2.15

FREE-ENERGY CALCULATION USING NONEQUILIBRIUM SIMULATIONS

Maurice de Koning[1] and William P. Reinhardt[2]
[1] *University of São Paulo São Paulo, Brazil*
[2] *University of Washington Seattle, Washington, USA*

1. Introduction

Stimulated by the progress of computer technology over the past decades, the field of computer simulation has evolved into a mature branch of modern scientific investigation. It has had a profound impact in many areas of research including condensed-matter physics, chemistry, materials and polymer science, as well as in biophysics and biochemistry. Many problems of interest in all of these areas involve complex many-body systems and analytical solutions are generally not available. In this light, atomistic simulations play a particularly important role, giving detailed insight into the fundamental microscopic processes that control the behavior of complex systems at the macroscopic level. They provide key and effective tools for providing *ab initio* predictions, interpreting complex experimental data, as well as conducting computational "experiments" that are difficult or impossible to realize in a laboratory.

In this article, we will discuss one of the most fundamental and difficult applications of atomistic simulation techniques such as Monte Carlo (MC) [1] and molecular dynamics (MD) [2, 3]; the determination of those thermodynamic properties that require determination of the entropy. The entropy, the chemical potential, and the various free energies are examples of *thermal* thermodynamic properties. In contrast their *mechanical* counterparts such as the enthalpy, thermal quantities *cannot* be computed as simple time, or ensemble, averages of functions of the dynamical variables of the system and, therefore, are not directly accessible in MC or MD simulations. Yet, the free energies are often the most fundamental of all thermodynamic functions. Under appropriate constraints they control chemical and phase equilibria, and transition state estimates of the rates of chemical reactions. Examples of applications

range from determination of the influence of crystal defects on the mechanical properties of materials, to the mechanisms of protein folding. The development of efficient and accurate techniques for their calculation has therefore attracted considerable attention during the past fifteen years, and is still a very active field of research [4].

As detailed in the previous chapter [4], the evaluation of free energies (or, more specifically free-energy *differences*) requires simulations that collect data along a sequence of states on a *thermodynamic path* linking two equilibrium states. If the system is at equilibrium at every point along such a path, the simulated process is quasistatic and *reversible*, and standard thermodynamic results may be used to interpret collected data and to estimate the free-energy difference between the initial and final equilibrium states. The present chapter generalizes this approach to the case where data is collected during nonequilibrium, and thus *irreversible*, processes. Several important themes will emerge, making clear why this generalization is of interest, and how nonequilibrium calculations may be set up to provide both upper and lower bounds (and thus *systematic* in addition to *statistical* error estimates) to the desired thermal quantities. Additionally, the irreversible process may be *optimized* in a variational sense so as to improve such bounds. The statistical–mechanical theory of nonequilibrium systems within the regime of linear response will prove particularly helpful in this endeavor. Finally, newly developed re-averaging techniques have appeared that, in some cases, allow quite precise estimates of equilibrium thermal quantities directly from nonequilibrium data. The combination of such techniques with near-optimal paths can give well converged results from relatively short computations.

In the illustrations that follow, for sake of conciseness, we will limit ourselves to the application of nonequilibrium methods within the realm of the classical canonical ensemble. For this representative case the relevant thermodynamic variables are the number of particles N, the volume V, and the temperature T; and the appropriate free energy is the Helmholtz free energy, $A(N, V, T) = E(N, V, T) - T S(N, V, T)$, E and S being the internal energy and entropy, respectively. However, appropriate generalizations of nonequilibrium methods to other classical ensembles, as well as to quantum systems, are readily available.

2.   Equilibrium Free-Energy Simulations

The calculation of thermodynamic quantities by means of atomistic simulation is rooted in the framework of equilibrium statistical mechanics [5], which provides the link between the microscopic details of a system and its macroscopic thermodynamic properties. Let us consider a system consisting

of N classical particles with masses m_i. A microscopic configuration of the system is fully specified by the set of N particle momenta $\{\mathbf{p}_i\}$ and positions $\{\mathbf{r}_i\}$, and its energy is described in terms of a potential-energy function $U(\{\mathbf{r}_i\})$. Statistical mechanics in the canonical ensemble then tells us that the distribution of the particle positions and momenta is given by

$$\rho(\mathbf{\Gamma}) = \frac{1}{Z(N, V, T)} \exp(-\beta H(\mathbf{\Gamma})), \tag{1}$$

where $\mathbf{\Gamma} \equiv (\{\mathbf{p}\}, \{\mathbf{r}\})$ denotes a microstate of the system, $\beta = 1/k_B T$ (with k_B Boltzmann's constant) and $H(\mathbf{\Gamma})$ is the classical Hamiltonian. The denominator in Eq. (1) is referred to as the canonical *partition function*, defined as

$$Z(N, V, T) = \int d\mathbf{\Gamma} \exp[-\beta H(\mathbf{\Gamma})], \tag{2}$$

and guarantees proper normalization of the distribution function. The *mechanical* thermodynamic properties such as the internal energy, enthalpy and pressure, can be expressed as *ensemble averages* over the distribution function $\rho(\mathbf{\Gamma})$. Here, the attribute "mechanical" means that the quantity of interest, X, is associated with a specific function $X = X(\mathbf{\Gamma})$ of the microstate, $\mathbf{\Gamma}$, of the system and can be written as

$$\langle X \rangle = \int d\mathbf{\Gamma} \rho(\mathbf{\Gamma}) X(\mathbf{\Gamma}). \tag{3}$$

Standard atomistic simulation techniques such as Metropolis MC [1] and MD [2, 3] provide powerful algorithms for generating sequences of microstates $(\mathbf{\Gamma}_1, \mathbf{\Gamma}_2, \ldots, \mathbf{\Gamma}_M)$ that are distributed according the particular statistical–mechanical (e.g., canonical) distribution function of interest. In this manner, the average implied by Eq. (3) is easily estimated by averaging the function $X(\mathbf{\Gamma})$ over a sequence, $\mathbf{\Gamma}_j$, of microstates generated using MC or MD simulation,

$$\langle X \rangle = \lim_{M \to \infty} \frac{1}{M} \sum_{j=1}^{M} X(\mathbf{\Gamma}_j). \tag{4}$$

Although the partition function Z, itself, is not known this does not present a problem in the case one is interested in any of the mechanical properties of the system; since Z is *implicit* in the generation of the sequence of microstates, $\mathbf{\Gamma}_i$, it is not needed to perform the ensemble average of Eq. (3).

The calculation of thermal quantities is not so straightforward, however. For example, the Helmholtz free energy

$$A(N, V, T) = -\frac{1}{\beta} \ln Z(N, V, T) = -\frac{1}{\beta} \ln \left(\int d\mathbf{\Gamma} \exp[-\beta H(\mathbf{\Gamma})] \right), \tag{5}$$

is seen to be an *explicit* function of the partition function Z rather than an average of the type shown in Eq. 3. Therefore, as Z is not directly accessible in an MC or MD simulation, indirect strategies must be used.

The most widely adopted strategy is to construct a real or artificial *thermodynamic path* that consists of a continuous sequence of equilibrium states linking two states of interest of the system and then attempt to calculate the free-energy *difference* between them. Should the free energy of one of these states be exactly known, the free energy of the other may then be put on an absolute basis. This approach provides the basis for the common *thermodynamic integration* (TI) method. Usually TI relies on the definition of a thermodynamic path in the space of system Hamiltonians. Typically, this involves the construction of an "artificial" Hamiltonian $H(\Gamma, \lambda)$, which, aside from the usual dependence on the microstate Γ is also a function of some generalized coordinate or *switching parameter* λ. This generalized Hamiltonian is then constructed in such a way that it leads to a continuous transformation from the Hamiltonian of a system of interest to that of a reference system of which the free energy is known beforehand. Within the canonical ensemble, the Helmholtz free-energy difference between the initial and final states of the path, characterized by the switching coordinate values λ_1 and λ_2, respectively, is then given by

$$\Delta A \equiv A(\lambda_2; N, V, T) - A(\lambda_1; N, V, T)$$
$$= \int_{\lambda_1}^{\lambda_2} d\lambda' \left(\frac{\partial A(\lambda; N, V, T)}{\partial \lambda} \right)_{\lambda'} = \int_{\lambda_1}^{\lambda_2} d\lambda' \left\langle \frac{\partial H(\Gamma, \lambda)}{\partial \lambda} \right\rangle_{\lambda'} \equiv W_{\text{rev}}, \qquad (6)$$

where $A(\lambda; N, V, T)$ is the Helmholtz free energy of the system as a function of the switching coordinate λ for fixed N, V, and T, and the brackets in the second integral denote an average evaluated for the canonical ensemble associated with the generalized coordinate value $\lambda = \lambda'$.

From a thermodynamic standpoint, Eq. (6) may be interpreted in the following way. The free-energy difference between the initial and final states is equal to the *reversible work* W_{rev} done by the generalized thermodynamic driving force $\partial H(\Gamma, \lambda)/\partial \lambda$ along a *quasistatic*, or *reversible* process connecting both states. By quasistatic we mean that the process is carried out so slowly that the system remains in equilibrium at all times and the instantaneous driving force is equal to the associated equilibrium ensemble average. In this way, the TI method represents a numerical discretization of the quasistatic process; W_{rev} is estimated by computing the equilibrium ensemble averages of the driving force on a grid of λ-values on the interval $[\lambda_1, \lambda_2]$, after which the integration is carried out using standard numerical techniques. For further details of the TI method and its applications we refer to the chapter by Kofke and Frenkel [4].

3.   Nonequilibrium Free-Energy Estimation

3.1.   Establishing Free-Energy Bounds: Systematic and Statistical Errors

Nonequilibrium free-energy estimation is an alternative approach to measuring the reversible work W_{rev}. Instead of discretizing the quasistatic process in terms of a sequence of independent equilibrium states, the reversible work is estimated by means of a single, dynamical sequence of *nonequilibrium* states, explored along an out-of-equilibrium simulation. This is achieved by introducing an explicit "time-dependent" element into the originally static sequence of states by making $\lambda = \lambda(t)$ an explicit function of the simulation "time" t. Here we have used the quotes to emphasize that t should not always be interpreted as a real physical time. For instance, in contrast to MD simulations, typical displacement MC simulations do not involve a natural time scale, in case of which t is simply an index variable that orders the sequence of sampling operations, measured in simulation steps.

Suppose we choose $\lambda(t)$ such that $\lambda(0) = \lambda_1$ and $\lambda(t_{sim}) = \lambda_2$, so that λ varies between λ_1 and λ_2 in a time t_{sim}. Accordingly, the Hamiltonian $H(\Gamma, \lambda) = H(\Gamma, \lambda(t))$ also becomes a function of t, and is driven from the initial system H_1 to the final system H_2 in the same time. The *irreversible* work W_{irr} done by the driving force along this switching process, defined as

$$W_{irr} = \int_0^{t_{sim}} dt' \left(\frac{d\lambda}{dt}\right)_{t'} \left(\frac{\partial H}{\partial \lambda}\right)_{\lambda(t')},\tag{7}$$

provides an estimator for the *reversible* work W_{rev} done along the corresponding quasistatic process.

The point of this nonequilibrium procedure is that values of W_{irr} can be found, in principle, from a single simulation, because the integration in Eq. (7) involves *instantaneous* values of the function $\partial H / \partial \lambda$ rather than ensemble averages. If efficient, this would be much less costly than the TI procedure in Eq. (6), which requires a series of independent equilibrium simulations. But there is, of course, a trade-off. While the TI method is inherently "exact" in that the errors are associated only with statistical sampling and the discreteness of the mesh used for the numerical integration, the irreversible work procedure provides a *biased* estimator for W_{rev}. That is, aside from statistical errors arising from different choices of initial configurations for calculation of Eq. (7), the irreversible estimator W_{irr} is subject to a systematic error $\Delta\mathcal{E}_{syst}$. Both types of error are due to the inherently irreversible nature of the nonequilibrium process.

The statistical errors originate from the fact that, for a fixed and finite simulation time t_{sim}, the value of the integral in Eq. (7) depends on the initial

conditions of the nonequilibrium process. In other words, for different initial conditions, $\mathbf{\Gamma}_j(t=0)$, and a finite simulation time t_{sim}, the value of W_{irr} in Eq. (7) is not unique. Instead, it is a stochastic quantity characterized by a distribution function with a finite variance, giving rise to statistical errors of the sort arising in any MC or MD simulation. The systematic error manifests itself in terms of a shift of the mean of the irreversible work distribution with respect to the value of the ideal quasistatic work W_{rev}. This shift is caused by the dissipative entropy production characteristic of irreversible processes [6]. Because the entropy always increases, the systematic error $\Delta\mathcal{E}_{\text{diss}}$ is always positive, regardless of the sign of the reversible work W_{rev}. In this way, the average value $\langle W_{\text{irr}}\rangle$ of many measurements of the irreversible work will yield an upper bound to the reversible work W_{rev}, provided the average is taken over an ensemble of *equilibrated* initial conditions $\mathbf{\Gamma}_j(t=0)$ at the starting point, $t=0$. The importance of satisfying the latter condition was demonstrated by Hunter *et al.* [7]. From a purely thermodynamic point of view, the bounding error is simply a consequence of the Helmholtz *inequality*. Starting from an equilibrium initial state, for instance at $\lambda=\lambda_1$, the irreversible work upon driving the system to $\lambda=\lambda_2$ is always an upper bound to the actual free-energy change between the equilibrium states of initial and final systems, i.e.,

$$\langle W_{\text{irr}}\rangle \geq \Delta A = A(\lambda_2; N, V, T) - A(\lambda_1; N, V, T). \tag{8}$$

Only in the limit of an ideally quasistatic, or reversible process, represented by the $t_{\text{sim}} \to \infty$ limit, does the inequality in Eq. (8) become the *equality*, $W_{\text{rev}} = \Delta A$, as also manifested in Eq. (6). The preceding ideas are illustrated conceptually in Fig. 1(a) and (b), which show typical distribution functions of irreversible work measurements starting from an ensemble of equilibrated initial conditions. Figure 1(a) compares the results that might be obtained for irreversible work measurements for two different finite simulation times $t_{\text{sim}} = t_1$ and $t_{\text{sim}} = t_2$, with $t_2 > t_1$ to the ideally reversible $t_{\text{sim}} \to \infty$ limit. Both finite-time results show distribution functions with a finite variance and whose mean values have been shifted with respect to the reversible work value by a positive systematic error. Both the variance and systematic error for $t_{\text{sim}} = t_1$ are larger than the corresponding values for $t_{\text{sim}} = t_2$, given that the latter process proceeds in a slower manner, leading to smaller irreversibility. Figure 1(b) shows the irreversible work estimators obtained for the reversible work associated with a quasistatic process in which system 1 is transformed into system 2 as obtained in the forward $(1 \to 2)$ and backward $(2 \to 1)$ directions using the same simulation time t_{sim}. Given that the systematic error is always positive, the forward and backward processes provide upper and lower bounds to the reversible work value, respectively. However, in general, the systematic and statistical errors need not be equal for both directions.
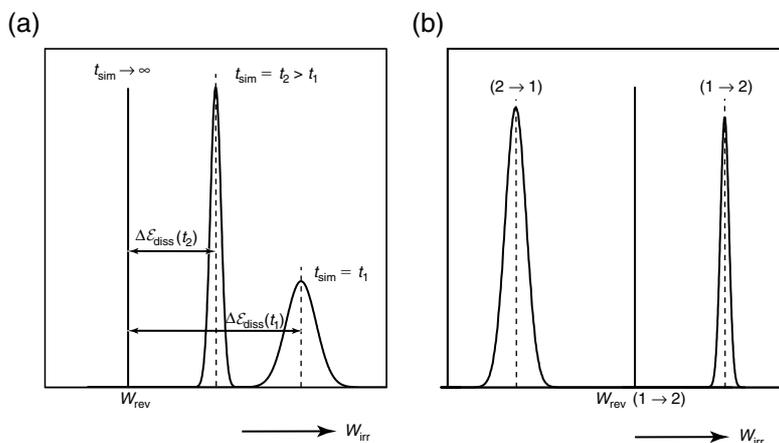
Figure 1. Conceptual illustration of typical irreversible work distributions obtained from nonequilibrium simulations. (a) compares the results that might be obtained for irreversible work measurements for two different finite simulation times $t_{\text{sim}} = t_1$ and $t_{\text{sim}} = t_2$, with $t_2 > t_1$ to the ideally reversible $t_{\text{sim}} \to \infty$ limit. (b) shows the irreversible work estimators obtained for the reversible work associated with a quasistatic process in which system 1 is transformed into system 2 as obtained in the forward $(1 \to 2)$ and backward $(2 \to 1)$ directions using the same simulation time t_{sim}.

3.2. Optimizing Free-Energy Bounds: Insight from Nonequilibrium Statistical Mechanics

A natural question that arises after considering the discussion in previous section is how one might tune the nonequilibrium process so as to minimize the systematic and statistical errors associated with the irreversibility for given initial and final equilibrium states and a given simulation time t_{sim}. To answer this question, it is useful to investigate the microscopic origin of entropy production in nonequilibrium processes. For this purpose, it is particularly helpful to consider the particular class of *close-to-equilibrium* nonequilibrium processes for which the instantaneous distribution functions of nonequilibrium states do not deviate too much from the ideally quasistatic equilibrium distribution functions and where theory of linear response [5] is appropriate. As we will see later on, it is not too difficult to reach this condition in practical situations. As described by Onsager's regression hypothesis [5], when a nonequilibrium state is not too far from equilibrium, the relaxation of any mechanical property can be described in terms of the proper *equilibrium* autocorrelation function. In other words, the hypothesis states that the relaxation of a nonequilibrium disturbance is governed by the same laws as the regression of spontaneous microscopic fluctuations in an equilibrium system.

Under the assumption of proximity to equilibrium, one can then derive the following expression for the mean dissipated energy, i.e., the systematic error $\Delta \mathcal{E}_{\mathrm{diss}}(t_{\mathrm{sim}})$, for a series a irreversible work measurements obtained from nonequilibrium simulations of duration t_{sim} [8–10]:

$$\Delta \mathcal{E}_{\mathrm{diss}}(t_{\mathrm{sim}}) = \frac{1}{k_B T} \int_0^{t_{\mathrm{sim}}} \mathrm{d}t' \, \left(\frac{\mathrm{d}\lambda}{\mathrm{d}t}\right)_{t'}^2 \, \tau[\lambda(t')] \, \mathrm{var}\left(\frac{\partial H}{\partial \lambda}\right)_{\lambda(t')}. \tag{9}$$

Aside from the switching rate, the integrand in Eq. (9) contains both the *correlation time* as well as the *equilibrium variance* of the driving force $\partial H/\partial \lambda$. These two factors describe, respectively, how quickly the fluctuations in the driving force decay and how large these fluctuations are in the equilibrium state. It is clear that the integral is positive-definite, as it must be. Moreover, it indicates that, for near-equilibrium processes, the systematic error should be the same for forward and backward processes. This means that, in the linear–response regime, one can obtain an *unbaised* estimator for the reversible work W_{rev} by combining the results obtained from forward and backward processes. More specifically, in this regime we have

$$\langle W_{\mathrm{irr}}(1 \rightarrow 2) \rangle = W_{\mathrm{rev}}(1 \rightarrow 2) + \Delta \mathcal{E}_{\mathrm{diss}}, \tag{10}$$

and

$$\langle W_{\mathrm{irr}}(2 \rightarrow 1) \rangle = -W_{\mathrm{rev}}(1 \rightarrow 2) + \Delta \mathcal{E}_{\mathrm{diss}}, \tag{11}$$

leading to the unbaised estimator (i.e., subject to statistical fluctuations only)

$$W_{\mathrm{rev}}(1 \rightarrow 2) = \tfrac{1}{2} \left(\langle W_{\mathrm{irr}}(1 \rightarrow 2) \rangle - \langle W_{\mathrm{irr}}(2 \rightarrow 1) \rangle \right). \tag{12}$$

Concerning minimization of dissipation, Eq. (9) tells us that one should attempt to reduce both the magnitude of the fluctuations in the driving force as well as the associated correlation times. This involves both a *static* component, i.e., the magnitude of the equilibrium fluctuations, and a *dynamic* one, namely the typical decay time of equilibrium correlations. This shows that not only the choice of the path, $H(\lambda)$, but also the simulation algorithm by which the system is propagated in "time" (i.e., MC or MD simulation) will affect the dissipation in the irreversible work measurements. Whereas the magnitude of the equilibrium fluctuations should be algorithm independent (as long as the algorithms sample the same equilibrium distribution function), the correlation time is certainly algorithm-dependent. In case of displacement MC simulation, as we will see below, the choice of the maximum displacement parameter affects the correlation time τ, and, consequently, the magnitude of the dissipation.

Finally, let us now assume that we have a prescribed path $H(\lambda)$ and a simulation algorithm to sample the nonequilibrium process between the systems $H(\lambda_1)$ and $H(\lambda_2)$. How do we now choose the functional form of the time-dependent switching function $\lambda(t)$ to minimize the dissipation? Equation (9) provides us with an explicit answer. To see this, we first perform a change of integration variable, setting $x' = t'/t_{\text{sim}}$, obtaining

$$\Delta \mathcal{E}_{\text{diss}}(t_{\text{sim}}) = \frac{1}{t_{\text{sim}}} \Delta \mathcal{E}_{\text{diss}}[\lambda(x)], \tag{13}$$

with

$$\Delta \mathcal{E}_{\text{diss}}[\lambda(x)] = \frac{1}{k_B T} \int\limits_0^1 dx' \left(\frac{d\lambda}{dx}\right)_{x'}^2 \tau(\lambda(x')) \, \text{var}\left(\frac{\partial H}{\partial \lambda}\right)_{\lambda(x')}. \tag{14}$$

Equation (14) is a *functional* of the common form [11]

$$S[\lambda(x)] = \int\limits_0^1 dx \, F(\lambda'(x), \lambda(x), x). \tag{15}$$

The minimization of the dissipation is thus equivalent to finding the function $\lambda(x)$ that minimizes a functional of the type (15) subject to the boundary conditions $\lambda(0)=\lambda_1$ and $\lambda(1)=\lambda_2$. Standard variational calculus then shows that the solution is obtained by solving the Euler–Lagrange equation [11] associated with the functional,

$$\frac{d}{dx} \frac{\partial F}{\partial \lambda'} = \frac{\partial F}{\partial \lambda}, \tag{16}$$

subject to the mentioned boundary conditions.

4. Applications of Nonequilibrium Free-Energy Estimation

To illustrate the discussion of the previous sections we will now discuss a number of applications of nonequilibrium free-energy estimation, demonstrating the bounding properties of irreversible-work measurements, as well as aspects of dissipation optimization.

4.1. Harmonic Oscillators

In the first application we consider the problem of computing the free-energy difference between two systems consisting of 100 identical, independent,

one-dimensional harmonic oscillators of unit mass with different characteristic frequencies [9]. In particular we will consider the path defined by

$$H(\lambda) = \frac{1}{2} \sum_{i=1}^{100} [(1-\lambda)\omega_1^2 + \lambda\omega_2^2] x_i^2, \tag{17}$$

with $\omega_1 = 4$ and $\omega_2 = 0.5$ at a temperature $k_B T = 2$. Note that we are considering only the potential energy of the oscillators here and have neglected any kinetic energy contributions. We can do this because the free-energy difference between two harmonic oscillators at a fixed temperature is determined only by the configurational part of the partition function. The value of the desired reversible work W_{rev} per oscillator associated with a quasistatic modification of the frequency from ω_1 to ω_2 is known analytically:

$$W_{\text{rev}}(\omega_1 \rightarrow \omega_2) = -k_B T \ln \frac{\omega_1}{\omega_2} = -4.15888. \tag{18}$$

The simulation algorithm we utilize is standard Metropolis displacement MC with a fixed maximum trial displacement $\Delta x_{\text{max}} = 0.3$. First we consider the statistics of the irreversible work measurements as a function of the simulation "time" t_{sim}, which here stands for the number of MC sweeps (one sweep corresponds to one trial displacement per oscillator) per process, for a *linear* switching function. The results are shown as the dashed line curves in Fig. 2(a) and (b), in which each data point represents the mean value of W_{irr} over 50 independent initial conditions. Figure 2(a) shows that the upper and lower
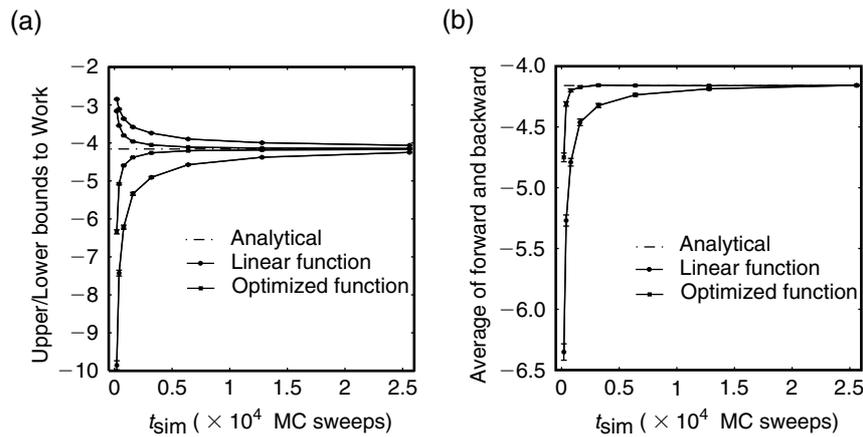
Figure 2. Results of irreversible-work measurements per oscillator as a function of the switching time t_{sim} for the linear (dashed lines) and optimal (solid lines) switching function. The analytical reversible work value is also shown (dot dashed line). (a) shows the results of the forward (upperbounds) and backward (lowerbounds) directions. (b) shows the values of the combined estimator of Eq. (12).

limit do converge toward the reversible value W_{rev}, although they do so quite slowly. The slow convergence becomes more apparent when we consider the behavior of the combined estimator of Eq. (12) in Fig. 2(b). If the process were sufficiently slow for linear–response theory to be accurate, the combined estimator should be unbiased and show no systematic deviation. It is clear that this is only the case for the slowest process, at $t_{\text{sim}} = 2.56 \times 10^4$ MC sweeps. All shorter simulations show a systematic deviation, indicating that the associated processes remain quite far from equilibrium, hampering convergence.

Next, we attempt to minimize dissipation in the simulation by using the switching function $\lambda(x)$ that satisfies the Euler–Lagrange Eq. (16). For this purpose we first measured the equilibrium variance in the driving force and the characteristic correlation time of decay as a function of λ from a series of equilibrium simulations (i.e., fixed λ), after which we numerically solved Eq. (16), subject to the boundary conditions $\lambda(0) = 0$ and $\lambda(1) = 1$. The equilibrium variances, correlation times and the resulting optimal switching function are shown in Fig. 3(a)–(c), respectively. The results in Fig. 3(a) and (b) indicate that the main contribution to the dissipation originates from the region $\lambda \approx 1$, where both the magnitude as well the characteristic decay time of the fluctuations in the driving force increase sharply. The optimal switching function in Fig. 3(c) captures this effect, prescribing a slow switching rate where one should and going faster where one can. The results obtained with this function for the irreversible work measurements are shown as the red lines in Fig. 2(a) and (b). The improvement compared to the linear switching function is quite significant. Figure 2(b), for instance, shows that for t_{sim} as short as 3.2×10^3 MC sweeps, the nonequilibrium process has already reached the linear–response regime.

The above optimization procedure is useful in cases where the thermodynamic path $H(\lambda)$ is prescribed beforehand. This is the case, for instance, for
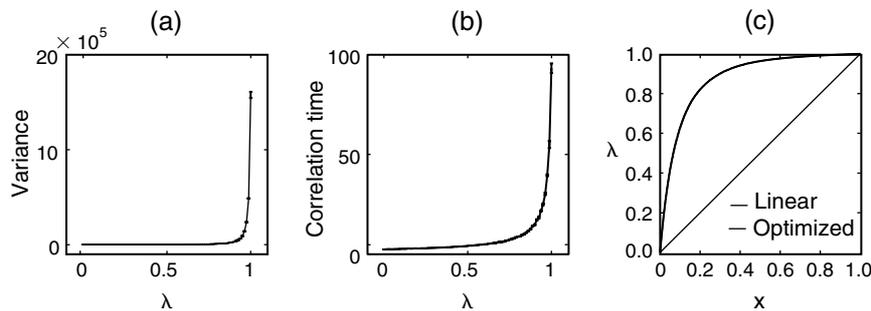
Figure 3. (a) The equilibrium variance ($\partial H / \partial \lambda$), and (b) the correlation decay time (in MC sweeps) as a function of λ. (c) shows the optimal switching function, as determined by numerically solving Euler–Lagrange equation (16).

the reversible-scaling method [12], in which each state along the fixed path $H(\lambda) = \lambda V$ (V is the interatomic interaction potential) represents the physical system of interest in a different temperature state. In this manner, a single irreversible-work simulation along the scaling path provides a continuous series of estimators of the system's free energy on a finite temperature interval. If one has some information about the behavior of the magnitude of the and correlation-decay times of the fluctuations of the driving force, one may use the variational method described above to optimize the switching function and minimize dissipation effects.

4.2. Compression of Confined Lennard–Jones Particles

In the following application we consider a system consisting of 30 Lennard–Jones particles, constrained to move on the x-axis only. In addition, the particles are subject to an external field whose strength is controlled by an external parameter L. More specifically, we consider the path

$$H(L) = \epsilon \left[\left(\frac{\sigma}{x_{ij}} \right)^{12} - \left(\frac{\sigma}{x_{ij}} \right)^{6} \right] + \left(\frac{2x_i}{L} \right)^{26}, \tag{19}$$

where x_i describes the position of particle i on the x-axis and $x_{ij} \equiv |x_i - x_j|$ is the distance between particles i and j. The second term in Eq. (19) is the external field, which is a very steeply rising potential and has the effect of confining the particles through very strong interactions with the first and last particles, effectively causing the 30 particles to lie approximately evenly spaced between $x = \pm L/2$. Now consider the compression process wherein L changes from $L_0 = 30\sigma$ to $L_1 = 26\sigma$, forcing the line of particles to undergo a one-dimensional compression. As in the previous example, we will attempt to compute the reversible work associated with this process by measuring the irreversible work $\langle W_{\mathrm{irr}} \rangle$ for both process directions. Once again we utilize the Metropolis MC algorithm, but instead of fixing the algorithm parameter $\Delta x_{\max}$, describing the maximum trial displacement, we now consider the effects of changing the sampling algorithm on the convergence of the upper and lower bounds. Although the variance of the driving force $\mathrm{var}\,(\partial H/\partial \lambda)$ will not be affected, the correlation time will certainly depend on the choice of $\Delta x_{\max}$. This is illustrated in Fig. 4, which shows the convergence of the upper and lower bounds to the reversible work as obtained for 3 different values of $\Delta_{\max}$ at a temperature $k_B T = 0.35\epsilon$: $\Delta x_{\max} = 0.6\sigma$, 0.1σ, and 0.04σ, respectively. Effectively, the variation of this algorithm parameter may be thought of as changing the strength of the coupling between the MC "thermostat" and the system of particles. We utilized the linear switching function which varies L linearly between L_0 and L_1 in t_{sim} MC sweeps (each sweep
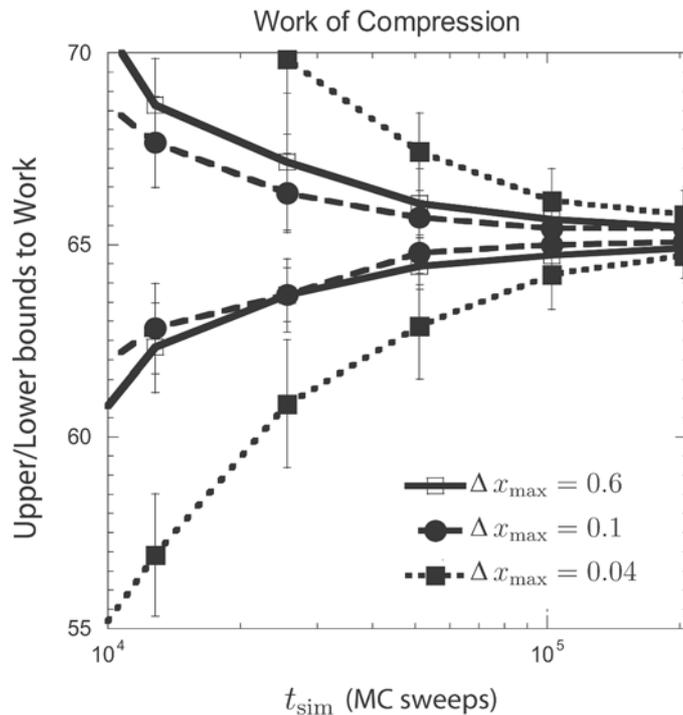
Figure 4. Results of forward (upperbound) and backward (lowerbound) irreversible-work measurements (in units of ϵ) as a function of the switching time t_{sim} for the linear switching function for three different values of the MC algorithm parameter Δx_{max}.

consisting of 30 MC single-particle trial moves). Each data point and corresponding error bar (± 1 standard deviation) were obtained from a set of 21 irreversible work measurements initiated for independent, equilibrated initial conditions. It is also useful to note that it is not necessary to explicitly compute the work W_{irr} by using (7). All that is needed, through the first law of thermodynamics which applies equally to reversible and irreversible processes, is to calculate the work as $W_{\text{irr}} = \Delta E - Q$, where ΔE is the difference in internal energies of the system between the first and last switching steps, and Q is the heat accumulated during the switching process. This heat, Q, is simply the sum of energies added to, or subtracted from, the system as MC configurations evolve during a simulation. Given that these energies, $\Delta \varepsilon_i$, are already calculated in determining whether moves for particle i are to be accepted or rejected according to the canonical $\exp(-\Delta \varepsilon_i / k_B T)$, no extra programming is needed to calculate W_{irr}.

It is immediately seen that the strength of the system-thermostat coupling through the algorithm parameter Δ_{max} is indeed a variational parameter

for the free-energy computations. Accordingly, rather than selecting a pre-set acceptance ratio of trial moves, as is usually done in equilibrium MC simulations, Δx_{max} should be determined so as to minimize the difference between the upper and lower bounds to ΔA. The results show that for all three values of $\Delta x_{\max}$, the upper and lower bounds show convergence. Yet, the convergence properties are clearly different for the three parameter values, giving the best results for $\Delta x_{\max} = 0.1$ and the worst for $\Delta x_{\max} = 0.04$, indicating that the correlation decay time for the fluctuations in the driving force are the shortest for the former and the longest for the latter.

Nevertheless, the convergence of the bounds is still quite slow, in that hundreds of thousands of MC sweeps are required to obtain convergence of to within a few percent. This is a consequence of the strong interactions between the particles, as their hard cores interact during the compression from the "ends" of the line of particles and such hard core density gradients are typically slow to work themselves out through single particle MC moves. Contrary to the simple harmonic oscillator problem discussed in the previous section, this problem will be ubiquitous in most atomic and molecular systems in the condensed phase, seemingly rendering the free-energy computations on realistic systems of interest problematic.

The questions that now arise are as to whether we can estimate the systematic errors $\Delta \mathcal{E}_{\text{diss}}$ from data already in hand and use it to improve the estimates of Fig. 4; and/or if we can optimize the thermodynamic path to reduce dissipation and achieve better behavior at short switching times; or perhaps both?

4.3. Estimating Equilibrium Work from Nonequilibrium Data

Recently, Jarzynski [13] has generalized the Gibbs–Feynman identity,

$$\Delta A = A_1 - A_0 = -k_B T \, \ln\langle \exp[-(H_1 - H_0)/k_B T]\rangle_0 \qquad (20)$$

where $\langle \cdots \rangle_0$ denotes canonical averaging with respect to configurations generated by H_0, and which is the basis of thermodynamic perturbation theory [4], to finite-time processes. Equation (20) is an identity, but in practice it is useful only when the configurations generated by canonical sampling with respect to H_0 strongly overlap those generated by H_1. For hard core fluids this would be unusual unless H_1 and H_0 are quite "close", resulting in the perturbative use of Eq. (20). Jarzynski now allows H_0 to dynamically approach H_1 along a path, in analogy with the above discussions. The result, in the context discussed here, suggests that for a given set of N irreversible-work measurements $W_i \equiv W_{\text{irr}}(\Gamma_i, t = 0)$, with $i = 1, \ldots, N$, instead of estimating $\langle W_{\text{irr}} \rangle$ as the sim-

ple arithmetic mean of the W_i, one should calculate the Boltzmann weighted "Jarzynski" (or "Jz") average

$$\langle W \rangle_{\mathrm{Jz}} = \frac{1}{M} \sum_{i=1}^{M} \exp(-W_{\mathrm{i}}/k_B T), \qquad (21)$$

and then estimate the free energy change as

$$\Delta A_{\mathrm{Jz}} \equiv -k_B T \ln \langle W \rangle_{\mathrm{Jz}} . \qquad (22)$$

In this way bounding is sacrificed, but a more accurate result is not precluded given that, in principle, the Jz-average is *unbiased*. This approach has been shown to be effective both in the analysis of simulation data as well as finite-time polymer extension experiments, which are of course irreversible. An immediate concern, however, is that, although in the limit of complete sampling as in the Gibbs–Feynman identity, the Jarzynski results are exact in the context of a dissipation-free system, incomplete MC sampling may result in unsatisfactory results.
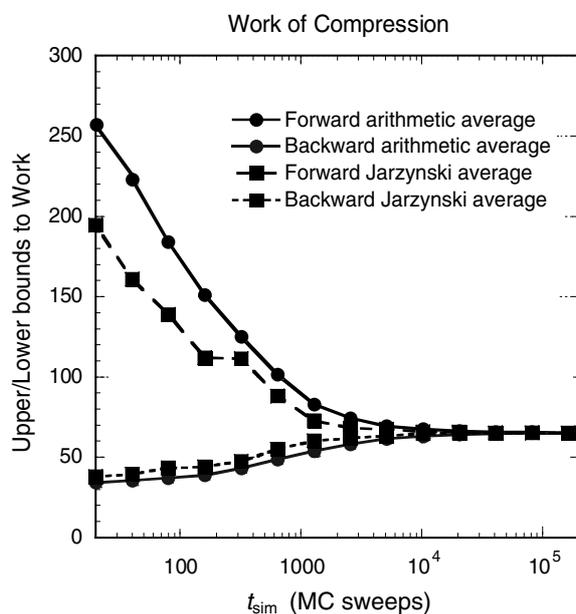
Figure 5. Results of forward and backward irreversible-work averages (in units of ϵ) for the 30-particle confined Lennard–Jones system as a function of the switching time t_{sim}. The results show both the simple arithmetic averages as well as the Boltzmann-weighted Jarzynski averages.

This is illustrated in Fig. 5, where data used to generate the bounds to ΔA in Fig. 4, are plotted over a much larger range of switching times t_{sim}, and compared to the ΔA_{Jz} estimates. Both the simple arithmetic as well as the Jarzynski averages for both directions were computed over the 21 independent initial conditions. It is evident that, although not giving bounds, the ΔA_{Jz} estimates indeed improve the upper and lower bounds compared to those calculated as simple averages. However, the Jarzynski averages become useful when the convergence of the simple arithmetic averages has reached the order of less than 1 $k_B T$ *per particle*. In this fashion, although a promising computational asset, the Jarzynski procedure still requires systematic procedures for finding more reversible paths.

4.4.   Path Optimization through Scaling and Shifting of Coordinates

As we have seen in the harmonic oscillator and Lennard–Jones problems, the choice of the thermodynamic path and the used switching function is quite crucial to the success of nonequilibrium free-energy estimation. In the case of the harmonic oscillator problem it was relatively straightforward to find a good switching function by explicitly solving the variational problem in Eqs. (15) and (16), which lead to an optimized simulation that "spends the right amount of time along each segment" of the already defined path. Here it is important to note that this variational optimization should be carried out over an *ensemble* averaged W_{irr}, being identical for every member of the ensemble, independently of any specific $\Gamma_i (t = 0)$. This is the reason why early attempts by Pearlman and Kollman [14] to determine paths "on the fly" by looking ahead and avoiding strong dissipative collisions in *specific* configurations may result in the unintentional introduction of a Maxwell demon [15], violating the second law of thermodynamics, which is of course the fundamental origin of the Helmholtz inequality.

Compared to the simple harmonic oscillator problem, the optimization of the nonequilibrium simulation of the confined Lennard–Jones system is significantly more challenging because of the strong interactions between the particles as during the compression of the system. Given that this type of interaction is expected to occur in most interesting problems, it is of interest to design thermodynamic paths that are different from the ones in which one simply follows $H(\lambda)$ as λ runs from an initial to a final value, like we did in the case of the harmonic oscillator problem.

We now present two approaches that follow this idea and lead to thermodynamic paths that are significantly more reversible. Both the coordinate scaling [16] and coordinate shifting methods discussed below derive from

the same fundamental thought: is there a (λ-dependent) coordinate system in which all particles are apparently at rest with relative to one another during the switching process? In such a coordinate system perhaps all particles will have little difficulty in remaining close to equilibrium during the whole switching process, with only the magnitude of their local fluctuations changing.

4.4.1. *Coordinate scaling*

Figure 6 illustrates the possibilities of such an approach, when applied to the simple problem of compression discussed above. Here, in an admittedly simple example, all particles should be compressed "uniformly," rather than by the nonuniform compression generated through the interactions of the confining potential with the particles at both ends of the line. This is accomplished by writing the coordinates as $s(\lambda) x_i$, where $s(\lambda)$ is a (common) scaling parameter, which may then be variationally optimized. The greatly improved bounds of Fig. 6 indicate that a better path has indeed been found. How does this fit the "at rest" criterion mentioned earlier? If one watches the MC dynamics in the unscaled "x_i" coordinates using an optimized $s(\lambda)$, rather than in the actual physical coordinates, $s(\lambda) x_i$, it appears that the equilibrium positions $\langle x_i \rangle$ do not change during the switching, and thus, indeed, the only irreversibility arises from the changes in the RMS fluctuations about the equilibrium positions. It should be noted, however, that, as these scalings may be regarded as a change in the metric that affects the length and volumes definitions, one should include a entropic (calculable) correction to obtain the desired free-energy difference.

Recently, there has been a variety of applications of the scaling approach [16–18], including the determination of the absolute free energy of Lennard–Jones clusters and a smooth metric scaling through a first order solid–solid phase transition, fcc to bcc, with no apparent hysteresis with its resulting irreversibility.

4.4.2. *Coordinate shifting*

In the applications of metric scaling, thermodynamic paths are often easily determined when a clear symmetry is present. Another approach, namely *coordinate shifting* is more useful when such symmetries are absent. As an alternative to writing a moving coordinate using the scaling relation $s(\lambda) x_i$, one can take $x_i = x_i^{\text{fluct}} + x_i^{\text{ref}}(\lambda)$. Here each particle moves in a concerted fashion along a λ-dependent reference path, chosen by symmetry, or by methods such as simulated annealing, to avoid strong hard core interactions or other

likely causes of irreversibility. As λ evolves, only the fluctuation coordinates x_i^{fluct} are subject to MC variations: should the physical environment of each particle remain at least roughly constant, one may hope that the fluctuations from the $x_i^{\text{ref}}(\lambda)$ do not depend strongly on λ. To the extent that this is the case, the fluctuation coordinates are always at equilibrium, and thus the path is reversible! Figure 7 illustrates the efficacy of this method for the linear compression problem.

As opposed to coordinate scaling, coordinate shifting does not change the metric, dispensing the need for entropic corrections and paving the way for applications involving inhomogeneous systems where the possible absence of symmetries obscures the choice of an appropriate metric obvious and complicates the computation of scaling entropy corrections. As is also clear from the results shown in Figure 7, the finite-time upper and lower bounds converge sufficiently quickly for the Jarzynski averaging to actually markedly improve even the shortest-time results. More general "non-linear" combinations of scaling and shifting may also be used to advantage, as in [19].
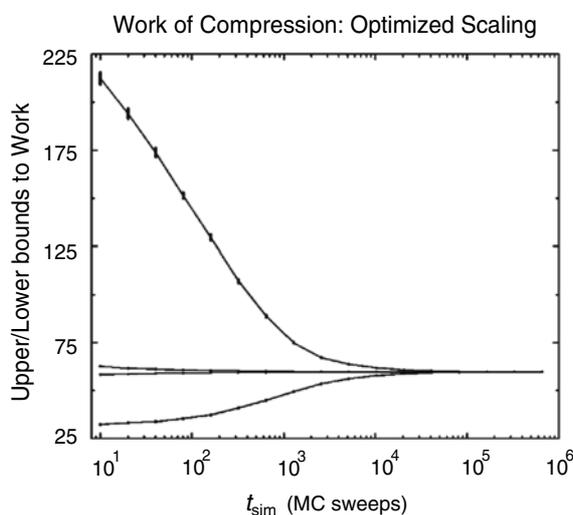
Figure 6. Convergence of upper and lower bounds to the free-energy change associated with the compression of the confined Lennard–Jones system at $k_B T = 0.35$ as a function of the switching time t_{sim}. The outer pair of lines are from standard finite-time switching, whereas the inner pair represents the results from finite-time switching using linear metric scaling. The vertical bars represent the standard error in the mean of 100 replicas.
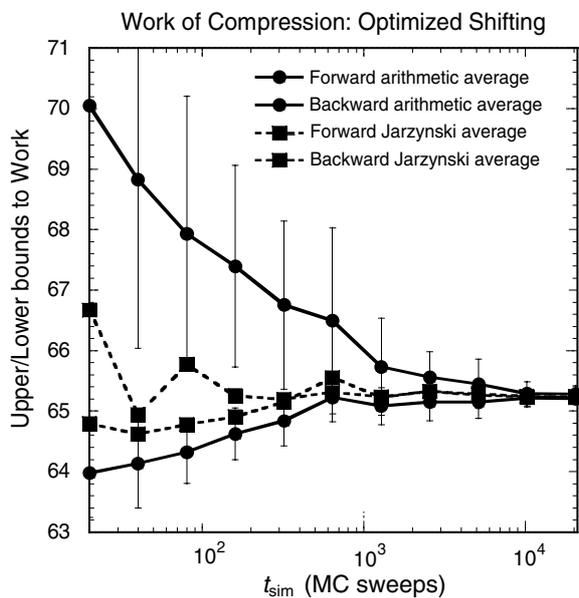
Figure 7. Convergence of upper and lower bounds to the free-energy change associated with the compression of the confined Lennard–Jones system at $k_B T = 0.35$ as a function of the switching time t_{sim} as obtained by optimized coordinate shifting. The vertical bars represent the standard error in the mean of 21 replicas. The results obtained with Jarzynski averages are also shown.

5. Outlook

One of the most fundamental and challenging applications of atomistic simulation techniques concerns the determination of those thermodynamic properties that require determination of the entropy, the chemical potential and the various free energies, which are all examples of thermal thermodynamic properties. In contrast to their mechanical counterparts (e.g., enthalpy, pressure) they cannot be computed as ensemble (or time) averages, and indirect strategies must be adopted.

Here, we have discussed the basic aspects of a particular strategy, that of using nonequilibrium simulations to obtain estimators of reversible work between equilibrium states. The point of this approach is that, in contrast to equilibrium methods such as thermodynamic integration, the desired value can, in principle, be estimated from a single simulation. But there is a trade-off, in that the nonequilibrium estimators are subject to both systematic and statistical errors, caused by the inherently irreversible nature of nonequilibrium processes.

Yet, the approach allows one to systematically obtain upper and lower bounds to the requested reversible result by exploring the nonequilibrium processes both in forward and backward directions. The bounds for a given process become tighter with decreasing process rates. But more importantly, it is possible to optimize the nonequilibrium process so as to minimize irreversibility and, for a given process time, decrease the bounds. We have discussed a number of methods by which to conduct this optimization task, including explicit functional optimization using standard variational calculus and techniques based on special coordinate transformations aimed at the reduction of irreversibility.

These techniques have been quite successful so far, allowing accurate free-energy measurements using relatively short nonequilibrium simulations. In this light, the idea of using nonequilibrium simulations has now grown into a robust and efficient computational approach to the problem of computing thermal thermodynamic properties using atomistic simulation methods. Nevertheless, further development remains necessary, in particular toward improving/generalizing the existing optimization schemes.

References

[1] G. Gilmer and S. Yip, *Handbook of Materials Modeling*, vol. I, chap. 2.14, Kluwer, 2004.
[2] J. Li, *Handbook of Materials Modeling*, vol. I, chap. 2.8, Kluwer, 2004.
[3] M.E. Tuckerman, *Handbook of Materials Modeling*, vol. I, chap. 2.9, Kluwer, 2004.
[4] D.A. Kofke and D. Frenkel, *Handbook of Materials Modeling*, vol. I, chap. 2.14, Kluwer, 2004.
[5] D. Chandler, *Introduction to Modern Statistical Mechanics*, Oxford University Press, Oxford, 1987.
[6] L.D. Landau and E.M. Lifshitz, *Statistical Physics, Part 1*, 3rd edn., Pergamon Press, Oxford, 1980.
[7] J.E. Hunter III, W.P. Reinhardt, and T.F. Davis, "A finite-time variational method for determining optimal paths and obtaining bounds on free energy changes from computer simulations," *J. Chem. Phys.*, 99, 6856, 1993.
[8] L.W. Tsao, S.Y. Sheu, and C.Y. Mou, "Absolute entropy of simple point charge water by adiabatic switching processes," *J. Chem . Phys.*, 101, 2302, 1994.
[9] M. de Koning and A. Antonelli, "Einstein crystal as a reference system in free energy estimation using adiabatic switching," *Phys. Rev. E*, 53, 465, 1996.
[10] M. de Koning and A. Antonelli, "Adiabatic switching applied to realistic crystalline solids: vacancy-formation free energy in copper," *Phys. Rev. B*, 55, 735, 1997.
[11] R. Courant and D. Hilbert, *Methods of Mathematical Physics*, vol. 1, Wiley, New York, 1953.
[12] M. de Koning, A. Antonelli, and S. Yip, "Optimized free energy evaluation using a single reversible-scaling simulation," *Phys. Rev. Lett.*, 83, 3973, 1999.
[13] C. Jarzynski, "Nonequilibrium equality for free energy differences," *Phys. Rev. Lett.*, 78, 2690, 1997.

[14] D.A. Pearlman and P.A. Kollman, "The lag between the Hamiltonian and the system configuration in free energy perturbation calculations," *J. Chem. Phys.*, 91, 7831, 1989.

[15] H.S. Leff and A.F. Rex, *Maxwell's Demon 2, Entropy, Classical and Quantum Information, Computing*, Institute of Physics Publishing, Bristol, U.K, 2002.

[16] M.A. Miller and W.P. Reinhardt, "Efficient free energy calculations by variationally optimized metric scaling: concepts and applications to the volume dependence of cluster free energies and to solid–solid phase transitions," *J. Chem. Phys.*, 113, 7035, 2000.

[17] L.M. Amon and W.P. Reinhardt, "Development of reference states for use in absolute free energy calculations of atomic clusters with application to 55-atom Lennard–Jones clusters in the solid and liquid states," *J. Chem. Phys.*, 113, 3573, 2000.

[18] W.P. Reinhardt, M.A. Miller, and L.M. Amon, "Why is it so difficult to simulate entropies, free energies and their differences?" *Accts. Chem. Res.*, 34, 607, 2001.

[19] C. Jarzynski, "Targeted free energy perturbation," *Phys. Rev. E*, 65, 046122, 2002.