A characteristic of Bennett’s acceptance ratio method

A.M. Hahn\textsuperscript{1} and H. Then\textsuperscript{1}

\textsuperscript{1}Institut für Physik, Carl von Ossietzky Universität, 26111 Oldenburg, Germany

A powerful and well-established tool for free-energy estimation is Bennett’s acceptance ratio method. Central properties of this estimator, which employs samples of work values of a forward and its time reversed process, are known: for given sets of measured work values, it results in the best estimate of the free-energy difference in the large sample limit. Here we state and prove a further characteristic of the acceptance ratio method: the convexity of its mean square error. As a two-sided estimator, it depends on the ratio of the numbers of forward and reverse work values used. Convexity of its mean square error immediately implies that there exists an unique optimal ratio for which the error becomes minimal. Further, it yields insight into the relation of the acceptance ratio method and estimators based on the Jarzynski equation. As an application, we study the performance of a dynamic strategy of sampling forward and reverse work values.

PACS numbers: 05.40.-a, 05.70.Ln

Keywords: fluctuation theorem, nonequilibrium thermodynamics

I. INTRODUCTION

A quantity of central interest in thermodynamics and statistical physics is the (Helmholtz) free-energy, as it determines the equilibrium properties of the system under consideration. In practical applications, e.g. drug design, molecular association, thermodynamic stability, and binding affinity, it is usually sufficient to know free-energy differences. As recent progress in statistical physics has shown, free-energy differences, which refer to equilibrium, can be determined via non-equilibrium processes \textsuperscript{1, 2}.

Typically, free-energy differences are beyond the scope of analytic computations and one needs to measure them experimentally or compute them numerically. Highly efficient methods have been developed in order to estimate free-energy differences precisely, including thermodynamic integration \textsuperscript{3, 4}, free-energy perturbation \textsuperscript{5}, umbrella sampling \textsuperscript{6, 7, 8}, adiabatic switching \textsuperscript{9}, dynamic integration \textsuperscript{3, 4}, free-energy perturbation \textsuperscript{5}, asymptotics of work distributions \textsuperscript{13}, optimal protocols \textsuperscript{14}, targeted and escorted free-energy perturbation \textsuperscript{15, 16, 17, 18, 19}.

A powerful \textsuperscript{20, 21, 22} and frequently \textsuperscript{23, 24, 25} used method for free-energy determination is two-sided estimation, i.e. Bennett’s acceptance ratio method \textsuperscript{20}, which employs a sample of work values of a driven nonequilibrium process together with a sample of work values of the time-reversed process \textsuperscript{27}.

The performance of two-sided free-energy estimation depends on the ratio

\[ r = \frac{n_1}{n_0} \]  \hspace{1cm} (1)

of the number of forward and reverse work values used. Think of an experimenter who wishes to estimate the free-energy difference with Bennett’s acceptance ratio method and has the possibility to generate forward as well as reverse work values. The capabilities of the experiment give rise to an obvious question: if the total amount of draws is intended to be \( N = n_0 + n_1 \), which is the optimal choice of partitioning \( N \) into the numbers \( n_0 \) of forward and \( n_1 \) of reverse work values, or equivalently, what is the optimal choice \( r_0 \) of the ratio \( r \)? The problem is to determine the value of \( r \) that minimizes the (asymptotic) mean square error of Bennett’s estimator when \( N = n_0 + n_1 \) is held constant.

While known since Bennett \textsuperscript{20}, the optimal ratio is underutilized in the literature. Bennett himself proposed to use a suboptimal equal time strategy, instead, because his estimator for the optimal ratio converges too slowly in order to be practicable. Even questions as fundamental as the existence and uniqueness are unanswered in the literature. Moreover, it is not always clear a priori whether two-sided free-energy estimation is better than one-sided exponential work averaging. For instance, Shirts et al. have presented a physical example where it is optimal to draw work values from only one direction \textsuperscript{28}.

The paper is organized as follows: in Secs. \textsuperscript{11} and \textsuperscript{11} we rederive two-sided free-energy estimation and the optimal ratio. We also remind that two-sided estimation comprises one-sided exponential work averaging as limiting cases for \( r \rightarrow \pm \infty \), a result that is also true for the mean square errors of the corresponding estimators.

The central result is stated in Sec. \textsuperscript{11}: the asymptotic mean square error of two-sided estimation is convex in the fraction \( r_0 \) of forward work values used. This fundamental characteristic immediately implies that the optimal ratio \( r_0 \) exists and is unique. Moreover, it explains the generic superiority of two-sided estimation if compared with one-sided, as found in many applications.

To overcome the slow convergence of Bennett’s estimator of the optimal ratio, which is based on estimating second moments, in Sec. \textsuperscript{11} we transform the problem into another form such that the corresponding estimator is entirely based on first moments, which enhances the
The density $p_\alpha(w)$ is a normalized harmonic mean of $p_0$ and $p_1$, $p_\alpha(w) = \frac{1}{U_\alpha} \frac{p_0(w) p_1(w)}{\alpha p_0(w) + \beta p_1(w)}$, and thus bridges between $p_0$ and $p_1$, see Fig. 1. In the limit $\alpha \to 0$, $p_\alpha(w)$ converges to the forward work density $p_0(w)$, and conversely for $\alpha \to 1$ it converges to the reverse density $p_1(w)$. As a consequence of the inequality of the harmonic and arithmetic mean, $\left[\frac{1}{p_1^0} + \frac{1}{p_0}\right]^{-1} \leq \alpha p_1 + \beta p_0$, $U_\alpha$ is bounded from above by unity,

$$U_\alpha \leq 1$$

$\forall \alpha \in [0,1]$. Except for $\alpha = 0$ and $\alpha = 1$, the equality holds if and only if $p_0 \equiv p_1$. Using the fluctuation theorem (2), $U_\alpha$ can be written as an average in $p_0$ and $p_1$,

$$U_\alpha = \left\langle \frac{1}{\alpha + \beta e^{-w+\Delta f}} \right\rangle_0 = \left\langle \frac{1}{\beta + \alpha e^{w-\Delta f}} \right\rangle_0,$$

(7)

where the angular brackets with subscript $\gamma \in [0,1]$ denote an ensemble average with respect to $p_\gamma$, i.e.

$$\left\langle g\right\rangle_\gamma = \int_\Omega g(w) p_\gamma(w) dw$$

(8)

for an arbitrary function $g(w)$.

In setting $\alpha = 1$, Eq. (7) reduces to the nonequilibrium work relation [1]

$$1 = \left\langle e^{-w+\Delta f} \right\rangle_0$$

(9)

in the forward direction, and conversely with $\alpha = 0$ we obtain the nonequilibrium work relation in the reverse direction,

$$1 = \left\langle e^{w-\Delta f} \right\rangle_1.$$  

(10)

The last two relations can, of course, be obtained more directly from the fluctuation theorem (2). An important application of these relations is the one-sided free-energy estimation: Given a sample $\{w_1^0 \ldots w_N^0\}$ of $N$ forward work values drawn from $p_0$, Eq. (4) is commonly used to define the forward estimate $\hat{\Delta f}_0$ of $\Delta f$ with

$$\hat{\Delta f}_0 = -\ln \frac{1}{N} \sum_{k=1}^{N} e^{-w_k^0}.$$  

(11)

Conversely, given a sample $\{w_1 \ldots w_N^1\}$ of $N$ reverse work values drawn from $p_1$, Eq. (10) suggests the definition of the reverse estimate $\hat{\Delta f}_1$ of $\Delta f$,

$$\hat{\Delta f}_1 = \ln \frac{1}{N} \sum_{l=1}^{N} e^{w_l^1}.$$  

(12)

If we have drawn both, a sample of $n_0$ forward and a sample of $n_1$ reverse work values, then Eq. (7) can serve...
us to define a two-sided estimate \( \hat{\Delta}f \) of \( \Delta f \) by replacing the ensemble averages with sample averages:

\[
\frac{1}{n_1} \sum_{i=1}^{n_1} \frac{1}{\alpha + \beta e^{-\Delta_f + \hat{\Delta}f}} = \frac{1}{n_0} \sum_{k=1}^{n_0} \frac{1}{\beta + \alpha e^{\hat{\Delta}f}}.
\]  

(13)

\( \hat{\Delta}f \) is understood to be the unique root of Eq. (13), which exists for any \( \alpha \in [0, 1] \). Different values of \( \alpha \) result in different estimates for \( \hat{\Delta}f \). Choosing

\[
\alpha = \frac{n_0}{N}, \quad \beta = \frac{n_1}{N},
\]

(14)

\( N = n_0 + n_1 \), the estimate (13) coincides with Bennett’s optimal estimate, which defines the two-sided estimate with least asymptotic mean square error for a given value \( \alpha = \frac{n_0}{n} \), or equivalently, for a given ratio \( r = \frac{\beta}{\alpha} = \frac{n_1}{n_0} \). We denote the optimal two-sided estimate, i.e. the solution of Eq. (13) under the constraint (14), by \( \hat{\Delta}f_{1-\alpha} \) and simply refer to it as the two-sided estimate. Note that the optimal estimator can be written in the familiar form

\[
\frac{1}{n_1} \sum_{i=1}^{n_1} \frac{1}{1 + e^{-\Delta_f + \hat{\Delta}f + \ln \frac{n_1}{n_0}}} = \frac{1}{n_0} \sum_{k=1}^{n_0} \frac{1}{1 + e^{\hat{\Delta}f - \ln \frac{n_1}{n_0}}}
\]

(15)

In the limit \( \alpha = \frac{n_0}{n} \to 1 \) the two-sided estimate reduces to the one-sided forward estimate (11), \( \hat{\Delta}f_{1-\alpha} \to \hat{\Delta}f_0 \), and conversely \( \hat{\Delta}f_{1-\alpha} \to \hat{\Delta}f_1 \). Thus the one-sided estimates are the optimal estimates if we have drawn data from only one of the densities \( p_0 \) or \( p_1 \).

A characteristic quantity to express the performance of the estimate \( \hat{\Delta}f_{1-\alpha} \) is the mean square error,

\[
\langle (\hat{\Delta}f_{1-\alpha} - \Delta f)^2 \rangle,
\]

(16)

which depends on the total sample size \( N = n_0 + n_1 \) and the fraction \( \alpha = \frac{n_0}{n} \). Here, the average is understood to be an ensemble average in the value distribution of the estimate \( \Delta f_{1-\alpha} \) for fixed \( N \) and \( \alpha \). In the limit of large \( n_0 \) and \( n_1 \), the asymptotic mean square error \( X \) (which then equals the variance) can be written [20] [26]

\[
X(N, \alpha) = \frac{1}{N} \frac{1}{\alpha^2} \left( \frac{1}{U_\alpha} - 1 \right)
\]

(17)

Provided the r.h.s. of Eq. (17) exists, which is guaranteed for any \( \alpha \in (0, 1) \), the \( N \)-dependence of \( X \) is simply given by the usual \( \frac{1}{N} \) factor, whereas the \( \alpha \)-dependence is determined by the function \( U_\alpha \) given in Eq. (19). Note that if a two-sided estimate \( \hat{\Delta}f_{1-\alpha} \) is calculated, then essentially the normalizing constant \( U_\alpha \) is estimated from two sides, 0 and 1, cf. Eqs. (17) and (13). With an estimate \( \hat{\Delta}f_{1-\alpha} \) we therefore always have an estimate of the mean square error at hand. However, the reliability of the latter naturally depends on the degree of convergence of the estimate \( \hat{\Delta}f_{1-\alpha} \). The convergence of the two-sided estimate can be checked with the convergence measure introduced in Ref. [19].

In the limits \( \alpha = \frac{n_0}{n} \to 1 \) and \( \alpha \to 0 \), respectively, the asymptotic mean square error \( X \) of the two-sided estimator converges to the asymptotic mean square error of the appropriate one-sided estimator [20],

\[
\lim_{\alpha \to 1} X(N, \alpha) = \frac{1}{N} \operatorname{Var}_0 \left( \frac{p_1}{p_0} \right) = \frac{1}{N} \operatorname{Var}_0 (e^{-\Delta f})
\]

(18)

and

\[
\lim_{\alpha \to 0} X(N, \alpha) = \frac{1}{N} \operatorname{Var}_1 \left( \frac{p_0}{p_1} \right) = \frac{1}{N} \operatorname{Var}_1 (e^{\Delta f})
\]

(19)

where \( \operatorname{Var}_\gamma \) denotes the variance operator with respect to the density \( p_\gamma \), i.e.

\[
\operatorname{Var}_\gamma (g) = \left\langle (g - \langle g \rangle_\gamma)^2 \right\rangle_\gamma
\]

(20)

for an arbitrary function \( g(w) \) and \( \gamma \in [0, 1] \).

### III. THE OPTIMAL RATIO

Now we focus on the question raised in the introduction: Which value \( \alpha_\gamma \) of \( \alpha \) in the range \( [0, 1] \) minimizes the mean square error (17) when the total sample size, \( N = n_0 + n_1 \), is held fixed?

Let \( M \) be the rescaled asymptotic mean square error given by

\[
M(\alpha) = N \cdot X(N, \alpha),
\]

(21)

which is a function of \( \alpha \) only. Assuming \( \alpha_\gamma \in (0, 1) \), a necessary condition for a minimum of \( M \) is that the derivative \( M'(\alpha) = \frac{dM}{d\alpha} \) of \( M \) vanishes at \( \alpha_\gamma \). Before calculating \( M \) explicitly, it is beneficial to rewrite \( M \) by using the identity

\[
U_\alpha = \int_\Omega \frac{p_0 p_1 (\alpha p_0 + \beta p_1)}{(\alpha p_0 + \beta p_1)^2} dw = \alpha \left\langle \frac{p_0^2}{(\alpha p_0 + \beta p_1)^2} \right\rangle_1 + \beta \left\langle \frac{p_1^2}{(\alpha p_0 + \beta p_1)^2} \right\rangle_0.
\]

(22)

Subtracting \( (\alpha + \beta)U_\alpha^2 = U_\alpha^2 \) from Eq. (22) and recalling the definition (3) of \( p_\alpha \), one obtains

\[
U_\alpha (1 - U_\alpha) = [\alpha \theta_1(\alpha) + \beta \theta_0(\alpha)] U_\alpha^2
\]

(23)

where the functions \( \theta_i \) are defined as

\[
\theta_1(\alpha) = \operatorname{Var}_1 \left( \frac{p_0}{p_1} \right) = \frac{1}{U_\alpha^2} \operatorname{Var}_1 \left( \frac{1}{\alpha + \beta e^{-\Delta f}} \right),
\]

(24)

\[
\theta_0(\alpha) = \operatorname{Var}_0 \left( \frac{p_0}{p_1} \right) = \frac{1}{U_\alpha^2} \operatorname{Var}_0 \left( \frac{1}{\beta + \alpha e^{\Delta f}} \right).
\]
\( \theta_0 \) and \( \theta_1 \) describe the relative fluctuations of the quantities that are averaged in the two-sided estimation of \( \Delta f \), cf. Eq. (13).

With the use of formula (22), \( M \) can be written
\[
M(\alpha) = \frac{\theta_0(\alpha)}{\alpha} + \frac{\theta_1(\alpha)}{\beta}
\]
and the derivative yields
\[
M'(\alpha) = \frac{\theta_1(\alpha)}{\beta^2} - \frac{\theta_0(\alpha)}{\alpha^2} + \frac{\beta \theta_0'(\alpha) + \alpha \theta_1'(\alpha)}{\alpha \beta}.
\]

The derivatives of the \( \theta \)-functions involve the first two derivatives of \( U_\alpha \), which will thus be computed first:
\[
U'_\alpha := \frac{d}{d\alpha} U_\alpha = \int_\Omega \frac{p_0 p_1 (p_1 - p_0) \beta}{(\alpha p_0 + \beta p_1)^2} \, dw
\]
and
\[
U''_\alpha := \frac{d^2}{d\alpha^2} U_\alpha = 2 \int_\Omega \frac{p_0 p_1 (p_1 - p_0)^2}{(\alpha p_0 + \beta p_1)^4} \, dw.
\]

From this equation it is clear that \( U_\alpha \) is convex in \( \alpha \), \( U''_\alpha \geq 0 \), with a unique minimum in \((0,1)\) (as \( U_0 = U_1 = 1 \)). We can rewrite the \( \theta \)-functions with \( U_\alpha \) and \( U'_\alpha \) as follows:
\[
\theta_1(\alpha) = \frac{U_\alpha - \beta U'_\alpha}{U''_\alpha} - 1,
\theta_0(\alpha) = \frac{U_\alpha + \alpha U'_\alpha}{U''_\alpha} - 1.
\]

Differentiating these expressions gives
\[
\theta_1'(\alpha) = \frac{1}{U''_\alpha} \left( U''_\alpha U_\alpha - 2 U'_\alpha \right),
\theta_0'(\alpha) = \frac{\alpha}{U''_\alpha} \left( U''_\alpha U_\alpha - 2 U'_\alpha \right).
\]

\( \theta_0 \) and \( \theta_1 \) are monotonically increasing and decreasing, respectively. This immediately follows from writing the term occurring in the brackets of Eqs. (30) as a variance in the density \( p_\alpha \),
\[
U''_\alpha U_\alpha - 2 U'^2_\alpha = 2 \text{Var}(\frac{p_1 - p_0}{\alpha p_0 + \beta p_1}) U'^2_\alpha,
\]
which is thus positive.

As a consequence of Eq. (30), the relation
\[
\beta \theta_0'(\alpha) + \alpha \theta_1'(\alpha) = 0 \quad \forall \alpha \in [0,1]
\]
holds and \( M' \) reduces to
\[
M'(\alpha) = \frac{\theta_1(\alpha)}{\beta^2} - \frac{\theta_0(\alpha)}{\alpha^2}.
\]

The derivatives of the \( \theta \)-functions do not contribute to \( M' \) due to the fact that the special form of the two-sided estimator (13) originates from minimizing the asymptotic mean square error, cf. (26). The necessary condition for a local minimum of \( M \) at \( \alpha_0 \), \( M'(\alpha_0) = 0 \), now reads
\[
\frac{\beta^2}{\alpha_0^2} = \frac{\theta_1(\alpha_0)}{\theta_0(\alpha_0)},
\]

where \( \beta_0 = 1 - \alpha_0 \) is introduced. Using Eqs. (24) and (2), the condition (34) results in
\[
\text{Var}(\frac{p_1 - p_0}{\alpha p_0 + \beta p_1}) \equiv \text{Var}(\frac{p_1 - p_0}{\alpha p_0 + \beta p_1}).
\]

This means, the optimal ratio \( r_\alpha \) is such that the variances of the random functions which are averaged in the two-sided estimation (15) are equal. However, the existence of a solution of \( M'(\alpha) = 0 \) is not guaranteed in general.

Writing Eq. (35) in the form
\[
\text{Var}(\frac{p_1 - p_0}{\alpha p_0 + \beta p_1}) = \text{Var}(\frac{p_1 - p_0}{\alpha p_0 + \beta p_1})\]
presents the equation from becoming a tautology.

**IV. CONVEXITY OF THE MEAN SQUARE ERROR**

**Theorem.** The asymptotic mean square error \( M(\alpha) \) is convex in \( \alpha \).

In order to prove the convexity, we introduce the operator \( \Gamma_\alpha(f) \) which is defined for an arbitrary function \( f(w) \) by
\[
\Gamma_\alpha(f) = \beta \text{Var}_\alpha(f) + \alpha \text{Var}_\alpha(f) - U_\alpha \text{Var}_\alpha(f) \cdot \gamma_\alpha(\alpha).
\]

**Lemma.** \( \Gamma_\alpha(f) \) is positive semidefinite, i.e.
\[
\Gamma_\alpha(f) \geq 0 \quad \forall f(w).
\]

For \( \alpha \in (0,1) \) and \( f(w) \neq \text{const.} \), the equality holds if and only if \( p_0 \equiv p_1 \).

**Proof of the Lemma.** Let \( \delta f_\gamma = f(w) - \langle f \rangle_\gamma, \gamma \in [0,1] \).

Then
\[
\Gamma_\alpha(f) = \int_\Omega \left( \delta f_0^2 p_0 + \alpha \delta f_1^2 p_1 - \delta f_0^2 \frac{p_0 p_1}{\alpha p_0 + \beta p_1} \right) \, dw
\]
\[
= \int_\Omega \left( \beta \delta f_0^2 p_0 + \alpha \delta f_1^2 p_1 \right) \frac{p_0 p_1}{\alpha p_0 + \beta p_1} \, dw
\]
\[
= \alpha \beta \int_\Omega \left( \delta f_1 p_1 - \delta f_0 p_0 \right)^2 \, dw
\]
\[
+ U_\alpha \left( \beta \langle f \rangle_0 + \alpha \langle f \rangle_1 - \langle f \rangle_\alpha \right)^2,
\]
which is clearly positive. Provided \( f \neq \text{const.} \) and \( \alpha \neq 0,1 \), the integrand in the last line is zero \( \forall w \) if and only if \( p_0 \equiv p_1 \). This completes the proof of the Lemma. □
Proof of the Theorem. Consulting Eqs. (33) and (32), the second derivative of $M$ reads

$$M''(\alpha) = 2 \left( \frac{\theta_1(\alpha)}{\alpha^3} + \frac{\theta_0(\alpha)}{\alpha^3} \right) - \frac{1}{\alpha^2} \theta_0'(\alpha).$$

(40)

Expressing $p_0 = p - \beta d$ and $p_1 = p + \alpha d$ in center- and relative “coordinates” $p = \alpha p_0 + \beta p_1$ and $d = p_1 - p_0$, respectively, gives

$$\theta_1(\alpha) = \frac{1}{U_\alpha^2} \text{Var}_1 \left( \frac{p_0}{p} \right) = \frac{\beta^2}{U_\alpha^2} \text{Var}_1 \left( \frac{d}{p} \right),$$

$$\theta_0(\alpha) = \frac{1}{U_\alpha^2} \text{Var}_0 \left( \frac{p_1}{p} \right) = \frac{\alpha^2}{U_\alpha^2} \text{Var}_0 \left( \frac{d}{p} \right),$$

$$\theta_0'(\alpha) = \frac{2\alpha}{U_\alpha} \text{Var}_\alpha \left( \frac{d}{p} \right).$$

(41)

Therefore, $\frac{1}{2} \alpha U_\alpha^2 M'' = \Gamma_{\alpha} \left( \frac{d}{p} \right)$, which is positive according to the Lemma.

The convexity of the mean square error is a fundamental characteristic of Bennett’s acceptance ratio method. This characteristic allows us to state a simple criterion for the existence of a local minimum of the mean square error in its derivatives at the boundaries. Namely, if

$$M'(0) = \text{Var}_1(e^{w-\Delta f}) - \text{Var}_0(e^{w-\Delta f})$$

(42)

is negative and

$$M'(1) = \text{Var}_1(e^{-w+\Delta f}) - \text{Var}_0(e^{-w+\Delta f})$$

(43)

is positive there exists a local minimum of $M(\alpha)$ for $\alpha \in (0,1)$. Otherwise, no local minimum exists and the global minimum is found on the boundaries of $\alpha$: if $M'(0) > 0$, the global minimum is found for $\alpha = 0$, thus it is optimal to measure work values in the reverse direction only and to use the one-sided reverse estimator \[12\]. Else, if $M'(1) < 0$, the global minimum is found for $\alpha = 1$, implying the one-sided forward estimator \[11\] to be optimal.

In addition, the convexity of the mean square error proves the existence and uniqueness of the optimal ratio, since a convex function has a global minimum on a closed interval.

Corollary. If a solution of $M'(\alpha) = 0$ exists, it is unique and $M(\alpha)$ attains its global minimum ($\alpha \in [0,1]$) there.

V. ESTIMATING THE OPTIMAL RATIO WITH FIRST MOMENTS

In situations of practical interest the optimal ratio is not available a priori. Thus, we are going to estimate the optimal ratio. There exist estimators of the optimal ratio since Bennett. In addition we have just proven that the optimal ratio exists and is unique. However there is still one obstacle to overcome. Yet, all expressions for estimating the optimal ratio are based on second moments, see e.g. Eq. (33). Due to convergence issues, it is not practicable to base any estimator on expressions that involve second moments. The estimator would converge far too slowly. For this reason, we transform the problem into a form that employs first moments, only.

Assume we have given $n_0$ and $n_1$ work values in forward and reverse direction, respectively, and want to estimate $U_a$, with $0 \leq a \leq 1$. According to Eq. (7) we can estimate the overlap measure $U_a$ by using draws from the forward direction,

$$\hat{U}_a(0) = \frac{1}{n_0} \sum_{k=1}^{n_0} \frac{1}{b + a e^{-w_k^{(1)}-\Delta f}},$$

(44)

where $b$ equals $1-a$ and for $\hat{U}_a(1)$ the best available estimate of $\Delta f$ is inserted, i.e. the two-sided estimate based on the $n_0 + n_1$ work values. Similarly, we can estimate the overlap measure by using draws from the reverse direction,

$$\hat{U}_a(1) = \frac{1}{n_1} \sum_{l=1}^{n_1} \frac{1}{a + b e^{-w_l^{(1)}+\Delta f}}.$$  

(45)

Since in general draws from both directions are available, it is reasonable to take an arithmetic mean of both estimates

$$\hat{U}_a = a \hat{U}_a(1) + b \hat{U}_a(0),$$

(46)

where the weighting is chosen such that the better estimate, $\hat{U}_a(0)$ or $\hat{U}_a(1)$, contributes stronger: with increasing $a$ the estimate $\hat{U}_a(1)$ becomes more reliable, as $U_a$ is the normalizing constant of the bridging density $p_a$, Eq. (3), and $p_a \overset{a \to 1}{\sim} p_1$; and conversely for decreasing $a$.

From the estimate of the overlap measure we can estimate the rescaled mean square error by

$$\hat{M}(a) = \frac{1}{ab} \left( \frac{1}{U_a} - 1 \right)$$

(47)

for all $a \in (0,1)$, a result that is entirely based on first moments. The infimum of $\hat{M}(a)$ finally results in an estimate $\hat{\alpha}_a$ of the optimal choice $\alpha_a$, of $\frac{n_a}{N}$,

$$\hat{\alpha}_a \iff \hat{M}(\hat{\alpha}_a) = \inf_a \hat{M}(a).$$

(48)

When searching for the infimum, we also take

$$\hat{M}(0) = \frac{1}{n_0} \sum_{k=1}^{n_0} e^{w_k^{(0)}-\Delta f} - \frac{1}{n_1} \sum_{l=1}^{n_1} e^{w_l^{(1)}-\Delta f},$$

$$\hat{M}(1) = \frac{1}{n_1} \sum_{l=1}^{n_1} e^{-w_l^{(1)}+\Delta f} - \frac{1}{n_0} \sum_{k=1}^{n_0} e^{-w_k^{(0)}+\Delta f}$$

into account which follow from a series expansion of Eq. (47) in $a$ at $a = 0$ and $a = 1$, respectively.
VI. INCORPORATING COSTS

The costs of measuring a work value in forward direction may differ from the costs of measuring a work value in reverse direction. The influence of costs on the optimal ratio of sample sizes is investigated here.

Different costs can be due to a direction dependent effort of experimental or computational measurement of work (unfolding a RNA may be much easier than folding it). We assume the work values to be uncorrelated, which is essential for the validity of the theory presented in this paper. Thus, a source of nonequal costs, which arises especially when work values are obtained via computer simulations, is the difference in the strength of correlations of consecutive Monte-Carlo steps in forward and reverse direction. To achieve uncorrelated draws, the “correlation-lengths” or “correlation-times” have to be determined within the simulation, too. However, this is advisable in any case of two-sided estimation, independent of the sampling strategy.

Let $c_0$ and $c_1$ be the costs of drawing a single forward and reverse work value, respectively. Our goal is to minimize the mean square error $X = \frac{1}{N}M$ while keeping the total costs $c = n_0c_0 + n_1c_1$ constant. Keeping $c$ constant results in

$$N(c, \alpha) = \frac{c}{\alpha c_0 + \beta c_1}$$

which in turn yields

$$X(c, \alpha) = \frac{1}{N(c, \alpha)}M(\alpha).$$

If a local minimum exists, it results from $\frac{\partial}{\partial \alpha}X(c, \alpha) = 0$ which leads to

$$\beta^2 = \frac{c_0\theta_1(\alpha)}{c_1\theta_0(\alpha)},$$

a result Bennett was already aware of. However, based on second moments, it was not possible to estimate the optimal ratio $r_\alpha$ accurately and reliably. Hence, Bennett proposed to use a suboptimal equal time strategy or equal cost strategy, which spends an equal amount of expenses to both directions, i.e. $n_0c_0 = n_1c_1 = \frac{c}{2}$ or

$$\frac{\beta_{\text{ec}}}{\alpha_{\text{ec}}} = \frac{c_0}{c_1},$$

where $\alpha_{\text{ec}} = 1 - \beta_{\text{ec}}$ is the equal cost choice for $\alpha = \frac{n_0}{N}$. This choice is motivated by the following result

$$X(c, \alpha) \geq \frac{1}{2}X(c, \alpha_{\text{ec}}) \quad \forall \alpha \in [0, 1]$$

which states that the asymptotic mean square error of the equal cost strategy is at most sub-optimal by a factor of 2. Note however that the equal cost strategy can be far more sub-optimal if the asymptotic limit of large sample sizes is not reached.

Since we can base the estimator for the optimal ratio $r_\alpha$ on first moments, see Sec. V, we propose a dynamic strategy that performs better than the equal cost strategy. The infimum of

$$\hat{X}(c, \alpha) = \frac{ac_0 + bc_1}{c} \hat{M}(a)$$

results in the estimate $\hat{\alpha}_o$ of the optimal choice $\alpha_0$ of $\frac{n_0}{N}$,

$$\hat{\alpha}_o : \Leftrightarrow \hat{X}(c, \hat{\alpha}_o) = \inf_{\alpha} \hat{X}(c, \alpha).$$

We remark that opposed to $M(\alpha)$, $X(c, \alpha)$ is not necessarily convex. However, a global minimum clearly exists and can be estimated.

VII. A DYNAMIC SAMPLING STRATEGY

Suppose we want to estimate the free-energy difference with the acceptance ratio method, but have a limit on the total amount of expenses $c$ that can be spend for measurements of work. In order to maximize the efficiency, the measurements are to be performed such that $\frac{n_0}{N}$ finally equals the optimal fraction $\alpha_0$ of forward measurements.

The dynamic strategy is as follows:

1. In absence of preknowledge on $\alpha_0$, we start with Bennett’s equal cost strategy as an initial guess of $\alpha_0$.

2. After drawing a small number of work values we make preliminary estimates of the free-energy difference, the mean square error, and the optimal fraction $\alpha_0$.

3. Depending on whether the estimated rescaled mean square error $\hat{M}(a)$ is convex, which is a necessary condition for convergence, our algorithm updates the estimate $\hat{\alpha}_o$ of $\alpha_o$.

4. Further work values are drawn such that $\frac{n_0}{N}$ dynamically follows $\hat{\alpha}_o$, while $\hat{\alpha}_o$ is updated repeatedly.

There is no need to update $\hat{\alpha}_o$ after each individual draw. Splitting the total costs into a sequence $0 < c^{(1)} < \cdots < c^{(p)} = c$, not necessarily equidistant, we can beforehand decide when and how often an update in $\hat{\alpha}_o$ is made. Namely, this is done whenever the actually spent costs reach the next value $c^{(\nu)}$ of the sequence.

The dynamic strategy can be cast into an algorithm.

\textbf{Algorithm.} Set the initial values $n^{(0)}_0 = n^{(0)}_1 = 0$, $\hat{\alpha}_0^{(1)} = \alpha_{\text{ec}}$. In the $\nu$-th step of the iteration, $\nu = 1, \ldots, p$, determine

$$n^{(\nu)}_0 = \lfloor \frac{\hat{\alpha}_o^{(\nu)}N^{(\nu)}}{c} \rfloor$$

$$n^{(\nu)}_1 = \lfloor \frac{c - n^{(\nu)}_0c_0}{c_1} \rfloor$$

where $\alpha_{\text{ec}} = 1 - \beta_{\text{ec}}$ is the equal cost choice for $\alpha = \frac{n_0}{N}$. This choice is motivated by the following result

$$X(c, \alpha) \geq \frac{1}{2}X(c, \alpha_{\text{ec}}) \quad \forall \alpha \in [0, 1]$$

which states that the asymptotic mean square error of the equal cost strategy is at most sub-optimal by a factor of 2. Note however that the equal cost strategy can be far more sub-optimal if the asymptotic limit of large sample sizes is not reached.
with
\[ N^{(\nu)} = \frac{e^{(\nu)}}{a^{(\nu)}_0 c_0 + b^{(\nu)}_0 c_1}, \tag{58} \]
where \( \lfloor \cdot \rfloor \) means rounding to the next lower integer. Then, \( \Delta n^{(\nu)}_0 = n^{(\nu)}_0 - n^{(\nu-1)}_0 \) additional forward and \( \Delta n^{(\nu)}_1 = n^{(\nu)}_1 - n^{(\nu-1)}_1 \) additional reverse work values are drawn. Using the entire present samples, an estimate \( \hat{\Delta} f^{(\nu)} \) of \( \Delta f \) is calculated according to Eq. (12). With the free-energy estimate at hand, \( \hat{M}^{(\nu)}(a) \) is calculated for all values of \( a \in [0,1] \) via Eqs. (44)–(47) and (49), discretized, say in steps \( \Delta a = 0.01 \). If \( \hat{M}^{(\nu)}(a) \) is convex, we update the recent estimate \( \hat{\alpha}^{(\nu)}_0 \) of \( \alpha_0 \) to \( \hat{\alpha}^{(\nu+1)}_0 \) via Eqs. (55) and (56). Otherwise, if \( \hat{M}^{(\nu)}(a) \) is not convex, the corresponding estimate of \( \alpha_0 \) is not yet reliable and we keep the recent value, \( \hat{\alpha}^{(\nu+1)}_0 = \hat{\alpha}^{(\nu)}_0 \). Increasing \( \nu \) by one, we iteratively continue with Eq. (57) until we finally obtain \( \hat{\Delta} f^{(\nu)} \) which is the optimal estimate of the free-energy difference after having spend all costs \( c \).

Note that an update in \( \hat{\alpha}^{(\nu)}_0 \) may result in negative values of \( \Delta n^{(\nu)}_0 \) or \( \Delta n^{(\nu)}_1 \). Should \( \Delta n^{(\nu)}_0 \) happen to be negative, we set \( n^{(\nu)}_0 = n^{(\nu-1)}_0 \) and
\[ n^{(\nu)}_1 = \left[ \frac{c^{(\nu)}_1 - c_0 n^{(\nu-1)}_0}{c_1} \right]. \tag{59} \]

We proceed analogously, if \( \Delta n^{(\nu)}_1 \) happens to be negative.

The optimal fraction \( \alpha_o \) depends on the cost ratio \( c_1/c_0 \), i.e. the algorithm needs to know the costs \( c_0 \) and \( c_1 \). However, the costs are not always known in advance and may also vary over time. Think of a long time experiment which is subject to currency changes, inflation, terms of trade, innovations, and so on. Of advantage is that the dynamic sampling strategy is capable of incorporating varying costs. In each iteration step of the algorithm one just inserts the actual costs. If desired, the breakpoints \( c^{(\nu)}_1 \) may also be adapted to the actual costs. Should the costs initially be unknown (e.g. the “correlation-length” of a Monte-Carlo simulation needs to be determined within the simulation first) one may use any reasonable guess until the costs are known.

**VIII. AN EXAMPLE**

For illustration of results we choose exponential work distributions
\[ p_i(w) = \frac{1}{\mu_i} e^{-\frac{w}{\mu_i}}, \quad w \in \Omega = \mathbb{R}^+, \tag{60} \]
\( \mu_i > 0, \ i = 0,1 \). According to the fluctuation theorem \( 29 \) we have \( \mu_1 = \frac{\mu_0}{1+\mu_0} \) and \( \Delta f = \ln(1 + \mu_0) \).

Exponential work densities arise in a natural way in the context of a two-dimensional harmonic oscillator with Boltzmann distribution \( \rho(x,y) = e^{-\frac{x^2+y^2}{2\sigma^2}}/Z \), where \( Z = 2\pi/\omega^2 \) is a normalizing constant (partition function) and \( (x,y) \in \mathbb{R}^2 \) \( 28 \). Drawing a point \( (x,y) \) from the initial density \( \rho = \rho_0 \), defined by setting \( \omega = \omega_0 \), and switching the frequency to \( \omega_1 > \omega_0 \) instantaneously amounts in the work \( \frac{1}{2}(\omega_1^2 - \omega_0^2)(x^2 + y^2) \). The probability-density of observing a specific work value \( w \) is given by the exponential density \( p_0 \) with \( \mu_0 = \frac{\omega^2 - \omega_0^2}{\omega_0^2} \). Switching the frequency in the reverse direction, \( \omega_1 \rightarrow \omega_0 \), with the point \( (x,y) \) drawn from \( \rho = \rho_1 \) with \( \omega = \omega_1 \), the density of work (with interchanged sign) is given by \( p_1 \) with \( \mu_1 = \frac{\omega^2 - \omega_0^2}{\omega_0^2} \). The free-energy difference of the states characterized by \( \rho_0 \) and \( \rho_1 \) is the log-ratio of their normalizing constants, \( \Delta f = -\ln \frac{\rho_1}{\rho_0} = \ln(1 + \mu_0) \).

A plot of the work densities for \( \mu_0 = 10 \) is enclosed in Fig. 2.

Now, with regard to free-energy estimation, is it better to use one- or two-sided estimators? In other words, we want to know whether the global minimum of \( M(\alpha) \) is on the boundaries \( \{0,1\} \) of \( \alpha \) or not. By the convexity of \( M \), the answer is determined by the signs of the derivatives \( M'(0) \) and \( M'(1) \) at the boundaries. The asymptotic mean square errors \( 18 \) and \( 10 \) of the one-sided estimators are calculated to be
\[ M(1) = \text{Var}_0(e^{-w+\Delta f}) = \frac{\mu_0^2}{1+2\mu_o} \tag{61} \]
for the forward direction and
\[ M(0) = \text{Var}_1(e^{w-\Delta f}) = \frac{\mu_0^2}{1-\mu_0}, \quad \mu_0 < 1, \tag{62} \]
for the reverse direction. For \( \mu_0 \geq 1 \) the variance of the reverse estimator diverges. Note that \( M(0) > M(1) \) holds for all \( \mu_0 > 0 \), i.e. forward estimation of \( \Delta f \) is
always superior if compared to reverse estimation. Furthermore, a straightforward calculation gives

\[ M'(1) = \frac{\mu_0^3(\mu_0 + \xi_-)(\mu_0 - \xi_+)}{(1 + 2\mu_0)^2(1 + 3\mu_0)}, \]  

(63)

where \( \xi_\pm = \frac{1}{2}(\sqrt{17} \pm 3), \) and

\[ M'(0) = -\frac{\mu_0^3(2 + (1 - 2\mu_0)\mu_0)}{(1 - \mu_0^3)^2(1 - 2\mu_0)}, \quad \mu_0 < \frac{1}{2}, \]  

(64)

and \( M'(0) = -\infty \) for \( \mu_0 \geq \frac{1}{2}. \) Thus, for the range \( \mu_0 \in (0, \xi_+) \) we have \( M'(0) < 0 \) as well as \( M'(1) < 0 \) and therefore \( \alpha_o = 1, \) i.e. the forward estimator is superior to any two-sided estimator in this range. For \( \mu_0 \in (\xi_+, \infty) \) we have \( M'(0) < 0 \) and \( M'(1) > 0, \) specifying that \( \alpha_o \in (0, 1), \) i.e. two-sided estimation with an appropriate choice of \( \alpha \) is optimal.

Numerical calculation of the function \( U_\alpha \) and subsequent evaluation of \( M(\alpha) \) allows to find the “exact” optimal fraction \( \alpha_o. \) Examples for \( U_\alpha \) and \( M \) are plotted in Fig. 3.

The behavior of \( \alpha_o \) as a function of \( \mu_0 \) is quite interesting, see Fig. 4. We can interpret this behavior in terms of the Boltzmann distributions as follows. Without loss of generality, assume \( \omega_0 = 1 \) is fixed. Increasing \( \mu_0 \) then means increasing \( \omega_1. \) The density \( \rho_1 \) is fully nested in \( \rho_0, \) cf. the inset of Fig. 2 (remember that \( \omega_1 > \omega_0 \)) and converges to a delta-peak at the origin with increasing \( \omega_1. \) This means that by sampling from \( \rho_0 \) we can obtain information about the full density \( \rho_1 \) quite easily, whereas sampling from \( \rho_1 \) provides only poor information about \( \rho_0. \) This explains why \( \alpha_o = 1 \) holds for small values of \( \mu_0. \) However, with increasing \( \omega_1 \) the density \( \rho_1 \) becomes so narrow that it becomes difficult to obtain draws from \( \rho_0 \) that fall into the main part of \( \rho_1. \) Therefore, it is better to add some information from \( \rho_1, \) hence, \( \alpha_o \) decreases. Increasing \( \omega_1 \) further, the relative number of draws needed from \( \rho_1 \) will decrease, as the density converges towards the delta distribution. Finally, it will become sufficient to make only one draw from \( \rho_1 \) in order to obtain the full information available. Therefore, \( \alpha_o \) converges towards 1 in the limit \( \mu_0 \to \infty. \)

In the following the dynamic strategy proposed in Sec. VII is applied. We choose \( \mu_0 = 1000 \) and \( c_0 = c_1. \) The equal cost strategy draws according to \( \alpha_{ec} = 0.5 \) which is used as initial value in the dynamic strategy. The results of a single run are presented in Figs. 5. Starting with \( N = 100, \) the estimate of \( \alpha_o \) is updated in steps of \( \Delta N = 100. \) The actual forward fractions \( \alpha \) together with the estimated values of the optimal fraction \( \alpha_o \) are shown in Fig. 4. The first three estimates of \( \alpha_o \) are rejected, because the estimated function \( \hat{M}(\alpha) \) is not yet convex. Therefore, \( \alpha \) remains unchanged at the beginning. Afterwards, \( \alpha \) follows the estimates of \( \alpha_o \) and
FIG. 6: (Color online) Displayed are estimated mean square errors $\hat{M}$ in dependence of $\alpha$ for different sample sizes. The global minimum of the estimated function $\hat{M}$ determines the estimate of the optimal fraction $\alpha_o$ of forward work measurements.

FIG. 7: (Color online) Comparison of a single run of free-energy estimation using the equal cost strategy versus a single run using the dynamic strategy. The errorbars are the square roots of the estimated mean square error $X$.

starts to fluctuate about the “exact” value of $\alpha_o$. Some estimates of the function $M$ corresponding to this run are depicted in Fig. 6. For these estimates $\alpha$ is discretized in steps $\Delta \alpha = 0.01$. Remarkably, the estimates of $\alpha_o$ that result from these curves are quite accurate even for relatively small $N$. Finally, Fig. 7 shows the free-energy estimates of the run (not for all values of $N$), compared with those of a single run where the equal cost strategy is used. We find some increase of accuracy when using the dynamic strategy.

In combination with a good a priori choice of the initial value of $\alpha$, the use of the dynamic strategy enables a superior convergence and precision of free-energy estimation, see Figs. 6 and 7. Due to insight into some particular system under consideration, it is not unusual that one has a priori knowledge which results in a better guess for the initial choice of $\alpha$ in the dynamic strategy than starting with $\alpha = \alpha_{ec}$. For instance, a good initial choice is known when estimating the chemical potential via Widom’s particle insertion and deletion [31]. Namely, it is a priori clear that inserting particles yields much more information than deleting particles, since the phase-space which is accessible to particles in the “deletion-system” is

FIG. 8: (Color online) Averaged estimates from 10,000 independent runs with dynamic strategy versus 10,000 runs with equal cost strategy in dependence of the total cost $c = n_0c_0 + n_1c_1$ spend. The cost ratio is $c_1/c_0 = 0.01$, $c_0 + c_1 = 2$, and $\mu_0 = 1000$. The errorbars represent one standard deviation. Here, the initial value of $\alpha$ in the dynamic strategy is $0.5$, while the equal cost strategy draws with $\alpha_{ec} \approx 0.01$. We note that $\alpha_o \approx 0.08$.

FIG. 9: (Color online) Displayed are mean square errors of free-energy estimates using the same data as in Fig. 8. In addition, the mean square errors of estimates with constant $\alpha = \alpha_o$ are included, as well as the asymptotic behavior, Eq. (51). The inset shows that the mean square error of the dynamic strategy approaches the asymptotic optimum, whereas the equal cost strategy is suboptimal. Note that for small sample sizes the asymptotic behavior does not represent the actual mean square error.
effectively contained in the phase-space accessible to the particles in the “insertion-system”, cf. e.g. [19]. A good a priori initial choice for $\alpha$ may be $\alpha = 0.9$ with which the dynamic strategy outperforms any other strategy that the authors are aware of. 

Once reaching the limit of large sample sizes, the dynamic strategy is insensitive to the initial choice of $\alpha$, since the strategy is robust and finds the optimal fraction $\alpha_0$ of forward measurements itself.

**IX. CONCLUSION**

Two-sided free-energy estimation, i.e. the acceptance ratio method [26], employs samples of $n_0$ forward and $n_1$ reverse work measurements in the determination of free-energy differences in a statistically optimal manner. However, its statistical properties depend strongly on the ratio $n_1/n_0$ of work values used. As a central result we have proven the convexity of the asymptotic mean square error of two-sided free-energy estimation as a function of the fraction $\alpha = \frac{n_1}{n_0}$ of forward work values used. From here follows immediately the existence and uniqueness of the optimal fraction $\alpha_0$, which minimizes the asymptotic mean square error. This is of particular interest if we can control the value of $\alpha$, i.e. we can make additional measurements of work in either direction. Drawing such that we finally reach $\frac{n_1}{n_0} = \alpha_0$, the efficiency of two-sided estimation can be enhanced considerably. Consequently, we have developed a dynamic sampling strategy which iteratively estimates $\alpha_0$ and makes additional draws or measurements of work. Thereby, the convexity of the mean square error enters as a key criterion for the reliability of the estimates. For a simple example which allows to compare with analytic calculations, the dynamic strategy has shown to work perfectly.

In the asymptotic limit of large sample sizes the dynamic strategy is optimal and outperforms any other strategy. Nevertheless, in this limit it has to compete with the near optimal equal cost strategy of Bennett which also performs very good. It is worth mentioning that even if the latter comes close to the performance of ours, it is worthwhile the effort of using the dynamic strategy, since the underlying algorithm can be easily implemented and does cost quite anything if compared to the effort required for drawing additional work values.

Most important for experimental and numerical estimation of free-energy differences is the range of small and moderate sample sizes. For this relevant range, it is found that the dynamic strategy performs very good, too. It converges significantly better than the equal cost strategy. In particular, for small and moderate sample sizes it can improve the accuracy of free-energy estimates by half an order of magnitude.

We close our considerations by mentioning that the two-sided estimator is typically far superior with respect to one-sided estimators: assume the support and $p_0$ and $p_1$ is symmetric about $\Delta f$ [22]; then, if the densities are symmetric to each other, $p_0(\Delta f + w) = p_1(\Delta f - w)$, the optimal fraction of forward draws is $\frac{n_0}{n_0} = \frac{1}{2}$ by symmetry. Therefore, if the symmetry is violated too strongly, the optimum will remain near 0.5. Continuous deformations of the densities change the optimal fraction $\alpha_0$ continuously. Thus, $\alpha_0$ does not reach 0 and 1, respectively, for some certain strength of asymmetry. It is exceptionally hard to violate the symmetry such that $\alpha_0$ hits the boundary 0 or 1. In consequence, in almost all situations, the two-sided estimator is superior.

**X. ACKNOWLEDGMENTS**

We thank Andreas Engel for a critical reading of the manuscript.

[1] C. Jarzynski, Phys. Rev. Lett. 78, 2690 (1997).
[2] G. E. Crooks, Phys. Rev. E 60, 2721 (1999).
[3] J. G. Kirkwood, J. Chem. Phys. 3, 300 (1935).
[4] A. Helman and X.-L. Meng, Stat. Science. 13, 163 (1998).
[5] R. W. Zwanzig, J. Chem. Phys. 22, 1420 (1954).
[6] G. M. Torrie and J. P. Valleau, J. Comput. Phys. 23, 187 (1977).
[7] M.-H. Chen and Q.-M. Shao, Annals of Stat. 25, 1563 (1997).
[8] H. Oberhofer and C. Dellago, Comput. Phys. Comm. 179, 41 (2008).
[9] M. Watanabe and W. P. Reinhardt, Phys. Rev. Lett. 65, 3301 (1990).
[10] S. X. Sun, J. Chem. Phys. 118, 5769 (2003).
[11] F. M. Ytreberg and D. M. Zuckerman, J. Chem. Phys. 120, 10876 (2004).
[12] C. Jarzynski, Phys. Rev E 73, 046105 (2006).
[13] S. von Egan-Krieger and A. Engel, arXiv:0807.4079
[14] H. Then and A. Engel, Phys. Rev. E 77, 041105 (2008).
[15] X.-L. Meng and S. Schilling, J. Comput. Graph. Stat. 11, 552 (2002).
[16] C. Jarzynski, Phys. Rev. E 65, 046122 (2002).
[17] H. Oberhofer, C. Dellago, and S. Boresch, Phys. Rev. E 75, 061106 (2007).
[18] S. Vaikuntanathan and C. Jarzynski, Phys. Rev. Lett. 100, 190601 (2008).
[19] A. M. Hahn and H. Then, Phys. Rev. E 79, 011113 (2009).
[20] X.-L. Meng and W. H. Wong, Stat. Sin. 6, 831 (1996).
[21] A. Kong, P. McCullagh, X.-L. Meng, D. Nicolae, and Z. Tan, J. R. Stat. Soc. B 65, 585 (2003).
[22] M. R. Shirts and J. D. Chodera, J. Chem. Phys. 129, 124105 (2008).
[23] D. M. Ceperley, Rev. Mod. Phys. 67, 279 (1995).
[24] D. Frenkel and B. Smit Understanding Molecular Simulation, 2nd ed. (Academic Press, London, 2002).
[25] D. Collin, F. Ritort, C. Jarzynski, S. B. Smith, I. Tinoco Jr and C. Bustamante, Nature (London) 437, 231 (2005).
[26] C. H. Bennett, J. Comput. Phys. 22, 245 (1976).
[27] G. E. Crooks, Phys. Rev. E 61, 2361 (2000).
[28] M. R. Shirts and V. S. Pande, J. Chem. Phys. 122, 144107 (2005).
[29] M. R. Shirts, E. Bair, G. Hooker, and V. S. Pande, Phys. Rev. Lett. 91, 140601 (2003).
[30] J. Gore, F. Ritort and C. Bustamante, Proc. Natl. Acad. Sci. 100, 12564 (2003).
[31] B. Widom, J. Chem. Phys. 39, 2808 (1963).
[32] which is not the case for the densities studied in Sec. VIII.