Model for small-sample bias of free-energy calculations applied to Gaussian-distributed nonequilibrium work measurements

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

(Received 27 July 2004; accepted 20 August 2004)

We present a model for the bias of free-energy differences when determined using the nonequilibrium work (NEW) formalism due to Jarzynski. Input to the model is the distribution of work values underlying the NEW calculation, and the bias is estimated by assuming that all of the inaccuracy is incurred by failure to sample work values beyond a specific point in the tail of the distribution. The model is formulated considering both small- and large-sample NEW calculations. It is then applied to the study of bias for cases in which the work distribution is Gaussian. The model is shown to give an excellent description of the bias in situations where the bias is a more significant source of error than the sample variance. A scaling law is presented and it is shown that the bias as a function of sampling can be reduced to a single universal curve, approximately valid for all Gaussian work distributions. This result is used to formulate a simple criterion that can be applied to test if a NEW calculation is effectively free of bias. The criterion is shown to be effective even though it uses the measured (and perhaps biased) free energy as an input. © *2004 American Institute of Physics.* [DOI: 10.1063/1.1806413]

I. INTRODUCTION

Methods for calculating free-energy differences by molecular simulation can be classified as work-based methods and density-of-states methods.[1] The latter relate the free energy to the probability distribution for an order parameter of a thermodynamic variable.[2,3] In the present study we focus on the former. Work-based free energy methods are brought together by the nonequilibrium work (NEW) formula introduced by Jarzynski[4–7]

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

where $\Delta F=F_B-F_A$ is the difference in free energy between two systems, labeled A and B, and $W_{A\rightarrow B}$ is the work involved in a process takes the system from A to B; $\beta=1/kT$ is the reciprocal temperature in energy units. The overbar indicates that the quantity beneath is averaged over many realization of the work process, each beginning from an initial condition sampled from the equilibrium ensemble of system A.

One of the remarkable features of Eq. (1) is that it relates an equilibrium free-energy difference to a process that is not necessarily reversible. More precisely, it relates the free-energy difference to an average over an ensemble of nonequilibrium work processes. Each NEW realization will in general result in a different value of $W_{A\rightarrow B}$, and it is natural then to consider the distribution of such values $p_A(W)$ where the A subscript indicates that the NEW process begins with system A. As the labels A and B are arbitrary, we can equally well consider the free-energy difference evaluated for the reverse process, sampling an ensemble of initial conditions appropriate to system B and transforming it into system A. Then analogous to Eq. (1) we have

$$\exp(+\beta\Delta F)=\overline{\exp(-\beta W_{B\rightarrow A})}, \tag{2}$$

where the sign is changed because we keep the same definition of $\Delta F=F_B-F_A$. We write the distribution of work values for this process as $p_B(W)$. We define W to be in the sense of going from A to B, so $p_A(W)\equiv p_A(W_{A\rightarrow B})$ and $p_B(W)=p_B(-W_{B\rightarrow A})$. Then in connection with Eqs. (1) and (2) the free-energy difference is given from these distributions by

$$\exp(-\beta\Delta F)=\int_{-\infty}^{+\infty}p_A(W)e^{-\beta W}dW, \tag{3a}$$

$$\exp(+\beta\Delta F)=\int_{-\infty}^{+\infty}p_B(W)e^{+\beta W}dW. \tag{3b}$$

The distributions are related[2,7,8]

$$p_A(W)e^{-\beta W}=p_B(W)e^{-\beta\Delta F}. \tag{4}$$

Calculation of the free energy by work-based approaches is prone to significant inaccuracies (perhaps many times kT), and this behavior presents a considerable problem in the application of these methods.[5,9–12] The difficulty is that the measurements can be inaccurate but also very precise, meaning that they can reproducibly give an incorrect result. Thus the usual error measures based on the variance of repeated measurements do not provide a reliable indication of the suitability of a given result. The problem is well recognized, and is becoming better understood. The averages given in Eqs. (1)–(3) clearly have significant weight given to the extreme values of W (smallest values of W in the case of $A\rightarrow B$ processes, and largest values of W for $B\rightarrow A$ processes). Failure to sample well these heavily weighted work values leads to inaccuracy in the calculated free energy.

Attempts have been made to address this issue through statistical analysis of the NEW formula.[9,12] A key quantity is the bias, which is the difference between the correct free energy and the expected value when measured using a finite sample of work values

$$B(M)=\langle \Delta F(M)\rangle - \Delta F, \tag{5}$$

where $\langle \Delta F(M)\rangle$ is the expected value of the NEW-measured ΔF when given by M work measurements. These analyses yield asymptotic relations, good for large sample sizes, in terms of the moments of the distribution of work values p_A or p_B. Unfortunately, such results are not useful in practice, because they do not become valid until sampling is sufficient to make the bias no longer a problem.[12] For this range of sampling, the accuracy is better than the precision of the calculations. One might expect this outcome, as the source of the bias is described in the tails of the work distributions, while the low-order moments are statistics heavily weighted by the bulk of the distributions.

Analyses are needed instead that can describe the bias in the small-sampling range, so that one can have some indication, from results in hand, whether the large-sampling, small-bias regime has been reached. Gore *et al.*[12] have presented an approach to this problem, particular to the case of Gaussian work distributions, and it is reviewed in more detail below. Lu and Kofke[10,11] developed an approach that is suitable for the small-sampling regime. In the following section we expand that treatment to include the very small-sampling regime. In Sec. III we test the generalized form in application to the now well-studied Gaussian-work model, and in doing so we aim to shed some more light on the nature of the Gaussian-work bias. We conclude in Sec. IV.

II. NEGLECTED-TAIL BIAS MODEL

Lu and Kofke[10] proposed a ‘‘neglected-tail’’ bias model that estimates the bias of a NEW calculation by considering explicitly the effect of poor sampling of the tails of the work distribution. This focus permits the approach to be effective in the regime of sampling where the accuracy of the result is of greater concern than its precision. However, the original development does make some assumptions regarding the amount of sampling, so it is not applicable in the very-small sampling regime (say, of the order of 10 work samples). Originally the method was formulated for application to free-energy perturbation[1,3,13] (FEP) calculations, which are not typically conducted with such a small sample. In the more general NEW calculation, each work measurement (be it via computation or experiment[14]) may be too time consuming to permit thousands of samples, so it is helpful for a bias model to apply in this regime too. In this section we develop a more general neglected-tail bias model that is applicable for any amount of sampling. We will find it useful to work with the cumulative distribution function, which for p_A is defined

$$C_A(W)\equiv\int_{-\infty}^{W} p_A(w)\,dw \tag{6}$$

with a similar definition of C_B in terms of p_B. Note that $1-C_A(W)=\int_W^{\infty} p_A(w)\,dw$.

The model is developed using the formalism provided by the work distribution functions. As applied to a general NEW calculation in the direction $A\rightarrow B$, the neglected-tail bias model asserts that *all* of the bias is due to the neglect of contributions below a particular value of W, designated W^*, such that no sampling is contributed below this value, and perfect sampling is achieved for $W>W^*$. Thus, if a free-energy calculation employs M work measurements sampling from the distribution $p_A(W)$, and the least (or most negative) of these has the value W^*, then the (inaccurate) free energy will in this particular case be given by

$$\Delta\hat{F}(W^*;M)=-kT\ln\left[\frac{1}{M}\left(e^{-\beta W^*}+(M-1)\times\frac{\int_{W^*}^{\infty} dW e^{-\beta W} p_A(W)}{1-C_A(W^*)}\right)\right]. \tag{7}$$

The first term of the sum in parentheses is the contribution from the work measurement W^*, which is stipulated to be among the M samples (this is one of the finer points not included in the original neglected-tail model[10]). The other term in the sum is the contribution from perfect sampling of work values greater than W^*, with $p_A(W)$ properly renormalized over this range by dividing by $1-C_A(W^*)$.

From Eq. (7), the bias for the M-sample calculation of given W^* can be written

$$\Delta\hat{F}(W^*;M)-\Delta F=-kT\ln\left[\frac{1}{M}\left(e^{-\beta W^*_{dis}}+(M-1)\times\frac{1-C_B(W^*)}{1-C_A(W^*)}\right)\right]. \tag{8}$$

Here we have employed Eq. (4) to rewrite the numerator on the right in terms of the conjugate distribution $p_B(W)$ (and then C_B), and we introduce the dissipated work defined as the difference between the work and the true free-energy difference:

$$W_{dis}=W-\Delta F.$$

If M is not very small, a reasonable approximation to Eq. (8) is[10]

$$\Delta\hat{F}(W^*;M)-\Delta F=-kT\ln[1-C_B(W^*)], \tag{9}$$

which expresses the bias in terms of the area of the $p_B(W)$ distribution lying below W^*. This area, and thus the bias, will be small to the extent that the p_A and p_B distributions overlap.

Now we turn to the issue of estimating W^*. This quantity will depend on the amount of sampling M, in that as more sampling is performed the greater the likelihood that more negative values of W^* will be encountered. A simple probabilistic argument can be applied to evaluate the probability distribution for W^* given $p_A(W)$ on which M work measurements are sampled.[10] The distribution is given by the probability density that the work value W^* is observed once, times the probability that all other $M-1$ work measurements are greater than W^*,

$$P_A^*(W^*)=Mp_A(W^*)[1-C_A(W^*)]^{M-1}. \tag{10}$$

The multiplicative M accounts for the different position that the W^* value could be found among the M measurements. This distribution is normalized over $W^* \in (-\infty,\infty)$.

The bias can be obtained as the average of the inaccuracy over the distribution of W^*

$$B(M)=\int_{-\infty}^{\infty} dW^* P_A^*(W^*)[\Delta\hat{F}(W^*;M)-\Delta F]. \quad (11)$$

With knowledge of the distribution $p_A(W)$ the bias can be evaluated using Eqs. (8) and (10) in a numerical integration of Eq. (11). Alternatively, we can estimate the bias using Eq. (8) evaluated for a single value of W^*, taken for example as the mean $\overline{W}^*$ or mode $\hat{W}^*$ of P_A^*. Lu and Kofke[10] used the latter choice, for which $\hat{W}^*$ is given by the solution of

$$\left.\frac{d\ln p_A(W)}{dW}\right|_{W=\hat{W}^*}=\frac{(M-1)p_A(\hat{W}^*)}{1-C_A(\hat{W}^*)} \quad (12)$$

or, with a small approximation

$$\left.\frac{d\ln p_A(W)}{dW}\right|_{W=\hat{W}^*}=(M-1)p_A(\hat{W}^*). \quad (13)$$

If the mean is used, then

$$\overline{W}_A^*=M\int_{-\infty}^{\infty} W p_A(W)[1-C_A(W)]^{M-1}dW \quad (14)$$

but we will not consider this choice, as the mode is much more convenient to employ. A less approximate approach estimates the least-work distribution as a Gaussian with mean given by the mode of P^*

$$P_A^*(W^*)\approx\frac{1}{(2\pi)^{1/2}\sigma^*}\exp[-(W^*-\hat{W}^*)^2/2\sigma^{*2}] \quad (15)$$

for which

$$\frac{1}{\sigma^{*2}}=-\left[\frac{\partial^2\ln p_A}{\partial W^2}-\frac{M}{M-1}\left(\frac{\partial\ln p_A}{\partial W}\right)^2\right]_{W=\hat{W}^*}. \quad (16)$$

Corresponding relations are given for the bias of the $B\rightarrow A$ work process.

Thus from the work distributions the expected performance of a simulation of given sampling length can be predicted. Tests have shown that the approximate bias model [Eq. (9) using the mode given by Eq. (13)] works well in describing the performance of large-M FEP calculations.[10]

III. GAUSSIAN WORK DISTRIBUTION

We now consider the neglected-tail bias model in application to the Gaussian-work model, which assumes that the distribution of work values is

$$p_A(W)=\frac{1}{(2\pi)^{1/2}\sigma}\exp[-(W-\overline{W})^2/2\sigma^2]. \quad (17)$$

From Eq. (4), $p_B(W)$ is also a Gaussian with the same variance but shifted by $\beta\sigma^2$

$$p_B(W)=\frac{1}{(2\pi)^{1/2}\sigma}\exp[-(W-\overline{W}+\beta\sigma^2)^2/2\sigma^2]. \quad (18)$$

The free energy difference is

$$\Delta F=\overline{W}-\beta\sigma^2/2. \quad (19)$$

The cumulative distribution functions are

$$C_A(W)=\frac{1}{2}\operatorname{erfc}\left(\frac{\overline{W}-W}{\sqrt{2}\sigma}\right), \quad (20)$$

$$C_B(W)=\frac{1}{2}\operatorname{erfc}\left(\frac{\overline{W}-\beta\sigma^2-W}{\sqrt{2}\sigma}\right), \quad (21)$$

where erfc(x) is the complementary error function.

The Gaussian-work model is important because it applies to all NEW calculations in the limit of a nearly-reversible process.[4,5,12] In a few cases it can apply to processes even when they are performed irreversibly.[12] However, it is also severely limited in its ability to characterize general NEW calculations. For example, the similarity of the p_A and p_B distributions indicates that the inaccuracy of a work calculation is independent of the direction it is performed ($A\rightarrow B$ versus $B\rightarrow A$), specifically that the bias for the two directions are equal in magnitude and opposite in sign. This implies that an accurate work value can be obtained by averaging (inaccurate) work measurements taken in opposite directions. For many irreversible work processes this is not even approximately valid, and in fact the result obtained from one direction is much more accurate than the other.

Gore *et al.*[12] proposed that the bias follows a power law:

$$B(M)=\overline{W}_{dis}/M^{\alpha}, \quad (22)$$

where $\overline{W}_{dis}$ is the mean dissipated work—the difference between the mean of work distribution and ΔF. The exponent α is evaluated by comparing the bias to that for the large-sampling regime, given by

$$B_{M\rightarrow\infty}(M)=\frac{e^{2\beta\overline{W}_{dis}}-1}{2\beta M} \quad (23)$$

and matching the values at a crossover that is observed to be insensitive to σ and determined empirically to satisfy $B(M)=(30\beta)^{-1}$. These workers also performed numerical experiments in which sets of 150 000 free-energy measurements (each sampling M work values) were averaged to determine the bias as a function of sampling length M. In comparison to their numerical calculations, Gore *et al.*[12] find that this model gives a good description of the trends in the bias with M and σ, with a slight overestimate that is diminished at small σ.

The bias according to the full neglected-tail model, Eqs. (8), (10), and (11), is presented in Fig. 1 where it is compared to the bias observed in numerical experiments performed by us and also reported by Gore *et al.* The agreement is excellent. The figure compares the model to the numerical data for values of the bias greater than $0.1kT$, and in fact the model shows significant deviation from the data for bias smaller than the lower bound of the figure. However, numeri-

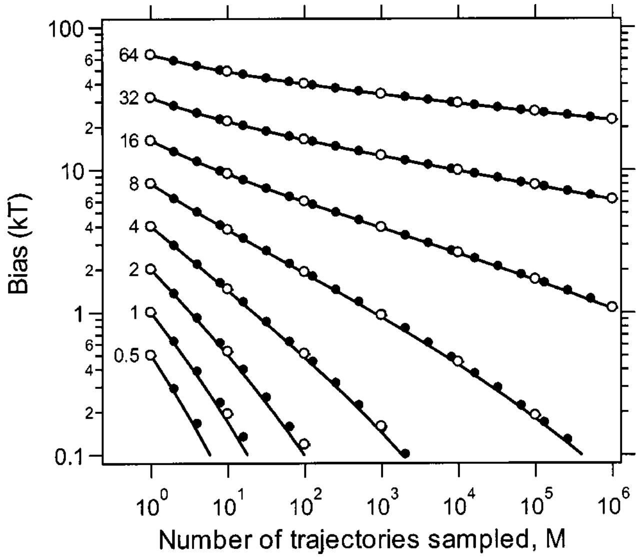

FIG. 1. Average bias as a function of sampling length for Gaussian work distributions of different widths. Numbers to the left of each curve indicate the average dissipated work $\overline{W}_{dis} = \beta\sigma^2/2$. Closed symbols are results of numerical calculations of the bias for 150 000 free-energy calculations each of length M, as reported by Gore *et al.* (Ref. 12). Open circles are results of similar calculations performed by us using 275 000 free-energy calculations. Lines are results of calculations using the full neglected-tail model, Eqs. (8), (10), and (11).

cal calculations by Gore *et al.*[12] show that in these cases the bias contributes a negligible amount to the overall error, and that the noise in the data will overwhelm any inaccuracy due to the sampling bias (in particular, the sample variance is more than ten times greater than the bias squared for bias less than about $0.1kT$ to $0.2kT$). Thus the neglected-tail model is effective in describing the inaccuracy in those situations where the inaccuracy is a significant factor in the overall error.

Figure 2 shows distributions of least-work values (W^*)

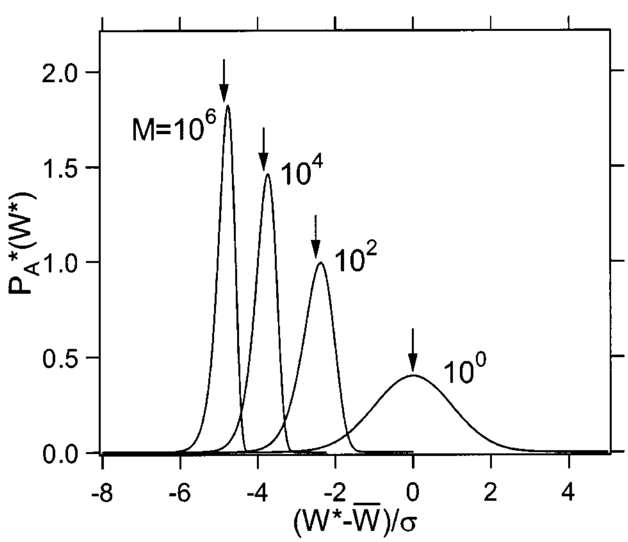

FIG. 2. Least-work distributions $P_A^*(W^*)$ [Eq. (10)] for Gaussian-distributed work distributions $p_A(W)$, scaled to be independent of the Gaussian width σ. Each curve corresponds to a different sampling length M, as indicated. Arrows point to the mean of each distribution.

as given by Eq. (10) for several values of M and scaled to remove the dependence on σ. Clearly as M is increased the least-work distribution shifts to more negative values of W^*, and becomes narrower. This behavior is consistent with the Gaussian approximation to P_A^*, Eq. (15). From Eq. (16) the width σ^* of the P_A^* distribution should go as

$$\frac{1}{\sigma^{*2}} = \frac{1}{\sigma^2}\left[1 + \frac{M}{M-1}\frac{(\hat{W}^* - \overline{W})^2}{\sigma^2}\right], \tag{24}$$

indicating that the distribution becomes narrower as the peak $\hat{W}^*$ of P_A^* moves away from the mean $\overline{W}$ of p_A. The figure also indicates the mean of each distribution for comparison to the mode, or value where the distribution peaks. They are not very different, indicating that a neglected-tail approach based on either Eqs. (12) or (14) should work equally well.

For the Gaussian work distribution, the mode of the least-work distribution is, approximately via Eq. (13),

$$\hat{W}^* = \overline{W} - \sigma\sqrt{\mathbf{W}_L\left[\frac{1}{2\pi}(M-1)^2\right]}, \tag{25}$$

where $\mathbf{W}_L(x)$ is the Lambert W function, defined as the solution to $x = we^w$. This expression for the mode differs from the exact mode by about 20% for $M=2$, but this error decreases quickly with increasing M, and is off by less than 1% for $M>10$. With this estimated mode used for W^* in Eq. (9), the estimate of the bias is

$$B(M) = -kT\,\ln\left\{\frac{1}{2}\operatorname{erfc}\left[-\frac{1}{\sqrt{2}}\left(\sqrt{\mathbf{W}_L\left[\frac{1}{2\pi}(M-1)^2\right]} - \beta\sigma\right)\right]\right\}. \tag{26}$$

Figure 3 compares this model for the bias to the numerical data. The bias is clearly overestimated, although somewhat less so at large σ. In approximating the average bias with that only at the mode of P^*, we overlook the small-bias contributions for $W^* < \hat{W}^*$. In fact, P^* (which is asymmetric) has a longer tail on the $W^* < \hat{W}^*$ side than on $W^* > \hat{W}^*$, as seen in Fig. 2.

Nevertheless, to the extent that the approximation given by Eq. (26) is correct, the Gaussian-work bias depends solely on the group

$$\Pi \equiv \sqrt{\mathbf{W}_L\left[\frac{1}{2\pi}(M-1)^2\right]} - \beta\sigma, \tag{27}$$

which can be thought of as a rescaling or shifting of M. Figure 4 presents the bias when plotted in terms of this group. To a very good approximation, the data collapse onto a single curve. The estimate of this universal curve as given by Eq. (26) is also presented on the figure, where the overestimate is again easily seen.

Still, let us stipulate from our observation of Fig. 4 that—within the constraint of a Gaussian work distribution—bias as a function of the scaled simulation length Π follows a universal curve. With this information, we can develop a

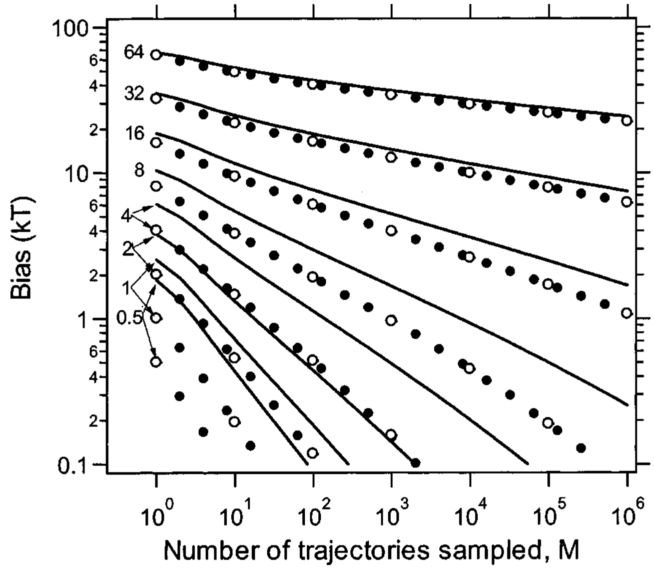

FIG. 3. Average bias as a function of sampling length for Gaussian work distributions of different widths. Numbers to the left of each curve indicate the average dissipated work $\overline{W}_{dis}=\beta\sigma^2/2$. Circles are results of numerical calculations as described in Fig. 1. Lines are results of calculations using the approximate neglected-tail model, Eq. (26).

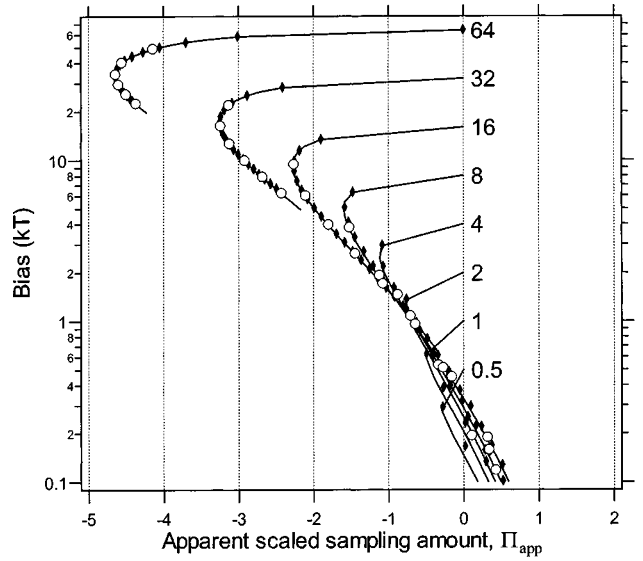

FIG. 5. Average bias as a function of apparent scaled sampling amount [defined in Eq. (29)] for Gaussian work distributions of different widths. Numbers on figure indicate the average dissipated work $\overline{W}_{dis}=\beta\sigma^2/2$ for each curve. Symbols are results of numerical calculations as described in Fig. 1. Lines are results of calculations using the full neglected-tail model, Eqs. (8), (10), and (11), and the left end of each corresponds to $M=9\times10^6$ (the right of each line at $\Pi=0$ corresponds to $M=1$).

means to determine if sufficient sampling has been performed to render a practically unbiased free-energy measurement. We observe in Fig. 4 that the bias becomes small (less than about 0.1, and thus smaller than the statistical noise) for $\approx\Pi>0.5$. Thus if σ in Eq. (27) is given via Eq. (19) from M NEW measurements of the free-energy difference ΔF, then the simulation data would indicate that Π is

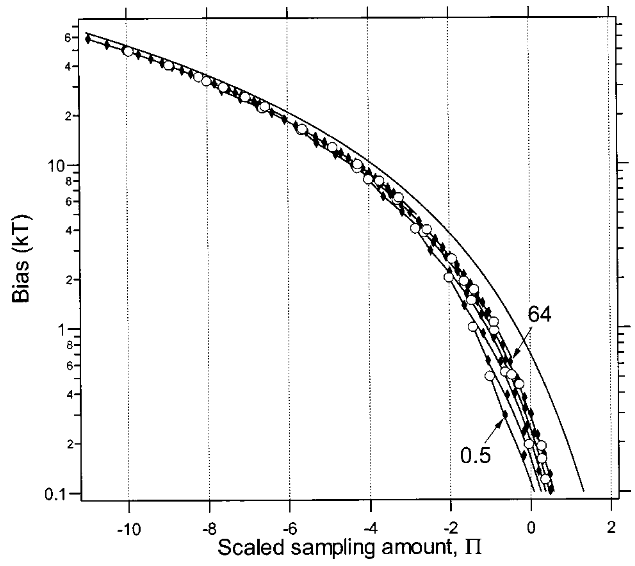

FIG. 4. Average bias as a function of scaled sampling amount [defined in Eq. (27)] for Gaussian work distributions of different widths. Numbers on figure indicate the average dissipated work $\overline{W}_{dis}=\beta\sigma^2/2$ for the leftmost and rightmost curves. Symbols are results of numerical calculations as described in Fig. 1. Lines among the data are results of calculations using the full neglected-tail model, Eqs. (8), (10), and (11). Single solid line above the others describes Eq. (26), which is independent of σ when expressed in terms of Π.

$$\Pi\equiv\sqrt{\mathbf{W}_L\left[\frac{1}{2\pi}(M-1)^2\right]}-\sqrt{2\beta\overline{W}_{dis}}. \qquad (28)$$

If the calculation is afflicted by significant bias it will cause ΔF to be overestimated, and therefore $\overline{W}_{dis}$ will be underestimated. Thus any bias in the calculation will lead to an *overestimate* of Π. Consequently, a test for bias based on the $\Pi>0.5$ criterion could indicate inappropriately that no significant bias exists in the calculation—the tendency of the bias is to hinder the ability of this measure to detect it. To examine the practical effectiveness of a $\Pi>0.5$ criterion for accuracy of the result, in Fig. 5 we plot the bias as a function of Π_{app}, the apparent value of Π that would be computed using the known M and the biased estimate of the free energy,

$$\Pi_{app}=\sqrt{\mathbf{W}_L\left[\frac{1}{2\pi}(M-1)^2\right]}-\sqrt{2\beta\{\overline{W}-[\Delta F+B(M;\sigma)]\}}, \qquad (29)$$

where $B(M;\sigma)$ is the bias as given in Fig. 1; note that $\overline{W}$ must also be given by the NEW data, but we assume that this average contains no bias. Interestingly, the change has actually tightened up the curves, such that they really do appear to follow a common form, at least when the bias is small (less than about kT). More important, the figure shows that any bias in the calculated ΔF would be insufficient to skew Π enough for the $\Pi>0.5$ criterion to provide a false-negative bias-detection result. From this we can conclude that the application of this test provides a reliable means for ascertaining the accuracy of a near-equilibrium NEW free-energy calculation.

In applying this bias detection scheme one should keep in mind that it could be limited by the precision (distinct from the accuracy) of the calculations. Accordingly, one should use the usual statistical methods to determine error bars on Π_{app}, and consider the data unbiased only if Π_{app} exceeds the threshold beyond these confidence limits.

Finally, we note that for large values of its argument,

$$\mathbf{W}_L(x) \sim \ln x - \left(1 - \frac{1}{\ln x}\right)\ln(\ln x) + O\left[\left(\frac{\ln(\ln x)}{\ln x}\right)^2\right], \quad x \to \infty, \tag{30}$$

indicating that at some point exponential increases in the amount of sampling are needed to lead to significant increases in Π, and therefore to reach the threshold that indicates elimination of bias. For large ΔF a single-stage calculation can be impractical, and staging approaches must be considered.[1,15]

IV. CONCLUDING REMARKS

The Gaussian-work model describes an important special case in the application of nonequilibrium work calculations. It applies generally in the near equilibrium limit, in which the work process is performed slowly. It is not generally applicable to fast work processes, and in particular, to the instantaneous switch involved in free-energy perturbation calculations. A notable failure of the Gaussian model in these cases is its inability to characterize asymmetric bias, in which the inaccuracy is greater when the work is performed in one direction versus its opposite.

The neglected-tail model has been shown here to be very effective in characterizing the bias involved in NEW calculations having Gaussian work distributions. Unlike moment-based methods, the approach works best in the cases in which the bias is the primary source of error in the calculation. The neglected-tail model has the benefit of being complete upon specification of only the work distribution. In contrast, the approach taken by Gore *et al.* is developed using some input from numerical studies of the bias. Consequently the results provided by them cannot be directly extended to cases involving non-Gaussian work distributions. The neglected-tail model presented here should be suitable for extension to these important cases.

The collapse of the bias curves when presented in terms of a suitably scaled sampling size is a useful outcome of this work. To summarize, the "recipe" for determining if a NEW measurement is free of bias is as follows:

(a) Perform M nonequilibrium work measurements, collecting work values $\{W_i\}$.

(b) Evaluate the average work $\overline{W} = M^{-1}\sum_{i=1}^{M} W_i$.

(c) Evaluate the free-energy difference $\beta\Delta F = -\ln[M^{-1}\sum_{i=1}^{M}\exp(-\beta W_i)]$.

(d) Evaluate (where $\mathbf{W}_L$ is the Lambert W function) $\Pi \equiv \sqrt{\mathbf{W}_L[1/2\pi(M-1)^2]} - \sqrt{2\beta(\overline{W}-\Delta F)}$.

(e) If $\Pi > 0.5$, the measured ΔF is free of bias; otherwise more sampling is needed. In applying this scheme one should keep in mind that it is based on the presumption of a Gaussian distribution of work values.

Presently there is no sure way to conclude that the underlying work distribution for a given NEW calculation adheres to the Gaussian form. One might develop a simple test for this, but instead it would be better to generalize the Π heuristic so that it can be applied to non-Gaussian work distributions as well. To do so we will need to identify a more general scaling relation, expressed in terms of a parameter that takes the role of σ in Eq. (27) and is distinct for the forward and reverse work distributions. With a suitable definition of such a scaling relation, it may be possible to formulate a universal bias curve that can indicate the likely inaccuracy from the estimated free energy difference and the amount of sampling performed to determine it. Steps in this direction were taken previously by Lu and Kofke[10,11] (scaling the sampling size by the entropy difference between systems A and B), but that development is based on assumptions that are likely to limit its applicability. Further investigations should lead to a more generally applicable scaling measure.

ACKNOWLEDGMENTS

This work was supported by the Division of Chemical Sciences, Office of Basic Energy Sciences of the U.S. Department of Energy (Contract No. DE-FG02-96ER14677). Computer resources have been provided by the University at Buffalo Center for Computational Research. The authors are grateful to Professor C. Bustamante and Jeff Gore for providing data from their numerical bias calculations.

[1] D. A. Kofke and D. Frenkel, in *Handbook of Molecular Modeling*, edited by S. Yip (Kluwer Academic, Dordrecht, 2004); D. A. Kofke, Mol. Phys. **102**, 405 (2004); D. A. Kofke, Fluid Phase Equil. (in press).
[2] C. H. Bennett, J. Comp. Phys. **22**, 245 (1976).
[3] D. Frenkel and B. Smit, *Understanding Molecular Simulation: From Algorithms to Applications*, 2nd ed. (Academic, San Diego, 2002).
[4] C. Jarzynski, Phys. Rev. Lett. **78**, 2690 (1997).
[5] C. Jarzynski, Phys. Rev. E **56**, 5018 (1997).
[6] G. E. Crooks, J. Stat. Phys. **90**, 1481 (1998).
[7] G. E. Crooks, Phys. Rev. E **61**, 2361 (2000).
[8] K. S. Shing and K. E. Gubbins, Mol. Phys. **46**, 1109 (1982).
[9] R. H. Wood, W. C. F. Mühlbauer, and P. T. Thompson, J. Phys. Chem. **95**, 6670 (1991); D. M. Zuckerman and T. B. Woolf, Phys. Rev. Lett. **89**, 180602 (2002).
[10] N. D. Lu and D. A. Kofke, J. Chem. Phys. **114**, 7303 (2001).
[11] N. D. Lu and D. A. Kofke, J. Chem. Phys. **115**, 6866 (2001).
[12] J. Gore, F. Ritort, and C. Bustamante, Proc. Natl. Acad. Sci. U.S.A. **100**, 12564 (2003).
[13] R. W. Zwanzig, J. Chem. Phys. **22**, 1420 (1954).
[14] J. Liphardt, S. Dumont, S. B. Smith, I. Tinoco, and C. Bustamante, Science **296**, 1832 (2002); G. Hummer and A. Szabo, Proc. Natl. Acad. Sci. U.S.A. **98**, 3658 (2001).
[15] N. D. Lu and D. A. Kofke, J. Chem. Phys. **111**, 4414 (1999); N. D. Lu, D. A. Kofke, and T. B. Woolf, J. Phys. Chem. B **107**, 5598 (2003); N. D. Lu, D. A. Kofke, and T. B. Woolf, J. Comput. Chem. **25**, 28 (2004); N. D. Lu, J. K. Singh, and D. A. Kofke, J. Chem. Phys. **118**, 2977 (2003); N. D. Lu, D. Wu, T. B. Woolf, and D. A. Kofke, Phys. Rev. E **69**, 057702 (2004).