Improving Free-Energy Estimates from Unidirectional Work Measurements: Theory and Experiment

Matteo Palassini[1,*] and Felix Ritort[1,2]

[1]*Departament de Física Fonamental, Universitat de Barcelona, Diagonal 647, E-08028 Barcelona, Spain*
[2]*CIBER-BBN de Bioingeniería, Biomateriales y Nanomedicina, Instituto de Salud Carlos III, Madrid, Spain*

We derive analytical expressions for the bias of the Jarzynski free-energy estimator from N nonequilibrium work measurements, for a generic work distribution. To achieve this, we map the estimator onto the random energy model in a suitable scaling limit parametrized by $(\log N)/\mu$, where μ measures the width of the lower tail of the work distribution, and then compute the finite-N corrections to this limit with different approaches for different regimes of $(\log N)/\mu$. We show that these expressions describe accurately the bias for a wide class of work distributions and exploit them to build an improved free-energy estimator from unidirectional work measurements. We apply the method to optical tweezers unfolding and refolding experiments on DNA hairpins of varying loop size and dissipation, displaying both near-Gaussian and non-Gaussian work distributions.

The accurate measurement of free-energy changes has important applications in physics, chemistry, and biology. Traditional measurement methods rely on reversible, near-equilibrium transformations, which, however, are often unfeasible. In recent years, new results in nonequilibrium statistical mechanics have suggested ways to measure free-energy changes from experiments (and simulations) far from equilibrium (see [1] for review). The Crooks fluctuation theorem (CFT) [2] states that the probability distribution $p(W)$ of the work W done on a system driven out of equilibrium following an arbitrary finite-time protocol obeys the relation $p(W)/p_R(-W) = e^{(W-\Delta F)/k_B T}$. Here, $p_R(W)$ is the work distribution (WD) for the corresponding time-reversed protocol, ΔF is the free-energy difference between the final and initial equilibrium states [3], and T is the temperature. Hence, ΔF can be estimated in bidirectional experiments by repeating many times the forward and reverse protocol, as demonstrated using single-molecule manipulation techniques [4,5]. An asymptotically unbiased estimator based on the CFT is the acceptance ratio (AR) estimator [6].

In many experimental settings, which we shall call unidirectional, the reverse work cannot be measured. Examples are found in pulling of biopolymers by atomic force microscopy [7,8], steered simulations [9], free-energy landscape reconstruction [10], and single-molecule experiments on protein unbinding, intercalation, specific cation binding, antigen-antibody interactions, and non-native protein conformations. In these cases, an alternative method is provided by a corollary of the CFT, the Jarzynski equality (JE) $\langle e^{-W/k_B T} \rangle = e^{-\Delta F/k_B T}$, where $\langle \cdot \rangle$ is the expectation over $p(W)$ [3]. Given N work measurements $W_1, \ldots, W_N$ under the same protocol, the Jarzynski estimator

$$\Delta F_N = -\log \frac{1}{N} \sum_{i=1}^{N} e^{-W_i} \qquad (1)$$

converges to ΔF from above as $N \to \infty$ (here and henceforth we set $k_B T = 1$ and express all work values in units of $k_B T$ at room temperature). In practice, convergence of ΔF_N requires that rare trajectories with $W_i < \Delta F$ be sufficiently represented, which in turn requires $N \gg \exp(D_{\text{typ}})$, where D_{typ} is the typical value of the dissipated work, $D = W - \Delta F$ [11]. Therefore, ΔF_N is a reliable estimator of ΔF only when D_{typ} is not much larger than $k_B T$. It is thus important to have a quantitative estimate of the bias $B_N = \Delta F_N - \Delta F$. The mathematical problem faced is that of calculating the distribution of a (log)sum of exponentials of independent, identically distributed (i.i.d.) random variables, Eq. (1), which depends on the system- and protocol-specific WD. No closed solution to this problem is available [12], even for a Gaussian WD (GWD). Expansions in N^{-1} [13,14] are only applicable when the bias is of order N^{-1}, i.e., smaller than the $O(N^{-1/2})$ statistical error and thus negligible. In the relevant regime $B_N \gg O(N^{-1})$, power-law interpolations in N [14] and other approximations [12] have been discussed, but no reliable analytical theory exists.

In this Letter, we derive analytical expressions for the bias expectation $\langle B_N \rangle$ for a wide class of WDs and validate them by comparison with exact numerical simulations, also in the regime of large bias. We use these results to build an improved unidirectional free-energy estimator by correcting for the bias of Eq. (1). We then discuss unfolding and refolding experiments on DNA hairpins, which allow us to test our method against the bidirectional AR estimator.

The experimental setup is shown in Fig. 1(a). We synthesized five hairpins (A, B, C, D, E) with identical stem and ($GAAA\dots$) loops of 4, 6, 12, 16, 20 bases, respectively [Fig. 1(b)]. The hairpins are inserted between two short (29 bp) dsDNA handles to improve signal-to-noise resolution [15]. The construct is tethered to two beads, one held by a pipette, the other by an optical trap created by counter-propagating laser beams [16]. The light deflected by the trapped bead provides a direct measurement of the force acting on the molecule. By moving the trap away from the pipette at constant velocity, the hairpin is stretched until it unfolds. Subsequent reversal of the velocity causes the hairpin to refold. By repeating this cycle ($\approx$ 200–1000 times per experiment) we collect the histogram of the WDs $p_{U,R}(W)$ for the work to unfold (U) and refold (R) the hairpin, measured by integrating the force-distance curves [Fig. 1(c)] for the forward and reverse part of each cycle (see Sec. 1 in [17] for details). We divide the data in blocks of N cycles, compute ΔF_N for each block, and average over the blocks to estimate $\langle \Delta F_N \rangle$ for U and R
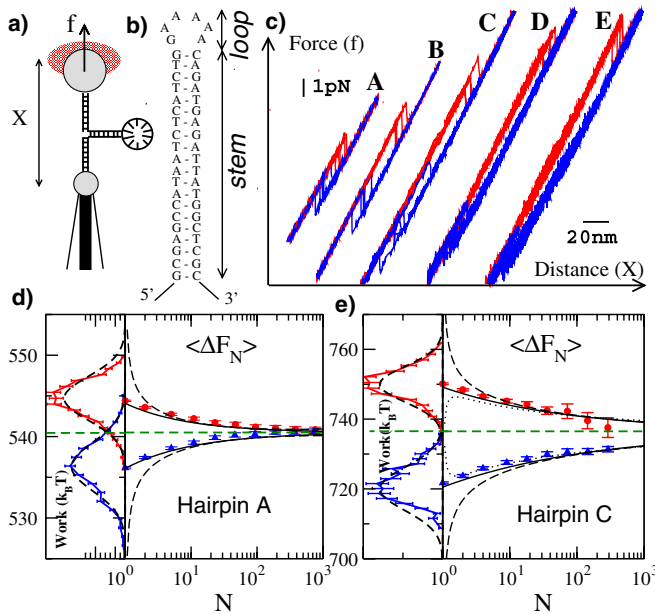

FIG. 1 (color online). Bias measurements in DNA hairpins. (a) Experimental setup. (b) Hairpin B sequence. (c) Examples of force-distance cycles for hairpins A–E. (d) Left: Histograms of $p_U(W)$ (upper) and $p_R(-W)$ (lower) for hairpin A pulled at 400 nm/s. The horizontal line is the AR estimate $\Delta F_{AR} = 540.5 k_B T$, giving $|\langle D \rangle| = |\langle W \rangle - \Delta F_{AR}| = 4.3 k_B T$ (U), $3.9 k_B T$ (R). The lines are GWDs fitting the lower (upper) tail of $p_U(W)$ [$p_R(-W)$]. Right: Jarzynski estimator $\langle \Delta F_N \rangle$ as a function of N for U and R. Errors are estimated by jackknife. The lines represent $\Delta F_{AR} + \langle B_N \rangle$ with $\langle B_N \rangle$ given by Eq. (6) (dashed line) and Eq. (7) (continuous line) for the GWD case, assuming $D_c = \langle D \rangle = \mu$, where μ is estimated from the GWD fit to the tails. (e) Same as (d) but for hairpin C pulled at 65 nm/s [$|\langle D \rangle| = 13.6 k_B T$ (U); $14.8 k_B T$ (R)]. Also shown is $\Delta F_{AR} + \langle B_N \rangle$ with $\langle B_N \rangle$ given by Eq. (5) (dotted line).

separately. As shown in Figs. 1(d) and 1(e) for hairpins A and C, $\langle \Delta F_N \rangle$ tends to the AR estimate ΔF_{AR} for large N, from opposite sides for U and R. Note that the dissipation increases with loop size and pulling speed.

We analyze theoretically the bias for a generic WD with finite mean and an unbounded lower tail which decays as

$$p(W) \sim q \frac{\Omega^{\alpha-1}}{|W - W_c|^\alpha} \exp\left(-\frac{|W - W_c|^\delta}{\Omega^\delta}\right), \quad (2)$$

for $W \ll W_c$, where W_c is a characteristic work value, $\Omega > 0$ measures the tail width, and q is a normalization constant. For the JE to hold, generally one must have $\delta > 1$ [18]. Two key parameters in the following are

$$\mu \equiv (\delta - 1)\left(\frac{\Omega}{\delta}\right)^{\delta/(\delta-1)}, \qquad \lambda \equiv (\delta - 1)\frac{\log N}{\mu}. \quad (3)$$

A saddle point calculation gives $\langle e^{-kD} \rangle \sim \exp(\mu k^{\delta/(\delta-1)} - kD_c)$ for large Ω, where $D_c = W_c - \Delta F$. Hence the JE implies $\mu \to D_c$ in this limit. An example of a WD obeying Eq. (2) is a GWD with mean $\langle W \rangle = W_c$ and variance $\sigma_W^2 = \Omega^2/2$ (i.e., $\alpha = 0$, $\delta = 2$, $q = \pi^{-1/2}$), for which $\langle e^{-kD} \rangle = \exp(\mu k^2 - k\langle D \rangle)$ and thus $\mu = \langle D \rangle = D_c = \sigma_W^2/2$ exactly for all Ω. This relation allows one to define another unidirectional estimator $\Delta F_v = \langle W \rangle - \sigma_W^2/2$ [19], since $\Delta F_v = \Delta F$ for the GWD.

Scaling limit.—Our strategy consists in computing first $\langle B_N \rangle$ in a suitable scaling limit, and then the finite-N corrections to this limit. We obtain the scaling limit by mapping the problem onto the random energy model (REM) [20] as $B_N = D_c + \log N - \log Z_N(\beta = \Omega/(\log_2 N)^{(\delta-1)/\delta})$, where $Z_N(\beta) = \sum_{i=1}^N \exp[-\beta(\log_2 N)^{(\delta-1)/\delta} E_i]$ is the REM partition function and the i.i.d. variables E_i have a distribution decaying as $|E|^{-\alpha} \exp(-|E|^\delta)$ for $E \ll -1$ [20,21]. From the known limit of $N^{-1}\langle \log Z_N(\beta) \rangle$ for $N \to \infty$ [21,22], we obtain $\langle B_N \rangle \to \mathcal{B}_{REM}$ in the scaling limit $(N, \Omega) \to \infty$ with λ finite. For $\lambda > 1$, all terms in Z_N give a finite contribution and we find $\mathcal{B}_{REM} = D_c - \mu$. For $\lambda < 1$, corresponding to the glass phase of the REM [21], Z_N is dominated by a finite number of terms and we obtain

$$\mathcal{B}_{REM} = D_c + \frac{\mu(\lambda - \delta\lambda^{1/\delta})}{\delta - 1}$$
$$= D_c + \log N - \Omega(\log N)^{1/\delta}. \quad (4)$$

Figure 2(a) shows the approach to the scaling limit as Ω increases, for the GWD case. Significant deviations occur for moderate Ω, from which the need to compute finite-N corrections is apparent.

Finite-N corrections.—One must distinguish three regimes, which require different analytical approaches: $\lambda > 1$, $\lambda \ll 1$, and $\lambda \lesssim 1$. For $\lambda > 1$, by partially resumming the $1/N$ expansion we obtain a closed expression for $\langle B_N \rangle$ that improves considerably over the truncated expansions previously considered [13], which are valid only for

$\lambda \gg 1$ (see Sec. 2.1 in [17]). However, the most relevant regime in applications of the JE is $\lambda < 1$, since in practice one usually has $N \ll \exp(D_{\text{typ}})$. In the limit $\lambda \ll 1$, using an extreme-value approach (Sec. 2.2 in [17]), we obtain to leading order

$$\langle B_N \rangle = \mathcal{B}_{\text{REM}} - \lambda^{(1-\delta)/\delta} \left[\gamma_E + \frac{1 - \alpha - \delta}{\delta} \log\log N \right. $$
$$\left. + \log(q/\delta) \right], \tag{5}$$

where γ_E is the Euler-Mascheroni constant. Cook and Derrida [23] were able to compute the finite-N corrections in the critical region $\lambda \lesssim 1$ for the GWD case in the context of the REM. We have extended their traveling-wave approach to the more general case of Eq. (2). In this way (see Sec. 2.3 in [17] for details) we recover Eq. (5) for $\lambda \ll 1$, while for $\lambda \lesssim 1$ we obtain

$$\langle B_N \rangle = \mathcal{B}_{\text{REM}} + \gamma_E - \lambda^{(1-\delta)/\delta} \left[\gamma_E + \frac{1 - \alpha - \delta/2}{\delta} \log\log N \right.$$
$$\left. + \frac{1}{2} \log \frac{\pi q^2}{2\delta(\delta - 1)} + \theta^2 + \log \text{erfc}(\theta) \right], \tag{6}$$
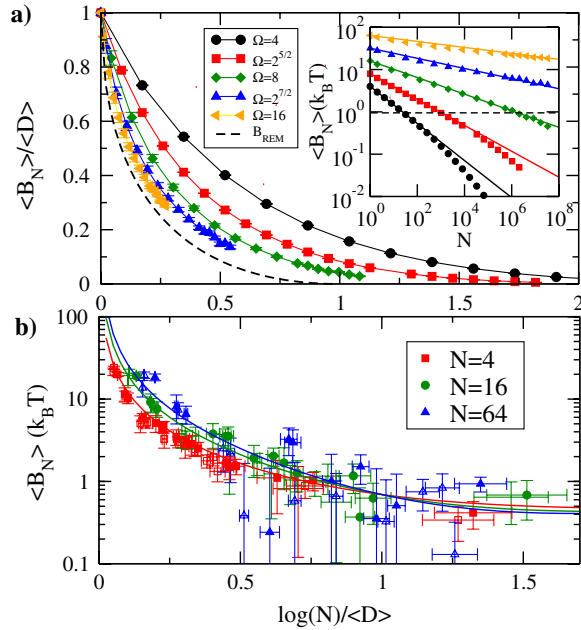

FIG. 2 (color online). Finite-N corrections to the bias for a GWD. (a) Convergence of $\langle B_N \rangle$ to its scaling limit. The points joined by lines are averages of $B_N = -\log N^{-1} \sum_i e^{-D_i}$ over many sets $(D_1, \dots, D_N)$ sampled from a GWD with variance $\Omega^2/2$ and mean $\langle D \rangle = \Omega^2/4$. The dashed line represents $\mathcal{B}_{\text{REM}} = \mu(1 - \lambda^{1/2})^2$ [Eq. (4) for $\delta = 2$]. Inset of (a): The unscaled data. The continuous lines represent Eq. (7). The horizontal dashed line $\langle B_N \rangle = 1$ indicates the accuracy limit common in biophysical studies. (b) Bias $|\langle B_N \rangle|$ (estimated as $|\langle \Delta F_N \rangle - \Delta F_{\text{AR}}|$) for all experiments on hairpins A, B, C and $N = 4, 16, 64$, including both U and R. The pulling speed is in the range 25–300 nm/s, and $|\langle D \rangle|$ is in the range 1–20$k_B T$. The lines show Eq. (6) for the GWD.

where erfc is the complementary error function and $\theta = (\lambda^{(1-\delta)/\delta} - 1) \sqrt{\delta \log N} / \sqrt{2(\delta - 1)}$.

Numerical test.—Equations (5) and (6) provide, via Eqs. (3) and (4), explicit expressions for $\langle B_N \rangle$ as a function of D_c, $\log N$, and the shape parameters δ, Ω, α, q of the WD tail. To illustrate the validity of these expressions, we computed numerically $\langle B_N \rangle$ by sampling W_i from the GWD and from the Weibull WD (WWD): $p(W) = \delta \Omega^{-\delta} |W - W_c|^{\delta - 1} \exp -(|W - W_c|/\Omega)^\delta$ for $W \leq W_c$; $p(W) = 0$ for $W > W_c$, where W_c is fixed numerically by imposing the JE. The WWD satisfies Eq. (2) ($\alpha = 1 - \delta$, $q = \delta$) and allows us to model tails falling faster ($\delta > 2$) or more slowly ($\delta < 2$) than a GWD. In this case $\langle W \rangle = W_c - \Omega \Gamma(1 + \delta^{-1})$, $W_{\text{typ}} = W_c - \Omega (\frac{\delta - 1}{\delta})^{1/\delta}$.

In their respective range of validity in λ, Eqs. (5) and (6) agree very well with the numerical data for the entire range tested ($1.1 \leq \delta \leq 3$, $1.41 \leq \Omega \leq 16$) also for large bias, as shown for some cases in Fig. 3(a) (for more examples see Fig. S4 in [17]). Furthermore, substituting D_c with μ in Eq. (4) worsens only slightly the agreement, an important observation for the following (see Sec. 2.5 in [17]). In the special case of the GWD, Eq. (6) gives $\langle B_N \rangle = \log 2$ at $\lambda = 1$. The empirical one-parameter power law
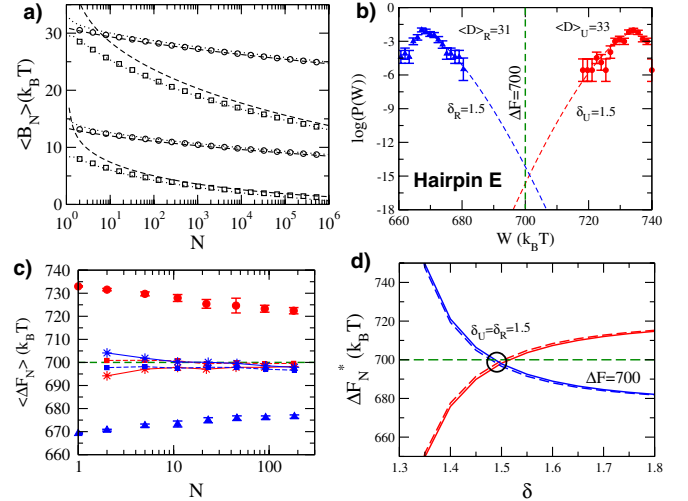

FIG. 3 (color online). Free-energy recovery for non-Gaussian WDs. (a) Analytical estimates of $\langle B_N \rangle$ as given in Eq. (5) (dashed lines) and Eq. (6) (dotted lines) with D_c replaced by μ, tested against simulated data for the WWD (symbols). From top to bottom, $(\delta, \Omega, \langle D \rangle) = (1.1, 1.83, 31.0)$, $(1.1, 1.68, 13.5)$, $(1.5, 6, 30.1)$, $(1.5, 4, 8.8)$. (b) WD for hairpin E pulled at 130 nm/s and fitted WD tails assuming $\delta = 1.5$ (dashed lines). The vertical line represents ΔF_{AR}. (c) $\langle \Delta F_N \rangle$ for U (circles) and R (triangles). Note the very slow convergence. The improved estimator ΔF_N^* obtained by correcting for the bias with Eq. (5) (squares) and Eq. (6) (stars) using $\delta = 1.5$ converges quickly to a value consistent with $\Delta F_{\text{AR}} = 700 k_B T$ (dashed horizontal line). (d) ΔF_N^* recovered from U and R as a function of δ for $N = 182$. Using Eq. (5) (dashed lines) or Eq. (6) (continuous lines) gives nearly the same result.

$$\langle B_N \rangle = \langle D \rangle N^{-z}, \qquad z = -\langle D \rangle^{-1} \log(\log 2 / \langle D \rangle), \quad (7)$$

interpolating between $\lambda = 1$ and $\lambda = 0$ (for which $\langle B_N \rangle = \langle D \rangle$), also fits fairly well the GWD data, as shown in Fig. 2, although less so than Eq. (6) for large Ω (see also Fig. S4 in [17]). A power-law fit of the bias in N was proposed in Ref. [14]. Figure 2(b) shows that the bias of all our experiments with mild dissipation is well described by Eq. (6) for a GWD.

Free-energy recovery method.—The fact that Eqs. (5) and (6) describe well the bias when $D_c = \mu$ suggests an improved estimator ΔF_N^* applicable to any problem involving the logarithm of an exponential average: (i) given N measurements W_i, compute ΔF_N and the histogram of the WD; (ii) estimate the shape parameters Ω, δ, α, q (and thus μ) by fitting the histogram tail to Eq. (2), taking for instance the maximum of the WD for W_c [24]; (iii) define $\Delta F_N^* = \Delta F_N - \langle B_N \rangle$ taking for $\langle B_N \rangle$ either Eq. (5) or (6) (depending on the value of λ) and setting $D_c = \mu$.

In the special case of the Jarzynski estimator, we must take into account the stronger constraint imposed by the CFT, which implies that $\langle e^{-W} \rangle$ is dominated by values of W near the maximum $W^\dagger$ of the reverse WD [25]. Sampling these very rare events is usually unfeasible, so there is no guarantee that the fit to the measured $p(W)$ will continue to hold near $W^\dagger$. Nevertheless, we argue that a distinction should be made between near-Gaussian and non-Gaussian tails. In the former case (i.e., when the tail can be fitted with $\delta \approx 2$), it is reasonable to assume that the fit will hold near $W^\dagger$; hence, ΔF_N^* should give a good estimate. The distance from a GWD can be self-consistently quantified *a posteriori* with the ratio $r = \sigma_W^2 / (2 \langle D \rangle)$ ($r = 1$ for a GWD) by taking $\langle D \rangle = \langle W \rangle - \Delta F_N^*$.

We return now to our DNA experiments, for which we can compare the unidirectional estimators ΔF_N, ΔF_v, ΔF_N^* separately for U and R with the bidirectional estimator ΔF_{AR}. Figure 1(d) shows a case with mild dissipation, for which the bias of ΔF_N is small and all estimators converge. Figure 1(e) shows an experiment with intermediate dissipation. In this case sampling the tail of p_U near the maximum of p_R (or vice versa) would require an unfeasible number of cycles. Nevertheless, both tails are well fitted with $\delta = 2$, $\alpha = 0$. [They are not perfect GWD, being slightly asymmetric. Using $\langle D \rangle = \langle W \rangle - \Delta F_{AR}$, we obtain $r = 0.67$ (U), 0.81 (R).] The curves in Fig. 1(e) represent $\Delta F_{AR} + \langle B_N \rangle$ with $\langle B_N \rangle$ given by Eqs. (5)–(7). We find that ΔF_N^* agrees with ΔF_{AR} within its statistical error for $N \gtrsim 5$, for all three expressions. This represents a significant improvement over the variance estimator, which has a bias $\Delta F_v - \Delta F_{AR} = 4.3 \pm 0.8$ (U), -2.9 ± 0.8 (R), and over the uncorrected Jarzynski estimator. For instance, we have $\Delta F_N - \Delta F_{AR} = 5.8 \pm 3.0$ (U), -6.6 ± 2.0 (R) for $N = 36$ and $\Delta F_N - \Delta F_{AR} = 5.1 \pm 0.6$ (R) for $N = 289$. Furthermore, the fitted WD satisfy fairly well the CFT (see

Sec. 4 in [17]), giving further confidence in the consistency of the method.

Finally, we consider an experiment where the tails are far from a GWD and the dissipation is large (Fig. 3). In this case the WDs are wide apart [Fig. 3(b)] and $\langle \Delta F_N \rangle$ converges very slowly [Fig. 3(c)]. Equation (2) fits reasonably well the WD tails for a fairly broad range of δ values (we take $\alpha = 0$ for simplicity [24]). The estimator ΔF_N^* is shown in Fig. 3(d): the pronounced dependence on δ and the discrepancy between U and R confirm that the predictive power of Eqs. (5) and (6) relies on accurately knowing δ (see Sec. 3 in [17] for other examples). Note also that Eqs. (5) and (6) are ill-defined as the exponent δ approaches 1 [18]. As δ decreases, the fitted tails satisfy less and less the CFT (see Sec. 4 in [17]), signaling that δ must increase further out in the tails. It is an open problem to generalize our analytical approach to an effective δ varying with W.

In summary, we obtained analytical expressions for the bias of the Jarzynski estimator and showed that they can be used to obtain improved unidirectional estimates of the free-energy of mechanical unfolding of DNA hairpins, provided the WD tail is described by a compressed exponential over a wide enough range of work values. These results are applicable to many unidirectional experiments and simulations and are relevant to other contexts involving sums of random exponentials.

*Corresponding author.
palassini@ub.edu

[1] C. Jarzynski, Eur. Phys. J. B **64**, 331 (2008).
[2] G. E. Crooks, Phys. Rev. E **60**, 2721 (1999).
[3] C. Jarzynski, Phys. Rev. Lett. **78**, 2690 (1997).
[4] J. Liphardt *et al.*, Science **296**, 1832 (2002).
[5] D. Collin *et al.*, Nature (London) **437**, 231 (2005).
[6] M. R. Shirts, E. Bair, G. Hooker, and V. S. Pande, Phys. Rev. Lett. **91**, 140601 (2003).
[7] G. Hummer and A. Szabo, Proc. Natl. Acad. Sci. U.S.A. **98**, 3658 (2001).
[8] N. C. Harris and C.-H. Kiang, Phys. Rev. E **79**, 041912 (2009).
[9] M. O. Jensen, S. Park, E. Tajkhorshid, and K. Schulten, Proc. Natl. Acad. Sci. U.S.A. **99**, 6731 (2002).
[10] A. N. Gupta *et al.*, Nature Phys. (in press).
[11] C. Jarzynski, Phys. Rev. E **73**, 046105 (2006).
[12] M. Romeo, V. Da Costa, and F. Bardou, Eur. Phys. J. B **32**, 513 (2003).
[13] D. M. Zuckerman and T. B. Woolf, Phys. Rev. Lett. **89**, 180602 (2002); J. Stat. Phys. **114**, 1303 (2004).
[14] J. Gore, F. Ritort, and C. Bustamante, Proc. Natl. Acad. Sci. U.S.A. **100**, 12 564 (2003).

[15] N. Forns *et al.*, Biophys. J. **100**, 1765 (2011).

[16] J. M. Huguet *et al.*, Proc. Natl. Acad. Sci. U.S.A. **107**, 15431 (2010).

[17] See Supplemental Material at http://link.aps.org/supplemental/10.1103/PhysRevLett.107.060601 for the appendices.

[18] The JE can also be satisfied with $\delta = 1$, $\Omega < 1$. This marginal case requires a separate analysis.

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

[20] B. Derrida, Phys. Rev. B **24**, 2613 (1981).

[21] J.-P. Bouchaud and M. Mézard, J. Phys. A **30**, 7997 (1997).

[22] The scaling limit of the REM has been studied rigorously in G. Ben Arous, L. V. Bogachev, and S. A. Molchanov, Probab. Theory Relat. Fields **132**, 579 (2005); A. Bovier, I. Kurkova, and M. Löwe, Ann. Probab. **30**, 605 (2002).

[23] J. Cook and B. Derrida, J. Stat. Phys. **63**, 505 (1991).

[24] Since $|W - W_c|^\alpha$ is slowly varying compared to the exponential, in fitting Eq. (2) to experiments one can safely set $\alpha = 0$.

[25] F. Ritort, J. Stat. Mech. (2004) P10016.