# Bias and error in estimates of equilibrium free-energy differences from nonequilibrium measurements

**Jeff Gore*, Felix Ritort*†, and Carlos Bustamante*‡§¶**

*Department of Physics, University of California, Berkeley, CA 94720; †Department of Physics, University of Barcelona, Diagonal 647, 08028 Barcelona, Spain; ‡Department of Molecular and Cell Biology and Howard Hughes Medical Institute, University of California, Berkeley, CA 94720; and §Physical Biosciences Division, Lawrence Berkeley National Laboratory, Berkeley, CA 94720

Contributed by Carlos Bustamante, August 12, 2003

**In 1997, Jarzynski proved a remarkable equality that allows one to compute the equilibrium free-energy difference $\Delta F$ between two states from the probability distribution of the nonequilibrium work $W$ done on the system to switch between the states, $e^{-\Delta F/kT} = \langle e^{-W/kT}\rangle$, [Jarzynski, C. (1997) *Phys. Rev. Lett.* 87, 2690–2693]. The Jarzynski equality provides a powerful free-energy difference estimator from a set of $N$ irreversible experiments and is closely related to free-energy perturbation, a common computational technique for estimating free-energy differences. Despite the many applications of the Jarzynski estimator, its behavior is only poorly understood. In this article we derive the large $N$ limit for the Jarzynski estimator bias, variance, and mean square error that is correct for arbitrary perturbations. We then analyze the properties of the Jarzynski estimator for all $N$ when the probability distribution of work values is Gaussian, as occurs, for example, in the near-equilibrium regime. This allows us to quantitatively compare it to two other free-energy difference estimators: the mean work estimator and the fluctuation–dissipation theorem estimator. We show that, for near-equilibrium switching, the Jarzynski estimator is always superior to the mean work estimator and is even superior to the fluctuation–dissipation estimator for small $N$. The Jarzynski-estimator bias is shown to be the dominant source of error in many cases. Our expression for the bias is used to develop a bias-corrected Jarzynski free-energy difference estimator in the near-equilibrium regime.**

Accurate measurement and calculation of free-energy differences is central to our understanding of biological, chemical, and physical molecular processes. A common method of estimating the free-energy difference $\Delta F = F_B - F_A$ between two states, A and B, of a classical system in contact with a heat reservoir is to perturb the system to induce a transition between these states. The average work done in such a perturbation satisfies $\langle W\rangle \geq \Delta F$, where equality holds if and only if the perturbation is infinitely slow. An average of the work values obtained by any finite time experiment or simulation therefore overestimates the true $\Delta F$. Until recently, recovering the equilibrium free energy from trajectories arbitrarily far from equilibrium was therefore thought to be impossible.

In 1997, however, Jarzynski (1, 2) proved an identity that relates the probability distribution of nonequilibrium work values with the equilibrium free-energy difference between the two states:

$$e^{-\beta\Delta F} = \langle e^{-\beta W}\rangle \qquad [1]$$

where $\beta = (kT)^{-1}$. On the left-hand side of this equation is an exponential of the equilibrium free-energy difference, and the right-hand side is an exponentially weighted average over an infinite number of nonequilibrium work trajectories, all started from the same initial equilibrium state. Although the second law of thermodynamics requires that the average work over all possible trajectories be greater than the free-energy difference, the work for an individual trajectory will occasionally be less (3). The Jarzynski average heavily weights these rare trajectories at the tail of the work distribution in order to recover $\Delta F$. The Jarzynski equality is closely related to other recent advances in the theory of nonequilibrium statistical mechanics (4–7).

A recent single-molecule manipulation experiment verified the Jarzynski equality (8). A single folded RNA molecule was mechanically unfolded both reversibly and irreversibly, and the Jarzynski equality was used to obtain the free-energy change of the system (as estimated by the reversible work values) from the irreversible trajectories. This experiment was done at constant temperature and pressure, and therefore the Jarzynski equality was cast in terms of the Gibb's free energy $\Delta G$ rather than the Helmholtz free energy $\Delta F$. A two-state model describing and justifying the outcome of these experiments was proposed recently (9). The Jarzynski equality can be applied at any point along the pulling trajectory, making it possible to reconstruct the entire free-energy profile of the molecule along the pulling coordinate (10).

The free-energy perturbation technique uses an identical averaging procedure to estimate free-energy differences but obtains the work values computationally (11–13). If a system is perturbed infinitely quickly from state A with Hamiltonian $H_A$ to state B with Hamiltonian $H_B$ such that its configuration, $\vec{x}$, remains unchanged, then the work that must be performed is $H_B(\vec{x}) - H_A(\vec{x})$. The Jarzynski relation is correct for arbitrary perturbation speeds, and thus we can recover the free-energy difference by averaging these work values over the equilibrium ensemble of configurations of state A (1):

$$e^{-\beta\Delta F} = \langle e^{-\beta(H_B - H_A)}\rangle_A = \sum_{\vec{x}} P_A(\vec{x})\exp\{-\beta[H_B(\vec{x}) - H_A(\vec{x})]\}.$$

This result was known long before the Jarzynski equality was proven, and we can view it as the Jarzynski equality in the limit of infinitely fast switching.

The Jarzynski equality therefore has many applications, but unfortunately the properties of the Jarzynski free-energy estimator

$$\Delta\hat{F}_J(N) = -\beta^{-1}\ln\left[\sum_i^N e^{-\beta W_i}\right]$$

are largely unknown. How good is the Jarzynski estimator? If we sample from a given work distribution $P(W)$ a total of $N$ times (by unfolding an RNA molecule $N$ times, for example), then how close will our final free-energy difference estimate be to the true $\Delta F$? It is well known that the Jarzynski estimator is biased for finite $N$, meaning that the estimator has a systematic error. This bias can be the dominant source of error in estimating the free-energy difference, but the magnitude of the bias is generally unknown.

## The Near-Equilibrium Regime and Free-Energy Difference Estimators

There are three common free-energy difference estimators that can be obtained from a set of $N$ trajectories: the mean work estimator

Abbreviations: FD, fluctuation–dissipation; MSE, mean square error.

See commentary on page 12537.

[¶] To whom correspondence should be addressed. E-mail: carlos@alice.berkeley.edu.

www.pnas.org/cgi/doi/10.1073/pnas.1635159100

$\Delta\hat{F}_{\mathrm{MW}}$, the fluctuation–dissipation (FD) theorem estimator $\Delta\hat{F}_{\mathrm{FD}}$, and the Jarzynski estimator $\Delta\hat{F}_{\mathrm{J}}$

$$\begin{aligned}
\Delta\hat{F}_{\mathrm{MW}} &= \langle W\rangle_N = \frac{1}{N}\sum_i^N W_i \\
\Delta\hat{F}_{\mathrm{FD}} &= \langle W\rangle_N - \frac{1}{2}\beta\hat{\sigma}_W^2 = \frac{1}{N}\sum_i^N W_i - \frac{\beta}{2}\frac{1}{N-1}\sum_i^N (W_i - \langle W\rangle_N)^2 \\
\Delta\hat{F}_{\mathrm{J}} &= -\frac{1}{\beta}\ln\langle e^{-\beta W}\rangle_N = -\frac{1}{\beta}\ln\left[\frac{1}{N}\sum_i^N e^{-\beta W_i}\right]
\end{aligned} \qquad \textbf{[2]}$$

$\langle\ldots\rangle_N$ in these equations denotes an experimental average over $N$ trajectories, and $\Delta\hat{F}$ represents an estimator of $\Delta F$. The connection between these three estimators can be elucidated by writing Jarzynski's equality as a sum over the cumulants of the original work distribution (14):

$$\Delta F = \sum_{n=1}^{\infty}\frac{\kappa_n(-\beta)^{n-1}}{n!}. \qquad \textbf{[3]}$$

The first four cumulants are

$$\begin{aligned}
&\kappa_1 = \langle W\rangle \qquad \kappa_3 = \langle(W - \langle W\rangle)^3\rangle \\
&\kappa_2 = \langle(W - \langle W\rangle)^2\rangle = \sigma_W{}^2 \qquad \kappa_4 = \langle(W - \langle W\rangle)^4\rangle - 3\sigma_W^4.
\end{aligned} \qquad \textbf{[4]}$$

Note that the mean work estimator $\Delta\hat{F}_{\mathrm{MW}}$ is just an estimate of the first term on the right side of the cumulant expansion in Eq. **3** and is valid when the perturbation is done reversibly. The FD estimator $\Delta\hat{F}_{\mathrm{FD}}$ is simply an estimate of the first two terms of the cumulant expansion and is valid when the perturbation is done close to equilibrium.

The near-equilibrium regime is defined in this article as the regime in which the perturbation is sufficiently slow for the probability distribution of work $P(W)$ to be Gaussian. It is important to note that, although being close to equilibrium requires a Gaussian work distribution, the inverse is not true. For example, a bead dragged through water by a Hookean spring has a Gaussian work distribution regardless of how fast the perturbation is done (15). The "near-equilibrium" results derived in this article therefore may be relevant even when the perturbation pulls the system far from equilibrium.

The near-equilibrium regime is important because the results are not model-dependent; any system perturbed slowly will have a Gaussian work distribution. In addition, the behavior of the Jarzynski estimator in this regime is a first-order estimate of the behavior in any regime. A Gaussian distribution has the unique property among probability distributions that only a finite number of its cumulants are nonzero. Specifically, only the first two cumulants are nonzero, and Eq. **3** shows that, in this regime, $\Delta F = \langle W\rangle - (1/2)\beta\sigma_W^2$. The following important equality therefore holds in the near-equilibrium regime (16):

$$\bar{W}_{\mathrm{dis}} = \langle W\rangle - \Delta F = \frac{1}{2}\beta\sigma_w^2. \qquad \textbf{[5]}$$

We have defined the dissipated work as the difference between the work and the equilibrium free-energy difference: $W_{\mathrm{dis}} = W - \Delta F$. Eq. **5** is one manifestation of the FD theorem (17) and suggests the particular expression given for the FD estimator in Eq. **2**. In the near-equilibrium regime the variance of the work distribution yields, via the FD relation, an estimate of the dissipated work and hence of the expected overestimate inherent in the mean work estimator.

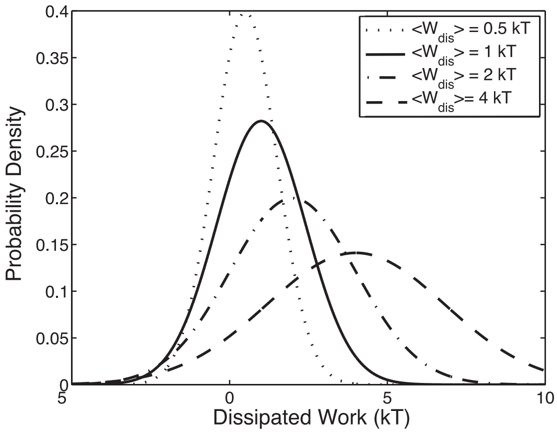

**Fig. 1.** Gaussian probability distributions $P(W_{\mathrm{dis}} = W - \Delta F)$ of dissipated work in the near-equilibrium regime for $\bar{W}_{\mathrm{dis}} \equiv \langle W_{\mathrm{dis}}\rangle$ of 0.5, 1, 2, and 4 kT. As $\bar{W}_{\mathrm{dis}}$ increases, the variance of the work distribution increases ($\sigma_W^2 = 2\bar{W}_{\mathrm{dis}}/\beta$), but $P(W_{\mathrm{dis}} < 0)$ decreases.

The Jarzynski equality can be expressed in terms of the dissipated work without reference to $\Delta F$: $1 = \langle e^{-\beta W_{\mathrm{dis}}}\rangle$. In the near-equilibrium regime the probability distribution of dissipated work $P(W_{\mathrm{dis}})$ is a Gaussian with mean $\bar{W}_{\mathrm{dis}}$ and variance $\sigma_W^2 = 2\bar{W}_{\mathrm{dis}}/\beta$ (see Eq. **5**). Fig. 1 plots this Gaussian $P(W_{\mathrm{dis}})$ for $\bar{W}_{\mathrm{dis}}$ of 0.5, 1, 2, and 4 kT. As $\bar{W}_{\mathrm{dis}}$ increases, the distribution becomes broader. The rare trajectories with negative dissipated work are those heavily weighted by the Jarzynski average, thus it is important to know how efficiently these trajectories are sampled. The probability of observing a trajectory with negative dissipated work is

$$P(W_{\mathrm{dis}} < 0) = \int_{-\infty}^{0} P(W_{\mathrm{dis}})dW_{\mathrm{dis}} = \frac{1}{2}\left[1 - erf(\sqrt{\bar{W}_{\mathrm{dis}}}/2)\right].$$

This is a sharply decreasing function of $\bar{W}_{\mathrm{dis}}$, implying that the efficiency of sampling falls rapidly with increasing $\bar{W}_{\mathrm{dis}}$.

The Jarzynski equality states that $e^{-\beta\Delta F} = \langle e^{-\beta W}\rangle = \int e^{-\beta W}P(W)dW$, so the contribution to the estimate of a given part of the probability distribution is $e^{-\beta W}P(W)$. In the case of Gaussian $P(W)$, $e^{-\beta W}P(W)$ is another Gaussian distribution with the same variance $\sigma_W^2 = 2\bar{W}_{\mathrm{dis}}/\beta$, but a mean given by $\bar{W} - \beta\sigma_W^2 = (\Delta F +$

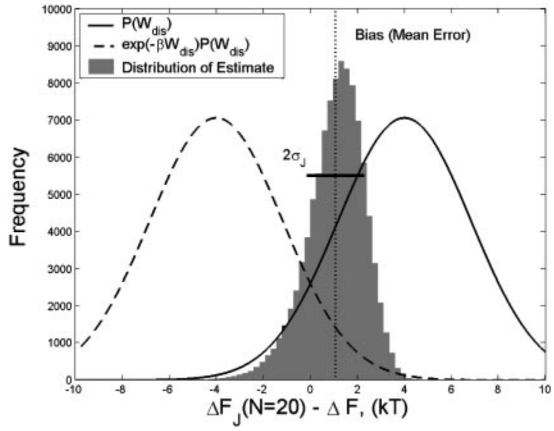

**Fig. 2.** Estimating $\Delta F$ from $N = 20$ near-equilibrium work values in the case of $\bar{W}_{\mathrm{dis}} = 4$ kT. The histogram plots $\Delta\hat{F}_{\mathrm{J}}(N = 20) - \Delta F$ for 100,000 different estimates of $\Delta F$, each obtained by a Jarzynski average of 20 work values sampled from a Gaussian work distribution (solid line). The dashed line corresponds to the contribution of the region to the average, $e^{-\beta W}P(W)$. The dotted line is the bias (mean error): $B_{\mathrm{J}}(N = 20) = \langle\Delta\hat{F}_{\mathrm{J}}\ (N = 20)\rangle - \Delta F \approx 1.07$ kT.

$\bar{W}_{\text{dis}}) - 2\bar{W}_{\text{dis}} = \Delta F - \bar{W}_{\text{dis}}$ (see Fig. 2). The most important part of the work distribution to sample therefore is the region

$$W_{\text{dis}} = -\bar{W}_{\text{dis}} \pm \sigma_W = -\bar{W}_{\text{dis}} \pm \sqrt{2\bar{W}_{\text{dis}}/\beta}.$$

Fig. 2 plots, for the case of $\bar{W}_{\text{dis}} = 4$ kT, the probability distribution of dissipated work $P(W_{\text{dis}})$ as well as the weighted probability distribution $e^{-\beta W_{\text{dis}}} P(W_{\text{dis}})$.

There are three important properties associated with any estimator of the free-energy difference:

$$\begin{aligned}
\text{Bias} \qquad & B(N) = \langle \Delta\hat{F}(N)\rangle - \Delta F \\
\text{Variance} \qquad & \sigma^2(N) = \langle (\Delta\hat{F}(N) - \langle \Delta\hat{F}(N)\rangle)^2\rangle \\
\text{Mean Square Error (MSE)} \qquad & \text{MSE}(N) = \langle (\Delta\hat{F}(N) - \Delta F)^2\rangle \\
& \qquad\qquad = \sigma^2(N) + B^2(N). \qquad [6]
\end{aligned}$$

The bias is simply the difference between the expectation value of the estimator and the true $\Delta F$ and therefore represents a systematic error. The expectation value of an estimator can be expressed by averaging over $M$ free-energy estimates (each obtained from a set of $N$ work values), and letting $M \to \infty$:

$$\langle \Delta\hat{F}(N)\rangle = \lim_{M\to\infty} \frac{1}{M} \sum_{k=1}^{M} \Delta\hat{F}_k(N). \qquad [7]$$

The standard measure for the quality of an estimator is the MSE, which is a combination of the statistical error associated with $\sigma^2(N)$ and the systematic error associated with $B(N)$. If the bias is zero, then $\langle \Delta\hat{F}\rangle = \Delta F$, and Eq. **6** shows that $\text{MSE}(N) = \sigma^2(N)$.

The Jarzynski estimator $\Delta\hat{F}_{\text{J}}(N) = -\beta^{-1} \ln\langle e^{-\beta W}\rangle_N$ is biased for all finite $N$ as a result of the nonlinear averaging. By "biased" we mean that the difference $\langle \Delta\hat{F}_{\text{J}}(N)\rangle - \Delta F \equiv B_{\text{J}}(N)$ is nonzero. This difference is caused by the imperfect sampling of the work distribution when only a finite number of trajectories are averaged. The presence of a bias is most obvious in the limiting case of a single trajectory being sampled ($N = 1$). In this case, the Jarzynski estimator is simply the work value measured because $\Delta\hat{F}_{\text{J}}(N = 1) = -\beta^{-1} \ln e^{-\beta W} = W$. The expectation value of the Jarzynski estimator in this case is therefore $\langle \Delta\hat{F}_{\text{J}}(N = 1)\rangle = \langle W\rangle = \Delta F + \bar{W}_{\text{dis}} > \Delta F$, yielding a bias of $B_{\text{J}}(N = 1) = \bar{W}_{\text{dis}}$.

Fig. 2 plots a histogram of $\Delta\hat{F}_{\text{J}}(N = 20) - \Delta F$ for near-equilibrium switching assuming $\bar{W}_{\text{dis}} = 4$ kT. The dotted line is the bias, or mean error, of the Jarzynski estimator: $B_{\text{J}}(N = 20) = \langle \Delta\hat{F}_{\text{J}}(N = 20)\rangle - \Delta F \approx 1.07$ kT. The variance of the estimator is the squared deviation about $B_{\text{J}}$ ($N = 20$), whereas the $\text{MSE}_{\text{J}}$ is the squared deviation about zero error (corresponding to the actual free-energy difference).

## The Jarzynski Estimator in the Large *N* Limit

In this section we consider the behavior of the Jarzynski estimator in the large $N$ limit for perturbations arbitrarily far from equilibrium. These results were derived independently by Zuckerman and Woolf (18). As shown previously, the Jarzynski bias $B_J$ starts at $B_J(N = 1) = \bar{W}_{\text{dis}}$. It then falls monotonically with increasing $N$ (19) and approaches zero in the limit of infinite $N$. Assuming that the variance $\text{Var}(e^{-\beta W})$ is finite, the central limit theorem guarantees that, for sufficiently large $N$, the random variable $Y = \langle e^{-\beta W}\rangle_N$ will be normally distributed with mean $\bar{Y} = e^{-\beta\Delta F}$ and variance $\sigma_Y^2 = \text{Var}(e^{-\beta W})/N$. We perform a linear expansion of ln ($Y$) around $Y = \bar{Y}$:

$$\begin{aligned}
\Delta\hat{F}_{\text{J}} &= -\frac{1}{\beta}\left[\ln(\bar{Y}) + \frac{(Y-\bar{Y})}{\bar{Y}} - \frac{(Y-\bar{Y})^2}{2!\bar{Y}^2} + \frac{2(Y-\bar{Y})^3}{3!\bar{Y}^3} - \cdots\right] \\
\langle \Delta\hat{F}_{\text{J}}\rangle &= \Delta F + \frac{\text{Var}(e^{-\beta W})}{2\beta e^{-2\beta\Delta F} N} + \cdots \\
B_{\text{J}}(N) &\simeq \frac{\text{Var}(e^{-\beta W})}{2\beta e^{-2\beta\Delta F} N} = \frac{\text{Var}(e^{-\beta W_{\text{dis}}})}{2\beta N}. \qquad [8]
\end{aligned}$$

This expression yields the bias for any well behaved [finite $\text{Var}(e^{-\beta W_{\text{dis}}})$] work distribution for $N \gg \text{Var}(e^{-\beta W_{\text{dis}}})$ or equivalently for $B_J(N) \ll$ kT.

To obtain the variance of the Jarzynski estimator (as defined in Eq. **6**) in the limit of large $N$, we once again expand ln ($Y$) about $Y = \bar{Y}$ and obtain

$$\sigma_J^2(N) = \frac{\text{Var}(e^{-\beta W_{\text{dis}}})}{\beta^2 N} = \frac{2B_J(N)}{\beta}. \qquad [9]$$

We see that, for large $N$ and under very general conditions, there exists a simple relationship between the bias and variance of the Jarzynski estimator.

The MSE is simply a combination of the bias and variance of the estimator: $\text{MSE}_{\text{J}}(N) = \sigma_{\text{J}}^2(N) + B_{\text{J}}^2(N)$. The large $N$ limit is defined as $N \gg \text{Var}(e^{-\beta W_{\text{dis}}})$, so Eq. **9** gives $\sigma_{\text{J}}^2(N) \gg B_{\text{J}}^2(N)$. This implies that $\text{MSE}_{\text{J}} \approx \sigma_{\text{J}}^2$, making the variance the dominant source of noise in the large $N$ limit.

## Near-Equilibrium Regime

For the remainder of the article we focus on probability distributions of work that is Gaussian. This is always the case for perturbations done close to equilibrium but can also be true for perturbations far from equilibrium. If $P(W)$ is Gaussian, then the mean $\langle W\rangle = \Delta F + \bar{W}_{\text{dis}}$ and variance $\sigma_W^2 = 2\bar{W}_{\text{dis}}/\beta$ have a simple relationship (Eq. **5**).

**Bias.** Knowing the magnitude of the Jarzynski bias in the large $N$ limit is useful, but in this limit, as shown above, the bias is not a significant source of error. In addition, we will show that the large $N$ limit is often inaccessible by any experimental or computational strategy. More important therefore is the behavior of the bias for modest $N$, where the bias can be the dominant source of error for any technique using the Jarzynski estimator to recover a free-energy difference. Wood *et al.* (20) performed a linear expansion to prove that for small variance $\sigma_W^2$, the bias is $B_{\text{J}}(N) \approx \beta\sigma_W^2/2N$. Unfortunately, this estimate is only valid for $\sigma_W^2 = 2\bar{W}_{\text{dis}}/\beta \ll (\text{kT})^2$, excluding almost all situations of interest. For example, if $\bar{W}_{\text{dis}} = 5$ kT, then at 50 sampled trajectories the above estimate yields a bias of only 0.1 kT, when in fact computational sampling from a Gaussian work distribution shows that the bias is ≈1 kT, i.e., 10 times larger. An improved estimate of the bias for larger $\bar{W}_{\text{dis}}$ and modest $N$ is therefore highly desirable.

In the near-equilibrium regime, a log–log plot of the numerically calculated small $N$ bias is approximately linear in $N$ (see Fig. 3). This power law behavior of the bias at small $N$ has been noted before and may be true even for systems perturbed far from equilibrium (21). We showed in The Near-Equilibrium Regime and Free-Energy Difference Estimators that $B_{\text{J}}(N = 1) = \bar{W}_{\text{dis}}$, implying that for small $N$ the bias can be approximated as

$$B_{\text{J}}(N) \approx \frac{\bar{W}_{\text{dis}}}{N^{\alpha}}. \qquad [10]$$

Systems with larger $\bar{W}_{\text{dis}}$ have a bias that falls off more slowly, meaning that $\alpha$ is a decreasing function of $\bar{W}_{\text{dis}}$. To determine an accurate approximation for $\alpha(\bar{W}_{\text{dis}})$, we consider the limit of large $N$.

In the near-equilibrium regime $P(W)$ is Gaussian and our large $N$ results (see Eq. **8**) can be evaluated immediately:

$$B_{\text{J}}(N) = \langle \Delta\hat{F}_{\text{J}}\rangle - \Delta F = \frac{(e^{\beta^2\sigma_W^2} - 1)}{2\beta N} = \frac{(e^{2\beta\bar{W}_{\text{dis}}} - 1)}{2\beta \text{N}}. \qquad [11]$$

As a consistency check, it is worth noting that, in the limit of small $\sigma_W^2$, our bias estimate reduces to the estimate of Wood *et al.* $B_{\text{J}}(N) \approx \beta\sigma_W^2/2N$. For small $\sigma_W^2$, the large $N$ limit occurs immediately, thus making the bias fall off as $1/N$ for all $N$. The first order term in our expansion of the bias has the form $(e^{2\beta\bar{W}_{\text{dis}}} - 1)/N$, which means that

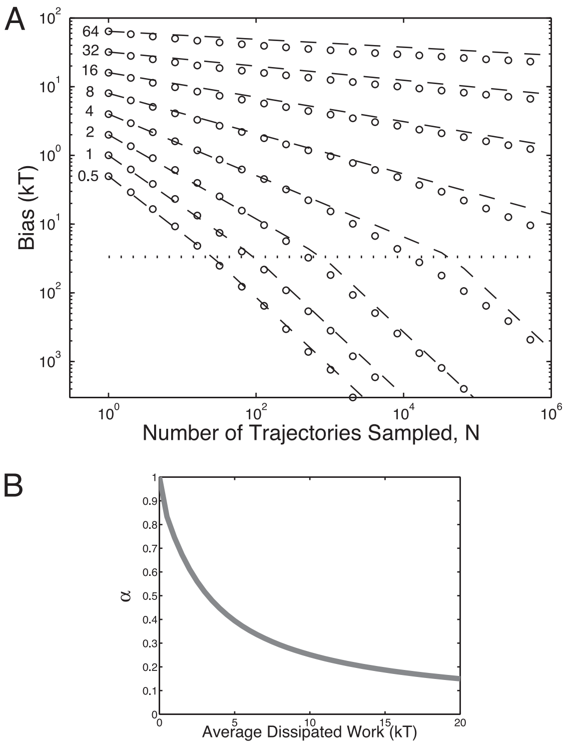

**Fig. 3.** (*A*) Jarzynski bias as a function of *N* for average dissipated work of 0.5, 1, 2, 4, 8, 16, 32, and 64 kT. Each circle is the average of 150,000 sets of work values, each of size *N* (*x* axis). The dashed lines are the approximation derived in the text, with $C = 15$ (see Eqs. **10** and **11**). The bias initially falls off as a power law $B_J(N) = \bar W_{dis}/N^\alpha$, with the fall-off rate $\alpha$ being a function of $\bar W_{dis}$. The horizontal dotted line corresponds to the boundary between the small and large *N* regimes. (*B*) $\alpha$ as a function of $\bar W_{dis}$. The bias falls off more slowly for larger $\bar W_{dis}$.

it is expected to be dominant if $N \gg (e^{2\beta\bar W_{dis}} - 1)$. We will therefore assume that the small $N$ limit intersects the large $N$ limit at $N_C = C(e^{2\beta\bar W_{dis}} - 1)$:

$$B_J(N_C) = \frac{\bar W_{dis}}{N_C^\alpha} = \frac{(e^{2\beta\bar W_{dis}} - 1)}{2\beta N_C} = \frac{1}{2\beta C}. \qquad [12]$$

$C \gg 1$ is a constant that defines how small the bias must be before the large $N$ limit is reached.

Fig. 3 plots $B_J(N)$ for $\bar W_{dis}$ of 0.5, 1, 2, 4, 8, 16, 32, and 64 kT. The data points (circles) correspond to the numerically calculated bias averaged over 150,000 sets of $N$ trajectories, and the dashed lines were obtained by assuming that the bias follows Eq. **10** in the small $N$ regime and Eq. **11** in the large $N$ regime (we have assumed $C = 15$). The single free parameter $C$ defines the intersection point for the two regimes and completely specifies all eight curves. The fit is quite good, although for $\bar W_{dis} \gg$ kT the bias initially falls more quickly than what would be predicted (see Eq. **13**), leading to an average error of ≈20%. In addition, the computed bias smoothly interpolates between the two regimes at $N \approx N_C$, thus falling below the crossing point of the estimated bias (especially for larger $\bar W_{dis}$). Note that we used our analytic expression for the bias in the large $N$ regime to determine the behavior of the bias for small $N$. This seems to work well despite the fact that the large $N$ limit is often impossible to reach. For example, in the case of $\bar W_{dis} = 64$ kT, the large $N$ limit does not begin until $N \approx 10^{57}$ work trajectories have been sampled. Most applications therefore fall in the small $N$ regime.

Eq. **12** can be used to solve for $\alpha$ ($\bar W_{dis}$), the rate at which the bias falls to zero in the small $N$ regime:

$$\alpha = \frac{\ln[2\beta C\bar W_{dis}]}{\ln[C\mathrm{Var}(e^{-\beta \bar W_{dis}})]} = \frac{\ln[2\beta C\bar W_{dis}]}{\ln[C(e^{2\beta\bar W_{dis}} - 1)]}. \qquad [13]$$

Fig. 3*B* plots $\alpha$ as a function $\bar W_{dis}$ assuming that $C = 15$. We see that the bias falls off more slowly as $\bar W_{dis}$ increases, and that $1 \geq \alpha \geq 0$. The bias for small $\bar W_{dis}$ falls off approximately as $1/N$, whereas $\bar W_{dis} = 5$ kT yields, via Eq. **13**, a value of $\alpha \approx 0.4$. It is important to be aware of the unexpectedly slow fall of the Jarzynski bias to avoid underestimating the bias and error in many applications (8, 22).

**Variance.** Once again it is helpful to consider the case of $N = 1$. The variance of the Jarzynski estimator then is simply the variance of the work distribution, $\sigma_J^2(N = 1) = \sigma_W^2 = 2\bar W_{dis}/\beta = 2B_J(N = 1)/\beta$. We therefore see that, in both the large $N$ and $N = 1$ limits, $\sigma_J^2(\mathrm{N}) = 2B_J(N)/\beta$ (see Eq. **9**), a relationship that does not hold at intermediate values of *N*. Nevertheless, for the regime $\bar W_{dis} \sim$ kT, a procedure similar to that used for the bias yields the small $N$ approximation,

$$\sigma_J^2(N) \approx \frac{\sigma_W^2}{N^{\alpha_v}} = \frac{2\bar W_{dis}}{\beta \mathrm{N}^{\alpha_v}}, \qquad [14]$$

where the best fit is obtained with $C_v \approx 50$ in Eq. **13** determining $\alpha$. The fact that $C_v > C$ means that the variance initially decays more quickly than the bias and hence that $\sigma_J^2 < 2B_J/\beta$ for small *N*. Fig. 4*A* plots $\sigma_J^2(N)$ for $\bar W_{dis}$ of 0.5, 1, 2, 4, 8, 16, 32, and 64 kT, with the small $N$ approximation of Eq. **14** shown as a dashed line. For $\bar W_{dis} \geq 5$ kT, the variance cannot be approximated accurately by a power law and initially falls off faster than predicted by Eq. **13**. The behavior of the variance for small $N$ and large $\bar W_{dis}$ is therefore an area for future research, although in this regime the variance is not a significant source of error.

**MSE.** The $\mathrm{MSE}_J(N)$ can be calculated by summing the variance of the estimator with the square of the bias. A summary of the properties of the Jarzynski estimator in the different regimes is available in Table 1. In most regimes the error is dominated by either the variance or bias of the Jarzynski estimator. Most applications are in the small $N$ regime, where the bias is a significant (and often dominant) source of noise.

The most complicated situation is $\bar W_{dis} \sim$ kT for small *N*, where both the variance and bias must be taken into account. In the near-equilibrium regime the $\mathrm{MSE}_J(N)$ is given by

$$\mathrm{MSE}_J(N) = \sigma_J^2(N) + B_J^2(N) = \frac{2\bar W_{dis}}{\beta N^{\alpha_v}} + \left(\frac{\bar W_{dis}}{N^\alpha}\right)^2. \qquad [15]$$

For simplicity, we will assume for now that a single compromise value of $C$ can be used to accurately describe the decay of both the variance $\sigma_J^2$ and the bias $B_J$, making $\alpha_v = \alpha$. In Fig. 4*B* we plotted $\mathrm{MSE}_J$ (*N*) for $\bar W_{dis}$ of 0.5, 1, 2, 4, 8, 16, 32, and 64 kT. The dashed line corresponds to Eq. **15** (assuming $C_M = 40$) in the small $N$ regime and $\mathrm{MSE}_J \approx \sigma_J^2$ in the large N regime (Eq. **9**). The expression for the variance used in Eq. 15 becomes invalid for $\bar{\mathrm{W}}_{dis} \gg$ kT, but in this regime the variance is not a significant source of error. Most of the overestimate of the error for $\bar{\mathrm{W}}_{dis} \gg$ kT is caused by the overestimate of the bias in Eq. **10**.

We can invert Eq. **15** for the $\mathrm{MSE}_J$ (*N*) to get N ($\mathrm{MSE}_J$), the number of trajectories that must be sampled to reach a given MSE:

$$N = \left(\frac{\beta\bar W_{dis}}{\sqrt{1 + \beta^2\mathrm{MSE}_J} - 1}\right)^{1/\alpha(\bar W_{dis})}. \qquad [16]$$

It can be shown that this expression for the number of trajectories necessary to average over to obtain a given level of accuracy increases exponentially with $\bar{\mathrm{W}}_{dis}$. This means that Jarzynski aver-

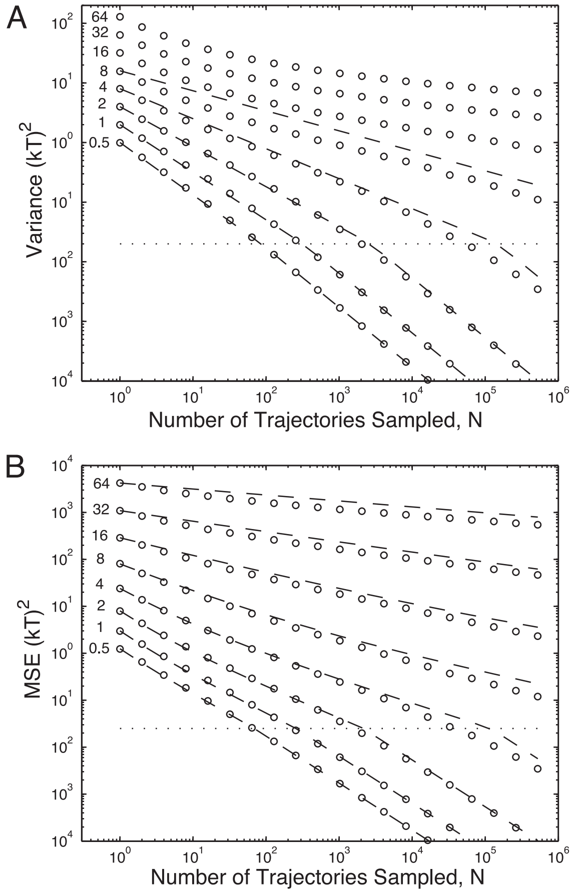

**Fig. 4.** Variance (*A*) and MSE (*B*) of the Jarzynski estimator in the near-equilibrium regime for $\bar{W}_{\mathrm{dis}}$ of 0.5, 1, 2, 4, 8, 16, 32, and 64 kT. Each point is the average of 150,000 sets of work values, each of size *N* (*x* axis). Dashed lines are the approximation derived in the text, with $C_V = 50$ for the variance curves and $C_M = 40$ for the MSE curves. The boundary between small and large *N* is denoted by the horizontal dotted line. The variance approximation in the small *N* regime is valid only for $\bar{W}_{\mathrm{dis}} \sim$ kT because, for $\bar{W}_{\mathrm{dis}} \gg$ kT, the small *N* variance does not follow power law behavior. In this regime the variance is not a significant source of noise.

aging becomes unwieldy for large molecular systems and is experimentally impossible with macroscopic samples.

A common question in computational studies is whether a given amount of computer time should be used to perform a few slow runs (which therefore are close to equilibrium) or many fast runs (which therefore are further from equilibrium) (18). The expression for $\mathrm{MSE_J}$ $(N)$ indicates that, as long as we are in the linear response regime (where $\bar{\mathrm{W}}_{\mathrm{dis}}$ is proportional to the perturbation rate and hence inversely proportional to the simulation time), a lower MSE is always obtained by performing fewer runs with smaller $\bar{\mathrm{W}}_{\mathrm{dis}}$.

**Table 1. Summary of the behavior of the Jarzynski estimator in the different regimes considered in this article**

| | Arbitrary perturbation | | Perturbations near equilibrium | | |
|---|---|---|---|---|---|
| | Small *N* | Large *N* | Small *N* | | Large *N* |
| | | | $\bar{W}_{\mathrm{dis}} \sim$ kT | $\bar{W}_{\mathrm{dis}} \gg$ kT | All $\bar{W}_{\mathrm{dis}}$ |
| Bias, $B_{\mathrm{J}}(N)$ | ? | $\dfrac{\mathrm{Var}(e^{-\beta W_{\mathrm{dis}}})}{2\beta N}$ | $\dfrac{\bar{W}_{\mathrm{dis}}}{N^{\alpha}}$ | $\dfrac{\bar{W}_{\mathrm{dis}}}{N^{\alpha}}$ | $\dfrac{e^{2\beta\bar{W}_{\mathrm{dis}}}-1}{2\beta N}$ |
| Variance, $\sigma_{\mathrm{J}}^2\,(N)$ | ? | $\dfrac{2B_{\mathrm{J}}(N)}{\beta}$ | $\dfrac{\sigma_W^2}{N^{\alpha_V}}\left(\le\dfrac{2B_{\mathrm{J}}}{\beta}\right)$ | $?\left(\le\dfrac{2B_{\mathrm{J}}}{\beta}\right)$ | $\dfrac{2B_{\mathrm{J}}(N)}{\beta}$ |
| MSE | ? | $\approx\sigma_{\mathrm{J}}^2(N)$ | $B_{\mathrm{J}}^2(N)+\sigma_{\mathrm{J}}^2(N)$ | $\approx B_{\mathrm{J}}^2(N)$ | $\approx\sigma_{\mathrm{J}}^2(N)$ |

In the case of the bias we have $\alpha = \ln[2\beta C\bar{W}_{\mathrm{dis}}]/\ln[C(e^{2\beta\bar{W}_{\mathrm{dis}}} - 1)$ with $C = 15$, but for the variance $\alpha_V \equiv \alpha$ $(C = 50)$.

## Comparing Three Free-Energy Difference Estimators

We are now able to compare the quality of the three free-energy estimators in the near-equilibrium regime(23):

$$\mathrm{MSE_{MW}} = \frac{\sigma_W^2}{N} + \frac{\sigma_W^4}{4} = \frac{2\bar{W}_{\mathrm{dis}}}{\beta N} + \bar{W}_{\mathrm{dis}}^2$$

$$\mathrm{MSE_{FD}} = \frac{\sigma_W^2}{N} + \frac{\sigma_W^4}{2(N-1)} = \frac{2\bar{W}_{\mathrm{dis}}}{\beta N} + \frac{2\bar{W}_{\mathrm{dis}}^2}{(N-1)} \qquad [17]$$

$$\mathrm{MSE_{J}} = \frac{\sigma_W^2}{N^{\alpha}} + \frac{\sigma_W^4}{4N^{2\alpha}} = \frac{2\bar{W}_{\mathrm{dis}}}{\beta N^{\alpha}} + \frac{\bar{W}_{\mathrm{dis}}^2}{N^{2\alpha}}.$$

These equations have several characteristics in common. The first term is proportional to $\bar{W}_{\mathrm{dis}}$ and falls off as either $1/N$ or $1/N^{\alpha}$, whereas the second term is proportional to $\bar{W}_{\mathrm{dis}}^2$. The first term in the $\mathrm{MSE_{MW}}$ and $\mathrm{MSE_{FD}}$ equations is the error associated with estimating $\langle W\rangle$. The second term in the $\mathrm{MSE_{MW}}$ equation is the contribution of the bias, whereas the second term in the $\mathrm{MSE_{FD}}$ equation is the error inherent in estimating $\sigma_W^2$.

The square root of the three MSE functions is plotted as a function of $N$ in Fig. 5 for the case of $\bar{W}_{\mathrm{dis}} = 2$ kT. The behavior of the MSE functions in this plot contains many general features. In the near-equilibrium regime, the Jarzynski estimator is always better than or equivalent to the mean work estimator. The errors are equal for $N = 1$, when the two estimators are identical, but the Jarzynski error then falls more quickly with increasing $N$. As $\bar{W}_{\mathrm{dis}} \to 0$, the Jarzynski average is essentially a linear average similar to the mean work estimator, and the two estimators become identical. For larger values of $\bar{W}_{\mathrm{dis}}$, the Jarzynski estimator quickly becomes superior to the mean work estimator. For example, at the relatively small dissipation of $\bar{W}_{\mathrm{dis}} = 4$ kT, we already have that $\mathrm{MSE_J}(N = 2) < \mathrm{MSE_{MW}}\ (N \to \infty) = \bar{W}_{\mathrm{dis}}^2$. This means that, for $\bar{W}_{\mathrm{dis}} \geq 4$ kT, it is better to do a Jarzynski average of two work values than to take the mean of an infinite number of work values. A simple conclusion of this analysis is that the Jarzynski estimator is always better than the mean work estimator for near-equilibrium switching.

The Jarzynski estimator is also superior to the unbiased FD estimator for small $N$. This is perhaps surprising because we are

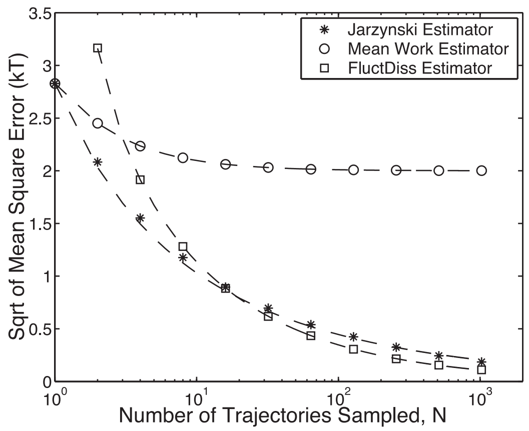

**Fig. 5.** Plot of the square root of the MSE for each of the three different estimators in the case of near-equilibrium switching and $\bar{W}_{\mathrm{dis}} = 2$ kT. Each point is the result of numerical calculation, and the dotted line is the theoretically predicted error (Eq. **17**). Note that the Jarzynski estimator is best for $N \leq 16$ despite the fact that we are considering the near-equilibrium regime where the FD estimator is unbiased. The Jarzynski estimator is always superior to the mean work estimator.

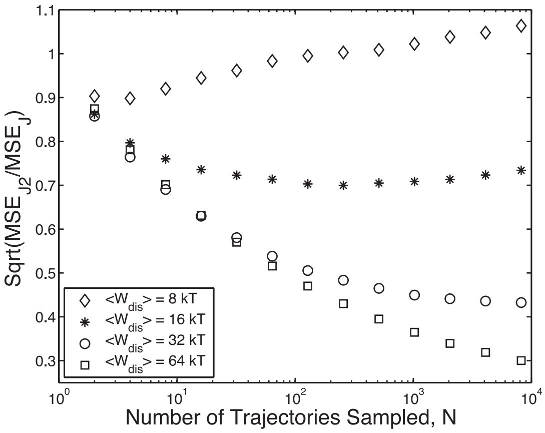

**Fig. 6.** Ratio of $\sqrt{\mathrm{MSE}_{\mathrm{J2}}}$ to $\sqrt{\mathrm{MSE}_{\mathrm{J}}}$ for average dissipated work of 8, 16, 32, and 64 kT (see Eq. **19**). This form of bias correction is more effective for large $\bar{W}_{\mathrm{dis}}$, partly because the bias is overcorrected for small $\bar{W}_{\mathrm{dis}}$ and large *N.* Correcting for the bias in the case of $\bar{W}_{\mathrm{dis}} \sim$ kT requires more complicated techniques.

considering the near-equilibrium regime, where the FD estimator is unbiased. The FD estimator works poorly for small *N* because of the inherent difficulty of estimating $\sigma_W^2$ from limited data. For larger *N* the FD estimator is better than the Jarzynski estimator in the near-equilibrium regime. For arbitrary perturbations the Jarzynski estimator is expected to be superior to the FD estimator for all *N,* because the FD estimator will become biased.

## Bias Correction

In this section we show that understanding the behavior of the bias allows us to construct an improved Jarzynski free-energy difference estimator. The most obvious bias correction is to use $\hat{\bar{W}}_{\mathrm{dis}} = \langle W \rangle - \Delta\hat{F}_{\mathrm{J}}$ as an estimate of the dissipated work, thus yielding $\hat{B}_{\mathrm{J1}} = \hat{\bar{W}}_{\mathrm{dis}}/N^{\alpha(\hat{\bar{W}}_{\mathrm{dis}})}$ as an estimate for the bias and making the final estimator

$$\Delta\hat{F}_{\mathrm{J1}} = \Delta\hat{F}_{\mathrm{J}} - \hat{B}_{\mathrm{J1}}. \qquad \textbf{[18]}$$

This improves the estimator, but $\hat{\bar{W}}_{\mathrm{dis}} = \langle W \rangle - \Delta\hat{F}_{\mathrm{J}}$ is an underestimate of $\bar{W}_{\mathrm{dis}}$ (because of the Jarzynski bias), and thus the bias correction will generally be too small.

A first-order correction is therefore to use $\hat{\bar{W}}_{\mathrm{dis2}} = \langle W \rangle - \Delta\hat{F}_{\mathrm{J}} + \hat{\bar{W}}_{\mathrm{dis}}/N^{\alpha(\hat{\bar{W}}_{\mathrm{dis}})}$, resulting in a final free-energy difference estimator of

$$\Delta\hat{F}_{\mathrm{J2}} = \Delta\hat{F}_{\mathrm{J}} - \hat{B}_{\mathrm{J2}} \qquad \hat{\mathrm{B}}_{\mathrm{J2}} = \hat{\bar{W}}_{\mathrm{dis2}}/N^{\alpha(\hat{\bar{W}}_{\mathrm{dis2}})}. \qquad \textbf{[19]}$$

The effect of using the bias-corrected estimator $\Delta\hat{F}_{\mathrm{J2}}$ in the near-equilibrium regime for $\bar{W}_{\mathrm{dis}}$ of 8, 16, 32, and 64 kT is shown in Fig. 6. Bias correction decreases the typical errors ($\sqrt{\mathrm{MSE}_{\mathrm{J}}}$), especially for larger *N* and $\bar{W}_{\mathrm{dis}}$. The error improvement deteriorates for small $\bar{W}_{\mathrm{dis}}$ and large *N,* because the bias becomes overcorrected. In this regime $\Delta\hat{F}_{\mathrm{J1}}$ has a lower error. Bias correction for $\bar{W}_{\mathrm{dis}} \sim$ kT must be done by using more sophisticated methods such as incorporating a technique called block averaging with the bias correction presented above [typical decreases in error are 20% (unpublished results)].

## Conclusions

Although Jarzynski averaging is an important and widespread technique for calculating free-energy differences, the behavior of the estimator is poorly understood. In particular, the most common method for estimating error is simply to examine the statistical error among repeated experiments. This technique can fail dramatically in the many situations where the bias is the dominant source of error. In addition, a substantial amount of effort must be invested before it can be determined whether a given strategy is practical. This article provides a quantitative estimate for the number of work trajectories that must be sampled to achieve a given MSE as a function of the average dissipated work.

The comparison between the three free-energy estimators in the near-equilibrium regime also leads to several concrete conclusions. First, the mean work estimator should never be used, even for very large systems with concomitantly large dissipated work. The Jarzynski estimator is expected to be optimal for systems that are perturbed violently. Less expected, however, is that the Jarzynski estimator should also be used in the near-equilibrium regime when insufficient data are available to accurately estimate the variance (which is necessary to use the FD estimator). This is often the case because the near-equilibrium regime is only accessed by perturbing the system slowly, thus requiring a large amount of computation.

A better understanding of the behavior of the Jarzynski estimator will allow researchers to develop improved sampling and averaging algorithms. As a preliminary step in that direction, we have shown that a quantitative understanding of the behavior of the Jarzynski estimator bias in the near-equilibrium regime makes it possible to improve the estimator by correcting for the bias.

An obvious direction for future research is to determine how well the methods presented here can be used to analyze the small *N* behavior of the Jarzynski bias and variance for perturbations arbitrarily far from equilibrium. Another important area of study is the behavior of the Jarzynski estimator when the work is measured in the presence of noise. The experimental realities of single-molecule techniques imply that measurement noise will often be an appreciable fraction of the desired signal.

We thank C. Jarzynski, J. Liphardt, H. Shroff, A. Edelstein, S. Dumont, and Z. Bryant for helpful discussions. J.G. thanks the Fannie and John Hertz Foundation for financial support. F.R. was supported by Spanish Ministerio de Ciencia y Tecnología Grant BFM2001-3525 and Generalitat de Catalunya. This research was supported by National Institutes of Health Grant GM-32543 and Department of Energy Grant DE-AC03-76DF00098.

1. Jarzynski, C. (1997) *Phys. Rev. Lett.* **78,** 2690–2693.
2. Jarzynski, C. (1997) *Phys. Rev. E Stat. Phys. Plasmas Fluids Relat. Interdiscip. Top.* **56,** 5018–5035.
3. Wang, G. M., Sevick, E. M., Mittag, E., Searles, D. J. & Evans, D. J. (2002) *Phys. Rev. Lett.* **89,** 050601.
4. Crooks, G. E. (2000) *Phys. Rev. E Stat. Phys. Plasmas Fluids Relat. Interdiscip. Top.* **61,** 2361–2366.
5. Yamada, T. & Kawasaki, K. (1967) *Prog. Theor. Phys.* **38,** 1031.
6. Evans, D. J., Cohen, E. G. D. & Morriss, G. P. (1993) *Phys. Rev. Lett.* **71,** 2401–2404.
7. Gallavotti, G. & Cohen, E. G. D. (1995) *Phys. Rev. Lett.* **74,** 2694–2697.
8. Liphardt, J., Dumont, S., Smith, S., Tinoco, I., Jr., & Bustamante, C. (2002) *Science* **296,** 1832–1835.
9. Ritort, F., Bustamante, C. & Tinoco, I., Jr. (2002) *Proc. Natl. Acad. Sci. USA* **99,** 13544–13548.
10. Hummer, G. & Szabo, A. (2001) *Proc. Natl. Acad. Sci. USA* **98,** 3658–3661.
11. Beveridge, D. & DiCapua, F. (1989) *Annu. Rev. Biophys. Biophys. Chem.* **18,** 431–492.
12. McCammon, J. A. (1991) *Curr. Opin. Struct. Biol.* **1,** 196–200.
13. Frenkel, D. & B. Smit, (1996) in *Understanding Molecular Simulation: From Algorithms to Applications* (Academic, San Diego).
14. Zwanzig, R. (1954) *J. Chem. Phys.* **22,** 1420–1426.
15. Mazonka, O. & Jarzynski, C. (1999) arXiv:cond-mat/9912121.
16. Hermans, J. (1991) *J. Phys. Chem.* **95,** 9029–9032.
17. Callen, H. B. & Welton, T. A. (1951) *Phys. Rev.* **83,** 34–40.
18. Zuckerman, D. & Woolf, T. (2002) *Phys. Rev. Lett.* **89,** 180602.
19. Zuckerman, D. & Woolf, T. (2002) arXiv:, cond-mat/0208015.
20. Wood, R. H., Muhlbauer, W. C. F. & Thompson, P. T. (1991) *J. Phys. Chem.* **95,** 6670–6675.
21. Zuckerman, D. & Woolf, T. (2002) *Chem. Phys. Lett.* **351,** 445–453.
22. Hendrix, D. & Jarzynski, C. (2001) *J. Chem. Phys.* **114,** 5974–5981.
23. Hummer, G. J. (2001) *J. Chem. Phys.* **114,** 7330–7337.