to point out the relationships between them.

We end by mentioning a closely related and important problem. In this paper the free energy has been considered to be a function of an externally controlled work parameter, λ (see Eq.3). In many cases of interest, however, it is more physically relevant to define the free energy as a function of an *order parameter* of the system, χ; then $F(\chi)$ is a *potential of mean force*[1]. Several of the free energy techniques discussed in this paper can be modified so as to allow for the computation of potentials of mean force. For instance, the *weighted histograms* method[28,29] is essentially an extension of the free energy perturbation method (or rather its refinement, umbrella sampling). More recently, schemes have been developed for reconstructing potentials of mean force from *steered molecular dynamics* simulations[30], roughly analogous to the slow growth method. Finally, Hummer and Szabo[23] have in effect introduced a fast growth method for using steered molecular dynamics to compute potentials of mean force. It bears mention that the choice of order parameter is itself not a trivial problem, especially in reactions in which the transition path from the initial to the final state is not obvious; for recent progress on this aspect of the problem, see Ref.[31].

REFERENCES

[1] D.Chandler, *Introduction to Modern Statistical Mechanics* (Oxford University, New York, 1987).

[2] D.Frenkel and B.Smit, *Understanding Molecular Simulation: From Algorithms to Applications*, Academic Press, San Diego (1996).

[3] M.Karplus and G.A.Petsko, Nature **347**, 631 (1990).

[4] N.Metropolis *et al*, J.Chem.Phys.**21**, 1087 (1953).

[5] T.P.Straatsma, H.J.C.Berendsen, and J.P.M.Postma, J.Chem.Phys. **85**, 6720 (1986).

[6] D.A.Pearlman and P.A.Kollman, J.Chem.Phys. **91**, 7831 (1989).

[7] R.Zwanzig, J.Chem.Phys.**22**, 1420 (1954).

[8] C.H.Bennett, J.Comp.Phys.**22**, 245 (1976).

[9] G.M.Torrie and J.P.Valleau, J.Comp.Phys.**23**, 187 (1977).

[10] R.J.Radmer and P.A.Kollman, J.Comp.Chem.**18**, 902 (1997).

[11] J.G.Kirkwood, J.Chem.Phys. **3**, 300 (1935).

[12] W.P.Reinhardt and J.E.Hunter III, J.Chem.Phys.**97**, 1599 (1992).

[13] J.E.Hunter III, W.P.Reinhardt, and T.F.Davis, J.Chem.Phys. **99**, 6856 (1993).

[14] L.D.Landau and E.M.Lifshitz, *Statistical Physics*, 3rd ed., Part 1, section 15 (Pergamon Press, Oxford, 1990).

[15] Moreover, the larger the system, the smaller the probability of randomly generating a trajectory which violates Eq.31; thus, in the macroscopic limit we recover the statement that W is "strictly" greater than ΔF for irreversible processes.

[16] M.A.Miller and W.P.Reinhardt, J.Chem.Phys.**113**, 7035 (2000).

[17] J.Hermans, J.Phys.Chem.**95**, 9029(1991). See Refs.[18] and[19] for closely related results.

[18] R.H.Wood, W.C.F.Mühlbauer, and P.T.Thompson, J.Phys.Chem.**95**, 6670 (1991).

[19] L.-W.Tsao, S.-Y.Sheu, and C.-Y.Mou, J.Chem.Phys.**101**, 2302 (1994).

[20] C.Jarzynski, Phys.Rev.Lett.**78**, 2690 (1997).

[21] C.Jarzynski, Phys.Rev.E **56**, 5018 (1997).

[22] G.E.Crooks, J.Stat.Phys. **90**, 1481 (1998).

[23] G.Hummer and A.Szabo, Proc.Natl.Acad.Sci.(US) **98**, 3658 (2001).

[24] D.A.Hendrix and C.Jarzynski, J.Chem.Phys.**114**, 5974 (2001).

[25] D.Frenkel, private communication.

[26] G.Hummer, "Fast-growth thermodynamic integration: results for sodium ion hydration", preprint.

[27] H.Hu, R.H.Yun, and J.Hermans, "Reversibility of free energy simulations: slow growth may have a unique advantage. (With a note on use of Ewald summation.)", preprint.

[28] B.Roux, Comput.Phys.Comm. **91**, 275 (1995).

[29] S.Kumar, D.Bouzida, R.H.Swendsen, P.A.Kollman, and J.M.Rosenberg, J.Comp.Chem. **13**, 1011 (1992).

[30] See, for instance, J.R.Gullingsrud, R.Braun, and K.Schulten, J.Comp.Phys. **151**, 190 (1999), and references therein.

[31] Dellago, C., Bolhuis, P.G., Csajka, F.S. & Chandler, D., *J.Chem.Phys.* **108**, 1964 (1998); Bolhuis, P., Dellago, C. & Chandler, D., *Faraday Discussion Chem.Soc.* **110**, 421 (1998); Geissler, P. L., Dellago, C. & Chandler, D., *J.Phys.Chem.* **B103**, 3706 (1999); Geissler, P.L., Dellago, C., Chandler, D., Hutter, J. & Parinello, M., *Science*, in press (2001).