Phase transition in the Jarzynski estimator of free energy differences

Alberto Suárez,[*] Robert Silbey,[†] and Irwin Oppenheim[‡]

Department of Chemistry, Massachusetts Institute of Technology, 77 Massachusetts Avenue, Cambridge, Massachusetts 02139, USA

The transition between a regime in which thermodynamic relations apply only to ensembles of small systems coupled to a large environment and a regime in which they can be used to characterize individual macroscopic systems is analyzed in terms of the change in behavior of the Jarzynski estimator of equilibrium free energy differences from nonequilibrium work measurements. Given a fixed number of measurements, the Jarzynski estimator is unbiased for sufficiently small systems. In these systems the directionality of time is poorly defined and the configurations that dominate the empirical average, but which are in fact typical of the *reverse* process, are sufficiently well sampled. As the system size increases the arrow of time becomes better defined. The dominant atypical fluctuations become rare and eventually cannot be sampled with the limited resources that are available. Asymptotically, only typical work values are measured. The Jarzynski estimator becomes maximally biased and approaches the exponential of minus the average work, which is the result that is expected from standard macroscopic thermodynamics. In the proper scaling limit, this regime change has been recently described in terms of a phase transition in variants of the random energy model. In this paper this correspondence is further demonstrated in two examples of physical interest: the sudden compression of an ideal gas and adiabatic quasistatic volume changes in a dilute real gas.

I. INTRODUCTION

The laws of thermodynamics summarize empirical observations about the approximate and most probable behavior of macroscopic systems [1–4]. They are formulated under the assumption of limited resources in the measurements of the quantities involved. In the words of Gibbs, thermodynamics laws "express the laws of mechanics for [systems of a great number of particles] as they appear to beings who have not the fineness of perception to enable them to appreciate quantities of the order of magnitude of those which relate to single particles, and who cannot repeat their experiments often enough to obtain any but the most probable results" [1]. This statement summarizes the conditions under which a proper thermodynamic description for a single system can be made: (i) The system has a large number of degrees of freedom, (ii) the quantities of interest involve averages over space and time on scales that are large compared to the corresponding molecular scales, (iii) there are limitations in the time span of the measurements, and (iv) there are limitations in the number of measurements made. For instance, the elimination of condition (iii) gives rise to the objection formulated by Zermelo, based on the fact that any isolated mechanical system will come arbitrarily close to its initial state provided that a sufficiently long period of time elapses (Poincaré recurrences) [3–5]. If the system is ergodic, averages over phase space can be replaced by time averages and conditions (iii) and (iv) are equivalent.

Improvements in measurement devices and techniques; an increasing interest in smaller scale systems in biology, physics, and chemistry [6–11]; and developments in dynamical systems and chaos theory [12] have prompted numerous researchers to develop interpretations and extensions of thermodynamics applicable to systems with small numbers of degrees of freedom [8,13–17]. In such systems it is not possible to uphold the standard interpretation of the thermodynamic quantities and of their relations. In particular, the measurements are made on molecular scales and produce values that are dominated by fluctuations. Nonetheless, reproducible results are obtained by performing averages over many independent measurements of small systems at equilibrium (e.g., in contact with a heat bath). Therefore, a proper thermodynamic description is recovered if the ensemble method, which was originally devised as an operational construct to obtain results for single macroscopic systems, is given a literal interpretation: Thermodynamic quantities are identified with ensemble averages, which are experimentally realized as averages over independent measurements under equilibrium conditions. Note that the condition of equilibrium requires that the small system be in contact with some other system with a large number of degrees of freedom (e.g., a heat reservoir for the canonical ensemble) so that correct results are obtained from the assumption that the initial state in a particular realization of the process can be treated as an independent sample from the appropriate ensemble. Thus conditions (i)–(iv) are required for the whole and are essential for a proper thermodynamic description of the small system as well.

In this work we investigate the transition between the regime in which the thermodynamic description is valid for a single system and the regime in which thermodynamic quantities need to be understood as averages over realizations of a given process. To this end we analyze the behavior of the empirical estimates of equilibrium free energy differences from nonequilibrium work measurements by means of the Jarzynski equality. In contrast to standard thermodynamic relations, the ensemble average that appears in the Jarzynski equality is dominated by rare extreme fluctuations [18]. Consequently, to obtain accurate estimations of the free energy

---

[*]Permanent address: Department of Computer Science, Escuela Politécnica Superior, Universidad Autónoma de Madrid, 28049 Madrid, Spain; alberto.suarez@uam.es

[†]Deceased.

[‡]irwin@mit.edu

change one needs to perform repeated work measurements in the nonequilibrium process. The number of measurements needed to obtain meaningful estimates increases exponentially with the size of the system [18,19]. Therefore, the average that appears in the Jarzynski equality cannot be realized in the macroscopic regime, in which definite values of thermodynamic quantities can be ascribed to single systems rather than to collections of systems.

For limited resources, when the number of measurements is fixed, the Jarzynski estimator of free energy differences becomes biased as the size of the system increases. In the appropriate scaling limit, the appearance of the bias is abrupt and corresponds to a phase transition in variants of the random energy model (REM) [20–24]. The connection between the random energy model and the Jarzynski estimator of free energy differences was first made in Refs. [25–27]. In those works expressions of the free energy in the low-temperature (small-M limit, where M is the number of nonequilibrium work measurements) and the high-temperature phases (large-M limit) of the random energy model, including finite-M corrections, were derived for a parametric family of work distributions. In another recent work the convergence of Monte Carlo estimates in terms of the random energy model was made in connection with the sign problem [28].

In the current paper we further establish the correspondence between the Jarzynski estimator of free energy differences and variants of the random energy model for the sudden compression of an ideal gas and for adiabatic quasistatic volume changes in a dilute real gas. Even though this correspondence is explicitly established for only a few particular physical systems, it is expected to obtain in more general situations. The origin of the phase transitions in sums of random exponentials is the interplay between classic limit theorems for sums of random variables (e.g., the central limit theorem and the law of large numbers) and extreme value statistics [29,30]. A contribution of this paper is to highlight the importance of the phase transition in the Jarzynski estimator of free energy differences to signal the change between two distinct regimes wherein the ensemble and the single-system interpretations of thermodynamic quantities are applicable, respectively.

The article is organized as follows. Section II introduces the Jarzynski equality and related concepts that are necessary for the subsequent study. The use of this equality in the estimation of equilibrium free energy differences from nonequilibrium work values is described in Sec. III. Section IV analyzes the change in behavior of the Jarzynski estimator as a function of system size and of the number of measurements performed in three illustrative examples. Finally, Sec. V discusses how this change in behavior corresponds to the emergence of classical macroscopic thermodynamics and a well-defined arrow of time as the size of the system is increased, when the number of measurements performed is fixed. The technical details of the analysis of the Jarzynski estimator in terms of variants of the random energy model are deferred to the Appendices.

II. JARZYNSKI EQUALITY

Consider a system characterized by a Hamiltonian $H(\Gamma; \lambda)$, where Γ denotes a point in phase space. The parameter λ can be modified by external manipulation to perform or extract work from the system. If the system is coupled to other degrees of freedom, the changes in λ are assumed to affect only the Hamiltonian of the system of interest and not the terms that describe its interaction with the environment or the environment itself.

Assume that the parameter λ is modified in the interval $[0, \tau]$ according to a specified protocol $\{\lambda(t); 0 \leqslant t \leqslant \tau\}$ from $\lambda(0) = \lambda_0$ to $\lambda(\tau) = \lambda_1$. It is possible to show that the work associated with the transition between an initial configuration of the system sampled from an equilibrium distribution with Hamiltonian $H_0(\Gamma) \equiv H(\Gamma; \lambda_0)$ at temperature β^{-1} and a final configuration corresponding to the Hamiltonian $H_1(\Gamma) \equiv H(\Gamma; \lambda_1)$ is a random variable W that satisfies the equality [31,32]

$$\langle e^{-\beta W} \rangle = e^{-\beta \Delta F}, \quad \Delta F = F_1 - F_0, \text{ where} \quad (1)$$

$$F_i = -\beta^{-1} \ln \int_\Omega d\Gamma \, e^{-\beta H_i(\Gamma)}, \quad i = 0,1. \quad (2)$$

The average in Eq. (1) is performed over trajectories whose initial point in phase space is sampled from a canonical distribution

$$\frac{e^{-\beta H_0(\Gamma)}}{\int_\Omega d\Gamma \, e^{-\beta H_0(\Gamma)}}, \quad (3)$$

where the integral is over Ω, the region of phase space accessible to the system. This region is assumed to remain unchanged during the process. Note that the state in which the system finds itself after the manipulation has been completed need not be an equilibrium one. In general, F_1 is not the free energy of such a state. Therefore, the free energy difference that appears on the right-hand side of Eq. (1) need not coincide with the actual free energy change in the process that is carried out [32,33]. It is the difference between the free energies of the system in equilibrium at temperature β^{-1} in two states characterized by Hamiltonians H_0 and H_1, respectively. If the system is strongly coupled with an environment that acts as a heat reservoir, $H(\Gamma; \lambda)$ is the potential of mean force associated with the variables of the system of interest [32,34]. If the system is isolated or weakly coupled with its environment during the external manipulation, $H(\Gamma; \lambda)$ can be identified with the Hamiltonian of the isolated system [31].

III. ESTIMATION OF FREE ENERGY DIFFERENCES FROM NONEQUILIBRIUM WORK MEASUREMENTS

Since it was derived the Jarzynski equality has attracted a fair amount of interest because it allows the computation of equilibrium free energy differences from measurements of work in general nonequilibrium processes

$$\Delta F = -\beta^{-1} \ln \langle e^{-\beta W} \rangle, \quad (4)$$

where the angular brackets denote an average over the canonical ensemble at temperature β^{-1} with respect to the initial Hamiltoninan H_0. However, this equality differs in crucial aspects from standard thermodynamic relations. Because of the molecular nature of the systems analyzed, thermodynamic quantities (energy density, number density, temperature, pressure, work, etc.) are fluctuating variables. In macroscopic systems, the fluctuations are small. This

allows us to identify these random thermodynamic quantities with their typical values, which are deterministic and can be measured in single experiments. Repetitions of these experiments yield values that are indistinguishable, within the error of the macroscopic measurement. Therefore, for standard thermodynamic quantities, typical and average values are close to each other and can be used interchangeably to characterize the macroscopic state of the system under study.

In contrast, the average that appears in the Jarzynski equality needs to be interpreted as a true ensemble average [31]. The work measured in a particular realization of a nonequilibrium process depends on the initial microscopic configuration of the system and is therefore a fluctuating quantity. In each of these measurements, the system is prepared in an initial state sampled from the equilibrium distribution. One then carries out the intervention that gives rise to the nonequilibrium process. The work involved in this process is recorded. Finally, to extract equilibrium free energy differences from these measurements, the average of the exponential of minus these work values is computed. Because of the exponential form of the summed quantities, typical and average values are, in the general case, very different. Unlike for standard thermodynamic relations, plugging in the typical value of the work on the left-hand side of Eq. (1) does not fulfill the equality. The reason is that the average is dominated by extreme events whose probability of occurrence is very low. Therefore, a sufficiently large number of measurements is needed so that these rare events are well represented and the equality can be empirically realized. The number of measurements required to provide a meaningful estimate of the average increases exponentially with the size of the system [18,19,35]. Therefore, it is not practicable in macroscopic systems. Nonetheless, the regime in which the equality can be experimentally verified is accessible for sufficiently small systems.

To realize the average, repeated experiments under the same conditions are carried out. The values of work obtained in each of these experiments are recorded and used to estimate the average

$$\langle e^{-\beta W} \rangle \approx \langle e^{-\beta W} \rangle_M \equiv \frac{1}{M} \sum_{m=1}^{M} e^{-\beta W_m}. \tag{5}$$

By means of the Jarzynski equality, this Monte Carlo average can be used to estimate the change in free energy

$$\Delta F_M \equiv -\beta^{-1} \ln \langle e^{-\beta W} \rangle_M. \tag{6}$$

Since the particular realization of work values $\{W_m\}_{m=1}^{M}$ is random, the estimate ΔF_M is also a random variable. Our goal is to understand the properties of this random variable as a function of M, the number of nonequilibrium work measurements, and N, the size of the system.

Because of the exponential form of the quantity averaged, the role of extreme fluctuations is very important. In contrast to usual thermodynamic averages, such as $\langle W \rangle$, the average work, which are dominated by configurations that are typical of the initial (equilibrium) state of the system, the average $\langle e^{-\beta W} \rangle$ is actually dominated by rare configurations that are typical of the system at equilibrium at temperature β^{-1} with respect to the *final* Hamiltonian [18].

Of particular interest is the block [36,37] or quenched [38] average

$$\mathbb{E}[\Delta F_M] \equiv -\beta^{-1} \mathbb{E}[\ln \langle e^{-\beta W} \rangle_M], \tag{7}$$

where the expectation $\mathbb{E}[\cdot]$ is with respect to independent realizations of M measurements, each of which corresponds to an independent sample from an initial equilibrium canonical distribution at temperature β^{-1}. In Refs. [36,37] $\mathbb{E}[\Delta F_M]$ is referred to as the *finite-data average free energy difference*. This quantity is an estimator of ΔF, albeit a biased one, in general. The bias is the difference between the expected value of this estimator and the actual value of the free energy difference

$$B_M \equiv \mathbb{E}[\Delta F_M] - \Delta F. \tag{8}$$

Using the law of large numbers, it is possible to show that in the limit $M \to \infty$ the block average converges to the free energy change

$$\lim_{M \to \infty} \mathbb{E}[\Delta F_M] = \Delta F. \tag{9}$$

Therefore, the estimator $\mathbb{E}[\Delta F_M]$ is asymptotically unbiased

$$\lim_{M \to \infty} B_M = 0. \tag{10}$$

For a single experiment $M = 1$ the estimator equals the average work performed on the system

$$\mathbb{E}[\Delta F_1] = \langle W \rangle. \tag{11}$$

Following Ref. [19], we define the dissipated work in a given realization of the experiment as

$$W_{\text{dis}} = W - \Delta F. \tag{12}$$

Using Jensen's inequality, it is possible to show that $\mathbb{E}[\Delta F_1]$ is a positively biased estimator of ΔF,

$$\langle W \rangle = \mathbb{E}[\Delta F_1] \geqslant \Delta F. \tag{13}$$

In fact, the convergence of $\mathbb{E}[\Delta F_M]$ to the asymptotic limit ΔF is monotonic [37],

$$\langle W \rangle = \mathbb{E}[\Delta F_1] \geqslant \mathbb{E}[\Delta F_M] \geqslant \mathbb{E}[\Delta F_{M+1}] \geqslant \mathbb{E}[\Delta F_\infty] = \Delta F,$$
$$1 < M < \infty. \tag{14}$$

In terms of the bias

$$\langle W_{\text{dis}} \rangle = \langle W \rangle - \Delta F = B_1 \geqslant B_M \geqslant B_{M+1} \geqslant B_\infty = 0,$$
$$1 < M < \infty. \tag{15}$$

Therefore, the maximum bias corresponds to $M = 1$ and coincides with the average dissipated work

$$B_{\text{max}} = B_1 = \langle W_{\text{dis}} \rangle. \tag{16}$$

The main contribution of this research is to explicitly show in several paradigmatic cases that the finite-sample estimate of free energy differences from the Jarzynski equality for a particular system exhibits a change of behavior as M, the number of repetitions of the experiment, increases. Alternatively, for a fixed number of measurements, the regime change occurs as a function of N, the system size. For small systems the Jarzynski estimator is unbiased. In these systems the nonequilibrium work measurements are dominated by fluctuations. Configurations that are typical of the reversed

process are well sampled, which means that the arrow of time is poorly defined. As the system size increases, the probability of sampling these configurations becomes exponentially small so that they are not observed in practice. The Jarzynski estimator of the free energy change becomes biased and asymptotically approaches the value of the average work. The suppression of these fluctuations also leads to the emergence of a well-defined arrow of time in the system. In the limit $M \to \infty$, $N \to \infty$ with $\ln M/N \to$ const the regime change is akin to a phase transition that arises in simplified models of spin glasses, such as the random energy model in both its continuous [20,21] and discrete [22,24,39] versions.

IV. PHASE TRANSITION IN THE JARZYNSKI ESTIMATOR OF FREE ENERGY DIFFERENCES

We now proceed to analyze the behavior of the Jarzynski estimator in three important cases. The first one corresponds to processes in which the nonequilibrium work distribution is Gaussian [19]. This is a particular case of the class of work distributions analyzed in Ref. [27] with $\delta = 2$. The second case is a compression experiment for an ideal gas [33,35,40]. Finally, we consider adiabatic and quasistatic volume changes for a dilute classical gas of interacting particles [41].

The change of regime is best analyzed in terms of the normalized bias

$$\tilde{B}_M \equiv \frac{B_M}{B_{\max}} = \frac{\mathbb{E}[\Delta F_M] - \Delta F}{\langle W \rangle - \Delta F}, \quad 0 \leqslant \tilde{B}_M \leqslant 1, \quad (17)$$

where the maximum bias is the difference between the average work in the actual nonequilibrium process (which is not necessarily isothermal) and the free energy difference in the corresponding isothermal process

$$B_{\max} = \langle W \rangle - \Delta F. \quad (18)$$

As a function of M, the normalized bias is a monotonically decreasing quantity of M, which is bounded between 0 and 1,

$$1 = \tilde{B}_1 \geqslant \tilde{B}_M \geqslant \tilde{B}_{M+1} \geqslant \tilde{B}_\infty = 0, \quad 1 < M < \infty. \quad (19)$$

To simplify the derivations we assume $\beta = 1$. It is straightforward to reintroduce this parameter in the final expressions by noting that setting $\beta = 1$ is equivalent to measuring energies in units of $\beta^{-1} = k_B T$, where k_B is the Boltzmann constant and T is the initial temperature.

A. Gaussian work distribution

In this section we illustrate the connection between the Jarzynski free energy difference estimator and the random energy model for a Gaussian work distribution. The results presented in this section were first derived in Ref. [27]. That reference gives explicit expressions for the bias of the Jarzynski estimator in different regimes for a general class of work distributions, which includes the Gaussian as a particular case. It also considers finite-size corrections, which are ignored in our analysis.

Assume that in the sample estimate

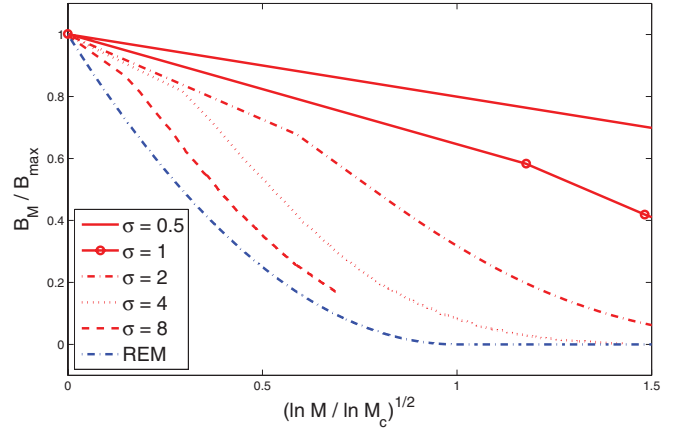$$\langle e^{-W} \rangle_M = \frac{1}{M} \sum_{m=1}^{M} e^{-W_m} \quad (20)$$

FIG. 1. (Color online) Approach of the normalized bias to the curve (24) as $M \to \infty$. The random energy model corresponds to the dash-dotted line.

the work values $\{W_m\}_{m=1}^{M}$ follow a normal distribution whose mean is $\langle W \rangle$ and whose variance is σ^2 [19]. In this case

$$\Delta F = -\ln\langle e^{-W} \rangle = \langle W \rangle - \tfrac{1}{2}\sigma^2. \quad (21)$$

Therefore, the maximum bias is

$$B_{\max} = \langle W \rangle - \Delta F = \tfrac{1}{2}\sigma^2. \quad (22)$$

This is an extensive quantity and scales with the size of the system. In the limit $\sigma \to \infty$, $M \to \infty$, and $\ln M/\sigma^2$ finite, the estimate of the free energy difference has an abrupt change of behavior (see Appendix A)

$$\mathbb{E}[\Delta F_M] = \begin{cases} \langle W \rangle - \tfrac{1}{2}\sigma^2, & M \geqslant \exp\{\tfrac{1}{2}\sigma^2\} \\ \langle W \rangle - \sqrt{2}\sigma\sqrt{\ln M} + \ln M, & M < \exp\{\tfrac{1}{2}\sigma^2\}. \end{cases} \quad (23)$$

These expressions correspond to those derived in Ref. [27] [Eq. (4) and the paragraph before this equation in that reference] for $\delta = 2$, $\Omega^2 = 2\sigma^2$, $N = M$, and $D_c = \sigma^2/2$. For a fixed value of σ the normalized bias is

$$\tilde{B}_M = \frac{B_M}{B_{\max}} = \begin{cases} 0, & M \geqslant M_c \\ (1 - \sqrt{\ln M/\ln M_c})^2, & M < M_c \end{cases} \quad (24)$$

with $M_c = \exp\{\tfrac{1}{2}\sigma^2\}$.

Figure 1 displays the dependence of the normalized bias [Eq. (24)] as a function of $\sqrt{\ln M/\ln M_c}$ with a fixed σ, for different values of σ. The curve corresponding to the asymptotic limit $\sigma \to \infty$ is plotted as a dash-dotted line. The remaining curves are averages over Monte Carlo simulations. For small numbers of measurements $M < M_c$, the Jarzynski estimator is biased. In this regime, the free energy difference estimator exhibits strong (of order 1) non-Gaussian fluctuations around its average (Theorem 1.6 from Ref. [23]). These fluctuations are driven by the Poisson process of the extremes of the random nonequilibrium work measurements. The range $M > M_c$ corresponds to a regime in which the estimate of the free energy change for sufficiently large σ is unbiased. The bias persists beyond this limit for small systems: When σ is small, the transition between regimes is more gradual. This is the region in which one expects to observe convergence to

TABLE I. Number of measurements needed to obtain an unbiased estimate of the free energy difference (M_c) and to reach a regime in which the fluctuations around this estimate are Gaussian (M'_c) as a function of $B_{\max}$, when the nonequilibrium work distribution is Gaussian.

$B_{\max}$	$k_B T$	$2k_B T$	$5k_B T$	$10k_B T$	$20k_B T$
M_c	3	8	149	22 027	485 165 196
M'_c	8	55	22 027	485 165 196	2.35×10^{17}

the Jarzynski limit in experiments. There is a second phase transition in the system: In the range $e^{\sigma^2/2} < M < e^{\sigma^2}$ the fluctuations can be expressed in terms of the Poisson process of extremes of the nonequilibrium work measurements. Beyond the threshold $M'_c = e^{\sigma^2}$, the central limit theorem holds and the fluctuations around the block average are approximately Gaussian (Theorem 1.5 from Ref. [23]). The graph presented corresponds to Fig. 2(a) in Ref. [27], with $\sigma^2 = \Omega^2/2$. The main difference is the square root in the abscissae, which does not modify the point at which the phase transition occurs in the REM limit or the qualitative picture.

Table I displays the number of measurements needed to obtain an unbiased estimate of the free energy change using the Jarzynski estimator (M_c) and to reach a regime in which the fluctuations around this estimate are Gaussian (M'_c) for the several values of $B_{\max}$ in the range explored in the experiments described in Ref. [27] (from $k_B T$ to $20k_B T$). The regime change can also be observed when the number of measurements is fixed and the size of the system, measured in terms of σ^2, increases. For a fixed M, the normalized bias is

$$\tilde{B}_M = \frac{B_M}{B_{\max}} = \begin{cases} 0, & \sigma \leqslant \sigma_c \\ (1 - \sigma_c/\sigma)^2, & \sigma > \sigma_c, \end{cases} \quad (25)$$

where $\sigma_c = \sqrt{2\ln M}$.

Figure 2 displays the curves that trace the dependence of the normalized bias [Eq. (25)] as a function of $\log_{10}(\sigma/\sigma_c)$ with M fixed, for different values of M. The curve corresponding to the random energy model ($M \to \infty$) is displayed as a dash-dotted line. In small systems $\sigma < \sigma_c$ the estimate of the free energy difference is unbiased. For sufficiently small systems $\sigma \to 0$, the bias scales as $B_M \sim B_{\max}/M$. Therefore, the bias is reduced by increasing M linearly. This is the region in which one expects convergence of the sample average to the Jarzynski equality limit. In the regime $\sigma > \sigma_c$, the empirical estimate is biased. As the system size increases the bias approaches its maximum value, which corresponds to measuring the typical value in most of the nonequilibrium work measurements, in agreement with the expected behavior of classical macroscopic systems. In this regime, linear increases in M do not lead to significant changes in the bias observed.

B. Ideal gas compression experiment

Consider an isolated system consisting of N noninteracting particles (ideal gas). The system is confined in the interval $[-L, L]$ in the X direction. The macroscopic state of the system is defined by the temperature and the value of an externally applied potential. The microscopic state of the system is characterized by n, the number of particles in region II ($0 < x \leqslant L$). Correspondingly, the number of number of particles in region I ($-L \leqslant x \leqslant 0$) is $N - n$.

Initially, the external potential is zero and the system is assumed to be in a homogeneous equilibrium state at temperature $\beta^{-1} = 1$. The probability of having n particles in region II in this state is

$$p(n; 0) = \frac{1}{2^N} \binom{N}{n}, \quad n = 0, 1, \ldots, N. \quad (26)$$

This is a binomial distribution with the parameters $(N, 1/2)$. It is peaked around the mean $n^*(0) = N/2$. Its standard deviation is $\sqrt{N}/2$. The corresponding free energy is

$$F_N(0) = -\ln\left[\frac{1}{2^N} \sum_{n=0}^{N} \binom{N}{n} \right] = 0. \quad (27)$$

Consider a compression experiment during which the system undergoes a sudden transition from the initial homogeneous equilibrium state to a state in which particles are more likely to be in region I because of the presence of a positive external potential in region II,

$$V(x; \epsilon) = \begin{cases} 0, & -L \leqslant x \leqslant 0 \quad \text{(region I)} \\ \epsilon, & 0 < x \leqslant L \quad \text{(region II)}. \end{cases} \quad (28)$$

The equilibrium distribution at temperature $\beta^{-1} = 1$, when the system is subject to this external potential is

$$p(n; \epsilon) = \frac{1}{(1 + e^{-\epsilon})^N} \binom{N}{n} e^{-n\epsilon}, \quad n = 0, 1, \ldots, N. \quad (29)$$

It is a binomial distribution with the parameters $[N, e^{-\epsilon}/(1 + e^{-\epsilon})]$. This distribution is peaked around its mean $n^*(\epsilon) = N/(1 + e^\epsilon)$. Its standard deviation is $\sqrt{N}e^{-\epsilon/2}/(1 + e^{-\epsilon})$. The corresponding free energy is

$$F_N(\epsilon) = -\ln\left[\frac{1}{2^N} \sum_{n=0}^{N} \binom{N}{n} e^{-n\epsilon} \right] = N \ln \frac{2}{1 + e^{-\epsilon}}. \quad (30)$$
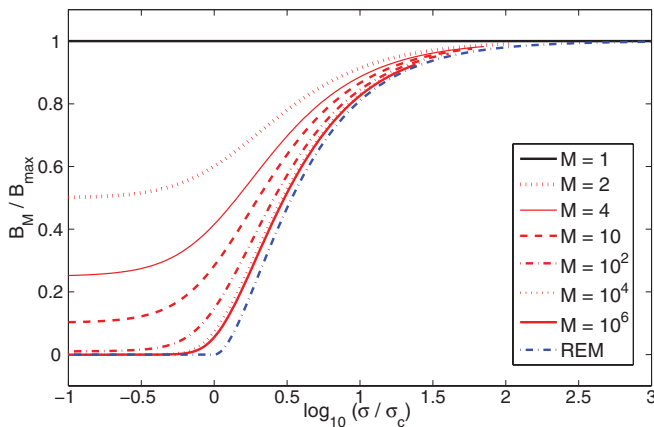

FIG. 2. (Color online) Approach of the normalized bias to the curve (25) as $M \to \infty$. The random energy model corresponds to the dash-dotted curve.

The change in free energy between the initial state, in which the system is in equilibrium at temperature $\beta^{-1} = 1$, in the absence of a external potential, and an equilibrium state at the same temperature with potential $V(x; \epsilon)$ is

$$\Delta F = F_N(\epsilon) - F_N(0) = N \ln \frac{2}{1 + e^{-\epsilon}}. \quad (31)$$

As noted by several authors, this is not the change of free energy between the equilibrium states corresponding to the initial and final values of the work parameter in the experiment that is being performed [32,33]. In particular, the configuration of the particles in the system at $t = 0^+$ is the same as in the initial state because there has not been any time to evolve. In most cases, this configuration is very atypical of the equilibrium state for the new constraints. In fact, the system does not have a well-defined temperature in this state. Nonetheless, using the equality

$$\Delta F = -\ln\langle e^{-W}\rangle, \quad (32)$$

it is possible to compute the change in free energy in an isothermal process [Eq. (31)] in terms of the external work performed on the system during the compression [40].

In practice, one needs to carry out a series of independent realizations of the experiment. In the mth realization, the configuration of the system is sampled from the equilibrium distribution $n_m \sim p(n; 0)$. The external work needed to perform the transition in this particular realization is $W_m = n_m \epsilon$. The average work over M realizations of the experiment is

$$\langle W\rangle_M = \frac{1}{M} \sum_{m=1}^{M} W_m = \frac{1}{M} \sum_{m=1}^{M} n_m \epsilon. \quad (33)$$

Since the $\{n_m\}_{m=1}^{M}$ are independent identically distributed random variables (IIDRVs), $\langle W\rangle_M$ is also a random variable

whose average is the thermodynamic work

$$\langle W\rangle = \mathbb{E}[\langle W\rangle_M] = \frac{1}{M} \sum_{m=1}^{M} \mathbb{E}[n_m]\epsilon = \mathbb{E}[n]\epsilon = \frac{1}{2}N\epsilon. \quad (34)$$

The average and the typical value of the work performed coincide.

Consider now the Jarzynski estimator of the free energy difference

$$\Delta F_M = -\ln\langle e^{-W}\rangle_M = -\ln\left(\frac{1}{M} \sum_{m=1}^{M} e^{-W_m}\right)$$

$$= -\ln\left(\frac{1}{M} \sum_{m=1}^{M} e^{-n_m \epsilon}\right). \quad (35)$$

This is also a random variable whose average is

$$\mathbb{E}[\Delta F_M] = -\mathbb{E}\left[\ln\left(\frac{1}{M} \sum_{m=1}^{M} e^{-n_m \epsilon}\right)\right]. \quad (36)$$

For $M = 1$ this average coincides with the average work

$$\mathbb{E}[\Delta F_1] = \langle W\rangle = \frac{1}{2}N\epsilon. \quad (37)$$

In the limit $M \to \infty$, it converges to the free energy change

$$\lim_{M \to \infty} \mathbb{E}[\Delta F_M] = \Delta F = N \ln \frac{2}{1 + e^{-\epsilon}}. \quad (38)$$

There is no closed-form expression for this average for other values of M. However, taking advantage of the correspondence between the ideal gas compression experiment and the discrete random energy model (see Appendix C), it is possible to show that there is a continuous but abrupt change in the block average

$$\mathbb{E}[\Delta F_M] = \begin{cases} N \ln \frac{2}{1+e^{-\epsilon}}, & \epsilon \leqslant \epsilon_c(\gamma) \\ N\left[\gamma \ln 2 + \frac{\epsilon}{2}\left(1 - \tanh \frac{\epsilon_c}{2}\right)\right] = N\left[\gamma \ln 2 + \frac{\epsilon}{1+e^{\epsilon_c}}\right], & \epsilon > \epsilon_c(\gamma) \end{cases} \quad (39)$$

in the limit $M \to \infty$, $N \to \infty$, with

$$\gamma = \frac{\ln M}{N \ln 2} \to \text{const}, \quad (40)$$

$$\epsilon_c(\gamma) = \begin{cases} \infty, & \gamma \geqslant 1 \\ \ln\left[1 - h_2^{-1}(1 - \gamma)\right] - \ln\left[h_2^{-1}(1 - \gamma)\right], & \gamma < 1. \end{cases} \quad (41)$$

The function $h_2^{-1}(y) \in [0, 1/2]$ is the inverse of the binary entropy

$$h_2(x) = -x \log_2 x - (1 - x)\log_2(1 - x). \quad (42)$$

In this limit the bias of $\mathbb{E}[\Delta F_M]$ as an estimator of ΔF is

$$B_M = \mathbb{E}[\Delta F_M] - \Delta F = \begin{cases} 0, & \epsilon \leqslant \epsilon_c(\gamma) \\ N\left[\gamma \ln 2 + \frac{\epsilon}{2}\left(1 - \tanh \frac{\epsilon_c}{2}\right) - \ln \frac{2}{1+e^{-\epsilon}}\right], & \epsilon > \epsilon_c(\gamma). \end{cases} \quad (43)$$

For a fixed value of ϵ the bias is maximum for $M = 1$ ($\gamma = 0$, $\epsilon_c = 0$),

$$B_{\max}(\epsilon) = B_1 = N\left[\frac{\epsilon}{2} - \ln \frac{2}{1+e^{-\epsilon}}\right] = N \ln \cosh \frac{\epsilon}{2}. \quad (44)$$

The normalized bias is

$$
\tilde{B}(\gamma,\epsilon) = \frac{B_M}{B_{\max}(\epsilon)} =
\begin{cases}
0, & \gamma \geqslant \gamma_c(\epsilon) \\
1 - \frac{\frac{\epsilon}{2}\tanh\frac{\epsilon_c}{2} - \gamma\ln2}{\ln\cosh\frac{\epsilon}{2}} = 1 - \frac{\epsilon - \epsilon_c}{2}\frac{\tanh\frac{\epsilon_c}{2}}{\ln\cosh\frac{\epsilon}{2}} - \frac{\ln\cosh\frac{\epsilon_c}{2}}{\ln\cosh\frac{\epsilon}{2}}, & \gamma < \gamma_c(\epsilon),
\end{cases}
\tag{45}
$$

where

$$
\gamma_c(\epsilon) = \epsilon_c^{-1}(\epsilon) = 1 - h_2\left(\frac{1}{1 + e^\epsilon}\right),
\tag{46}
$$

with $\gamma_c(0) = 0$ and $\lim_{\epsilon\to\infty}\gamma_c(\epsilon) = 1$.

The results of computer simulations of the ideal gas compression experiment at temperature $\beta^{-1} = 1$ are depicted in Fig. 3. The graphs display, for different values of ϵ, the dependence of the normalized bias as a function of system size, measured in terms of γ^{-1} with M fixed, for different values of M. The behavior observed is similar to the Gaussian case: The regime in which the Jarzynski equality can be realized (i.e., the Jarzynski free energy difference estimator is unbiased) corresponds to gases composed of a small number of particles. If the system size is increased, assuming that the number of measurements is kept fixed, a gradual change takes place to a regime in which the Jarzynski estimator becomes biased. Asymptotically, for large N, the estimator becomes maximally biased, which means that only typical work values are observed. Linear increases in M do not significantly reduce this bias. Therefore, in this regime, it is not possible to empirically realize the Jarzynski equality and measurements yield standard macroscopic thermodynamic values. The change in regime becomes more abrupt as ϵ, M, and N increase and is asymptotically well described by the phase transition that takes place in the discrete random energy model.

Table II displays the number of particles N_c at which the transition from a regime in which the Jarzynski estimator is unbiased $N < N_c$ to a regime in which the Jarzynski estimator is biased $N > N_c$, for several values of M and ϵ. Analyzing the results presented in this table, one can see that the critical system size is rather small and increases rather slowly (logarithmically) with M, the number of work measurements performed.

C. Adiabatic quasistatic volume change in a dilute gas

Consider a dilute gas of N interacting particles in d dimensions. Assume that this gas undergoes an adiabatic quasistatic volume change from V_0 to V_1. The work distribution in this process is

$$
p(W) = \frac{1}{|\alpha|\Gamma(K)}\left(\frac{W}{\alpha}\right)^{K-1} e^{-W/\alpha}\theta(\alpha W),
\tag{47}
$$

where $K = Nd/2$ and $\alpha = (V_0/V_1)^{2/d} - 1$ [41]. The Heaviside step function $\theta(\alpha W)$ guarantees that the work is positive for compression $\alpha > 0$ and negative for expansion $\alpha < 0$. The average work is $\langle W\rangle = (Nd/2)\alpha$. The typical value can be

identified with the mode of the distribution $W_{\text{typ}} = (Nd/2 - 1)\alpha$ for $Nd/2 > 1$. Note that for large systems ($N \to \infty$) typical and average work values are very similar.

Assuming a fixed number of measurements M, there is an abrupt change of behavior of the Jarzynski estimator as N, the size of the system, increases, for sufficiently large M and N (see Appendix B),

$$
\mathbb{E}[\Delta F_M] = -\mathbb{E}[\ln\langle e^{-W}\rangle_M] = \langle W\rangle + \ln M - \mathbb{E}\left[\ln\sum_{m=1}^{M} e^{-\alpha E_m}\right]
$$

$$
=
\begin{cases}
N\frac{d}{2}\ln(1 + \alpha), & N \leqslant N_c \\
N\frac{d}{2}\alpha[1 + x_l(N)] + \ln M, & N > N_c,
\end{cases}
\tag{48}
$$

where the system size at which the transition takes place is

$$
N_c = \frac{2\ln M}{d\left[\ln(1 + \alpha) - \frac{\alpha}{1+\alpha}\right]}
\tag{49}
$$

and $x_l(N)$ is the negative solution of

$$
x_l(N) - \ln[1 + x_l(N)] = \frac{2\ln M}{Nd}, \quad -1 < x_l(N) < 0.
\tag{50}
$$

This nonlinear equation has another solution at $x_u(N) > 0$. At the transition point

$$
x_l(N_c) = -\frac{\alpha}{1 + \alpha}.
\tag{51}
$$

The free energy difference for an *isothermal* volume change computed from the values of work measured in the *adiabatic* process is

$$
\Delta F = \lim_{M\to\infty}\mathbb{E}[\Delta F_M] = N\frac{d}{2}\ln(1 + \alpha) = N\ln\frac{V_0}{V_1}
\tag{52}
$$

in units of $\beta^{-1} = k_B T$. The bias in the Jarzynski estimator of the free energy difference is

$$
B_M =
\begin{cases}
0, & N \leqslant N_c \\
N\frac{d}{2}[\alpha(1 + x_l) - \ln(1 + \alpha)] + \ln M, & N > N_c.
\end{cases}
\tag{53}
$$

The maximum value of the bias corresponds to $M = 1$, which implies $x_l = x_u = 0$, $N_c = 0$, and

$$
B_{\max} = B_1 = \langle W\rangle - \Delta F = N\frac{d}{2}[\alpha - \ln(1 + \alpha)].
\tag{54}
$$

Therefore, the bias normalized by its maximum value is

$$
\tilde{B}_M = \frac{B_M}{B_{\max}} =
\begin{cases}
0, & N \leqslant N_c \\
1 + \frac{x_l(1 + \alpha) - \ln(1 + x_l)}{\alpha - \ln(1 + \alpha)}, & N > N_c.
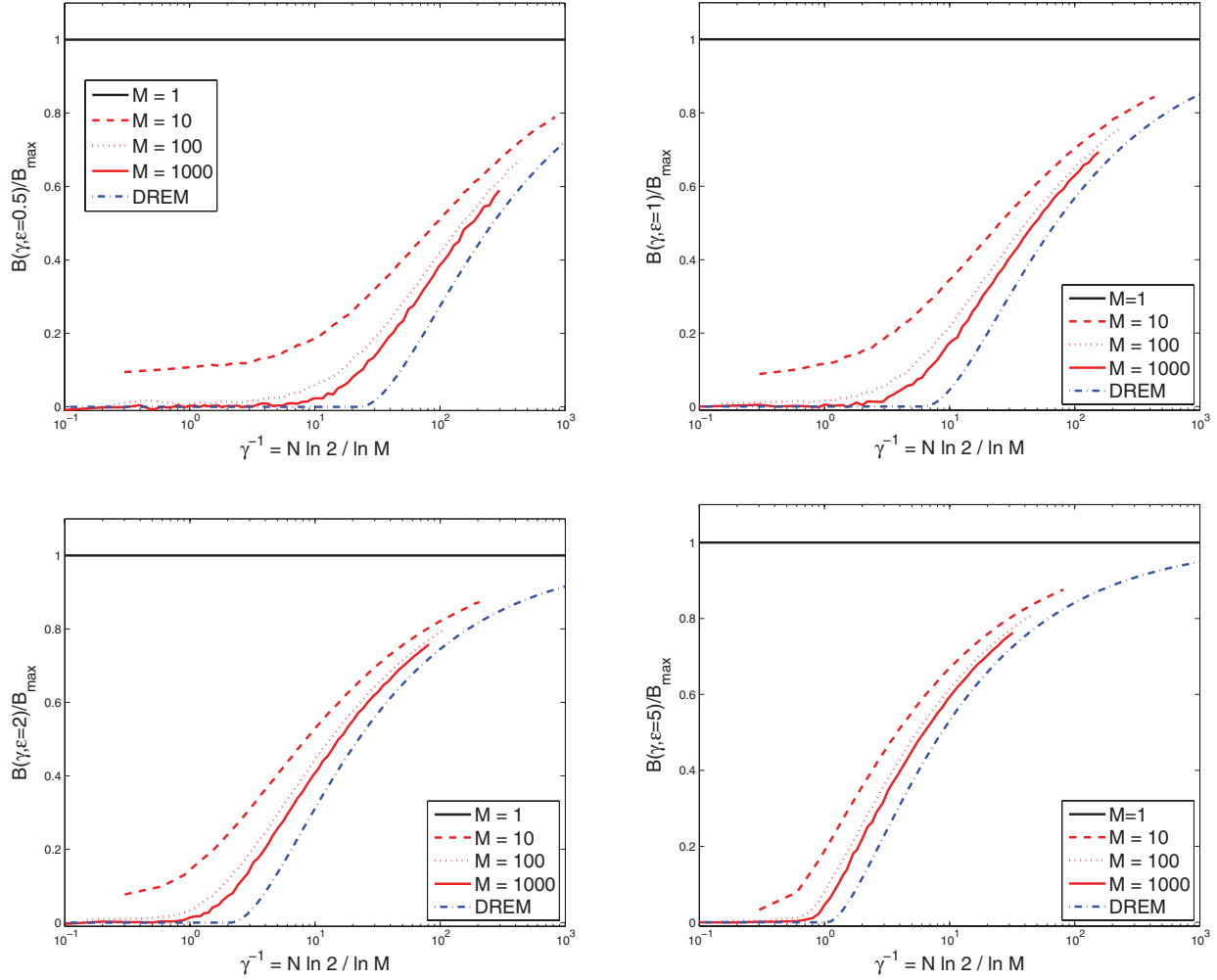\end{cases}
\tag{55}
$$

FIG. 3. (Color online) Dependence of the normalized bias on N for $\epsilon = 0.5$ (top left), $\epsilon = 1$ (top right), $\epsilon = 2$ (bottom left), and $\epsilon = 5$ (bottom right). The asymptotic value for the discrete random energy model (DREM) corresponds to the dash-dotted curve in the plots.

The results obtained in simulations of the adiabatic compression of a dilute real gas of N particles in $d = 3$ dimensions from a volume V_0 to a volume $V_1 = V_0/4$ are shown in Fig. 4. The plots display the normalized bias as a function of $\ln N / \ln N_c$ for different values of M, the number of measurements. In contrast with the Gaussian case, the curve for the random energy model changes, albeit slightly, as a function of M. The remaining curves in the plots correspond to averages over Monte Carlo simulations. The simulations cannot be performed for large values of N because

TABLE II. Number of particles N_c at which the transition between the two regimes of the Jarzynski estimator occurs for different values of M, the number of work measurements performed, and ϵ, the value of the external potential applied in the compression of an ideal gas.

M	$\epsilon = 0.5$	$\epsilon = 1$	$\epsilon = 2$	$\epsilon = 5$
10	76	21	8	4
100	152	42	15	8
1000	228	63	22	11
10 000	304	84	29	15

of numerical overflows. For fixed values of M, one observes a transition from a regime in which the bias is close to zero, for small systems, to a region in which the Jarzynski estimator is biased, when the system size is above a threshold N_c. The value of N_c has a logarithmic dependence on M, which means that the change in regime occurs for fairly small systems, even when the number of experiments is rather large ($N_c \approx 24$ for $M = 100\,000$). Beyond this threshold, the bias increases monotonically and asymptotically approaches its maximum. The regime in which the bias is maximal corresponds to the realm of classical thermodynamics. Within this regime the measured work can be identified with either the average $(Nd/2)\alpha$ or the typical value $(Nd/2 - 1)\alpha$. Typical and average values of work, as well as the Jarzynski estimator in a finite number of measurements, become indistinguishable as the number of gas particles increases.

The change between these two regimes, which is smooth for small M, becomes more abrupt as M increases. The variant of the REM model analyzed in Appendix B provides a good qualitative description of the transition that becomes more accurate as M, N, and the volume change become larger.
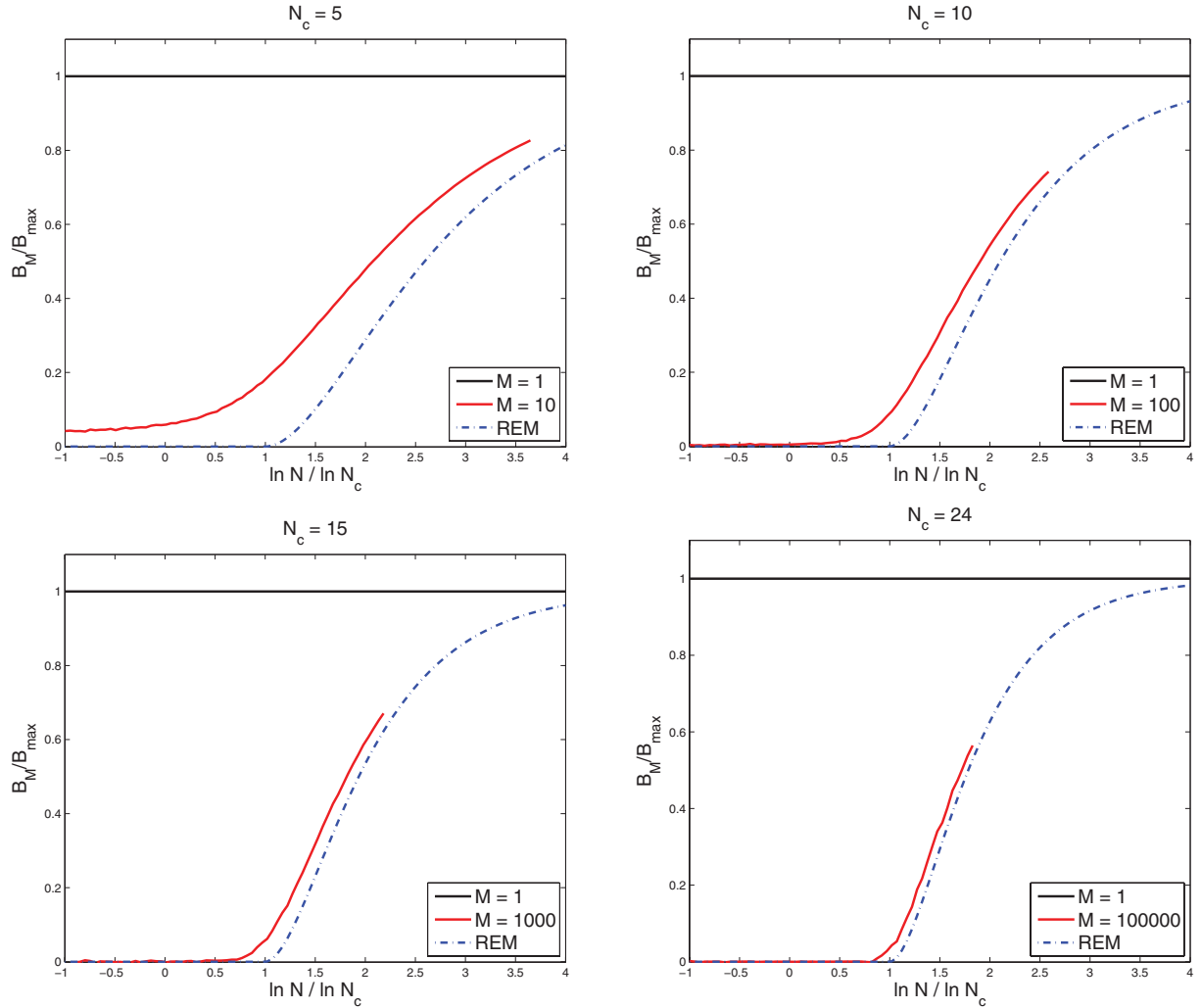
FIG. 4. (Color online) Dependence of the normalized bias on N for $M = 10$ (top left), $M = 100$ (top right), $M = 1000$ (bottom left), and $M = 100\,000$ (bottom right). The dash-dotted curve corresponds to the discrete random energy model and gives the limiting behavior as $M \to \infty$.

V. DISCUSSION AND CONCLUSIONS

There are two paradigms in which statistical mechanics has successfully bridged mechanical and thermodynamic laws in systems composed of large numbers of particles using a probabilistic description. On the one hand, individual macroscopic systems can be characterized by the values of thermodynamic quantities and their fluctuations. In these large systems the distribution of each of the thermodynamic quantities is sharply peaked around its mode. Therefore, the fluctuations are small and need to be probed by means of special experimental techniques (e.g., light scattering for mass density fluctuations) or in indirect measurements (e.g., using fluctuation-dissipation relations to analyze the relaxation of systems removed from equilibrium). The most probable value obtained in a single experiment agrees with the average over several experiments. On the other hand, in a small system coupled to a large number of degrees of freedom the standard thermodynamic description is valid for the system plus environment considered as a whole. However, measurements of reduced properties associated

with a small number of degrees of freedom are dominated by fluctuations. In consequence, thermodynamic relations cannot be used to describe individual measurements. Notwithstanding, they can be understood in terms of averages and fluctuations over repeated measurements. The coupling of the small system to the environment provides a mechanism for the reduced properties to approach their equilibrium values as given by the corresponding ensemble average [42–45]. Each measurement can be understood as sampling the initial state from the equilibrium distribution for the appropriate constraints (e.g., temperature, if the environment acts as a thermal reservoir). For standard thermodynamic quantities, these measurements over ensembles of small systems yield typical and average values that are close to each other. In contrast, typical values obtained in individual measurements of work in nonequilibrium processes do not fulfill the Jarzynski equality. In this equality, the average of the exponential of minus the work is computed by taking the mean of repeated independent measurements in systems prepared at equilibrium. By the law of large numbers, this Monte Carlo estimate

converges to the expected value in the limit of an infinite number of measurements. However, the sum is dominated by rare events that need to be sufficiently well sampled so that the Monte Carlo estimate is close to the asymptotic result. If the number of samples is below a certain threshold, the estimator is biased. It becomes unbiased only for sufficiently large sample sizes. In the appropriate scaling limit, the change of behavior corresponds to a phase transition in variants of the random energy model, a simplified model for spin glasses.

This phenomenon can be analyzed from an alternative point of view. Assuming that the resources available are limited and that the number of measurements M is fixed, the change of behavior in the estimator appears as the size of the system is increased. For small systems, the thermodynamic description needs to be understood in terms of ensemble averages, not of individual measurements. In this regime, the Jarzynski estimator is unbiased. It is equal to the exponential of minus the free energy difference in an isothermal evolution between the initial and final values of the work parameter, independently of other characteristics of the actual nonequilibrium process that takes place in the system. In particular, it does not depend on the final state of the system after the external manipulation is completed. The lack of sensitivity to the details of the manipulation is a striking reflection at the microscopic level of the Hamiltonian evolution of the system as a whole (including the degrees of freedom of the environment) as demonstrated in Ref. [32]. The mechanical character of the equality is also highlighted by the fact that the largest contributions to the average correspond to initial configurations of the system that are typical of the reverse process [18]. Since these configurations need to be sufficiently well sampled, this implies that, under these conditions, the arrow of time is only tenuously defined.

For large systems and fixed M, the Jarzynski estimator becomes biased. In fact, as the size of the system approaches infinity, the average of the exponential of minus the work values obtained in the nonequilibrium process approaches the exponential of minus the average work. Since the typical and the average are close to each other, a single experiment and an average over a number of experiments of the order of M yield results that are equivalent within measurement errors. Furthermore, this average depends on the details of the external manipulation. These features are in agreement with the standard thermodynamic description of an individual macroscopic system: The dominant configurations, which are rare for the initial state but typical of the final equilibrium state, are never sampled. The sample mean is dominated by typical configurations, which are characteristic of the forward process. Therefore, in this regime, the arrow of time becomes well defined.

The necessary condition to observe a transition in the Jarzynski estimator is that $\langle W \rangle$, the average work in the nonequilibrium process, be different from ΔF, the isothermal free energy change. If the nonequilibrium process is adiabatic, these quantities are different even when the external manipulation of the system is slow. This can be seen in Sec. IV C, which analyzes a quasistatic adiabatic volume change in a dilute gas. Another example is the adiabatic expansion of an ideal gas against a piston: The maximum bias $B_{\max} = \langle W \rangle - \Delta F$ is different from zero even when the velocity of the piston

approaches 0 (see Fig. 4 in Ref. [33]). The situation is different for isothermal processes. In this case $\langle W \rangle$ approaches ΔF in the limit of quasistatic driving. Therefore, in an isothermal process, one should expect a transition from a fast driving regime, in which the Jarzynski estimator of the free energy difference is biased, to a slow driving regime, in which the Jarzynski estimator is unbiased. Isothermal processes are currently under analysis.

The conclusions of this study are of course not new. The emergence of irreversibility and of a thermodynamic description in systems with many degrees of freedom is one of the central results of statistical mechanics [1–4,46]. The analysis carried out shows how, in the context of the Jarzynski equality, the emergence of this macroscopic picture can be understood in terms of a phase transition that appears in the proper scaling limit and when the measurement resources are limited as the size of the system increases.

APPENDIX A: THE RANDOM ENERGY MODEL

The random energy model was introduced in Refs. [20,21] as a simplified model for spin glasses that captures many salient properties of these types of disordered systems. In the random energy model, the system has $M = 2^K$ energy levels $\{E_i\}_{i=1}^{M}$. These energy levels are independent identically distributed random variables sampled from a normal distribution

$$p(E) = \frac{1}{\sqrt{\pi K}} e^{-E^2/K}. \tag{A1}$$

The canonical partition function for a particular system (i.e., for a particular realization of the M energy levels) at temperature β^{-1} is

$$Z_M(\beta) \equiv \sum_{i=1}^{2^K} e^{-\beta E_i}. \tag{A2}$$

In the limit of large $K \to \infty$, the system undergoes a second-order phase transition

$$\lim_{K \to \infty} \frac{1}{K} \mathbb{E}[\ln Z_M(\beta)] = \begin{cases} \frac{\beta^2}{4} + \ln 2, & \beta \leqslant \beta_c \\ \beta \sqrt{\ln 2}, & \beta > \beta_c, \end{cases} \tag{A3}$$

where $\beta_c = 2\sqrt{\ln 2}$.

To make the connection between the Gaussian REM and the empirical estimation of the free energy difference in a process in which the work distribution is Gaussian, $W_m \sim N(\langle W \rangle, \sigma^2)$, we make the change of variables

$$W_m = \langle W \rangle + \sqrt{\frac{2 \ln 2}{\ln M}} \sigma E_m,$$
$$E_m \sim N\left(0, \frac{\ln M}{2 \ln 2}\right), \tag{A4}$$

where we have used the fact that $K = \ln M/\ln 2$. In terms of these new variables

$$\langle e^{-W} \rangle_M = \frac{1}{M} \sum_{m=1}^{M} e^{-W_m} = \frac{1}{M} e^{-\langle W \rangle} \sum_{m=1}^{M} \exp\left(-\sqrt{\frac{2\ln 2}{\ln M}} \sigma E_m\right). \tag{A5}$$

Identifying $\sqrt{2\ln 2/\ln M}\,\sigma$ with β in the REM, we obtain

$$\mathbb{E}[\Delta F_M] = -\mathbb{E}[\ln\langle e^{-W} \rangle_M] = \langle W \rangle + \ln M$$

$$- \mathbb{E}\left\{ \ln\left[\sum_{m=1}^{M} \exp\left(-\sqrt{\frac{2\ln 2}{\ln M}} \sigma E_m\right) \right] \right\}$$

$$= \begin{cases} \langle W \rangle - \frac{1}{2}\sigma^2, & M \geqslant \exp\{\frac{1}{2}\sigma^2\} \\ \langle W \rangle - \sqrt{2}\sigma\sqrt{\ln M} + \ln M, & M < \exp\{\frac{1}{2}\sigma^2\} \end{cases} \tag{A6}$$

for $M \to \infty$ and $\sigma^2 \to \infty$ with $\ln M/\sigma^2 \to \text{const}$.

APPENDIX B: THE RANDOM ENERGY MODEL WITH ENERGY LEVELS THAT FOLLOW A GAMMA DISTRIBUTION

Consider a system with M random energy levels $\{E_i\}_{i=1}^{M}$ independently sampled from a gamma distribution

$$p(E) = \frac{(E+K)^{K-1}}{\Gamma(K)} e^{-(E+K)}, \quad E \in [-K,\infty). \tag{B1}$$

Define the parameter $\xi = \frac{\ln M}{K \ln 2}$. In terms of this parameter $M = 2^{\xi K}$. It is possible to derive an expression for the entropy in the limit $K \to \infty$ by analyzing the behavior of $\mathcal{N}(\epsilon, \epsilon + \delta)$, the number of energy levels in an interval $\mathcal{I}(\epsilon; \delta) = [K\epsilon, K(\epsilon + \delta)]$, with $\epsilon \geqslant -1$, $\delta > 0$ [38]. Since this count depends on the realization of the system, $\mathcal{N}(\epsilon, \epsilon + \delta)$ is a binomial random variable whose first two moments are

$$\mathbb{E}[\mathcal{N}(\epsilon, \epsilon + \delta)] = 2^{\xi K} \mathcal{P}_{\mathcal{I}}(\epsilon; \delta), \tag{B2}$$

$$\text{Var}[\mathcal{N}(\epsilon, \epsilon + \delta)] = 2^{\xi K} \mathcal{P}_{\mathcal{I}}(\epsilon; \delta)[1 - \mathcal{P}_{\mathcal{I}}(\epsilon; \delta)], \tag{B3}$$

where

$$\mathcal{P}_{\mathcal{I}}(\epsilon; \delta) = \frac{K^K}{\Gamma(K)} \int_{\epsilon}^{\epsilon+\delta} (x+1)^{K-1} e^{-K(x+1)} dx \tag{B4}$$

is the probability of an individual energy level to be in the interval $\mathcal{I}(\epsilon; \delta)$. As $K \to \infty$ these moments can be approximated to leading exponential order as

$$\mathbb{E}[\mathcal{N}(\epsilon, \epsilon + \delta)] \doteq \exp\left\{K \max_{[\epsilon,\epsilon+\delta]} s_a(x)\right\}, \tag{B5}$$

$$\frac{\text{Var}\{\mathcal{N}(\epsilon, \epsilon + \delta)\}}{\{\mathbb{E}[\mathcal{N}(\epsilon, \epsilon + \delta)]\}^2} \doteq \exp\left\{-K \max_{[\epsilon,\epsilon+\delta]} s_a(x)\right\}, \tag{B6}$$

with

$$s_a(x) = \xi \ln 2 + \ln(1+x) - x. \tag{B7}$$

Note that $s_a(x) \geqslant 0$ for $x_l \leqslant x \leqslant x_u$, where $-1 \leqslant x_l \leqslant 0 \leqslant x_u$ fulfill $s_a(x_l) = s_a(x_u) = 0$. The limiting behavior of this equation when $\xi \to 0$ is

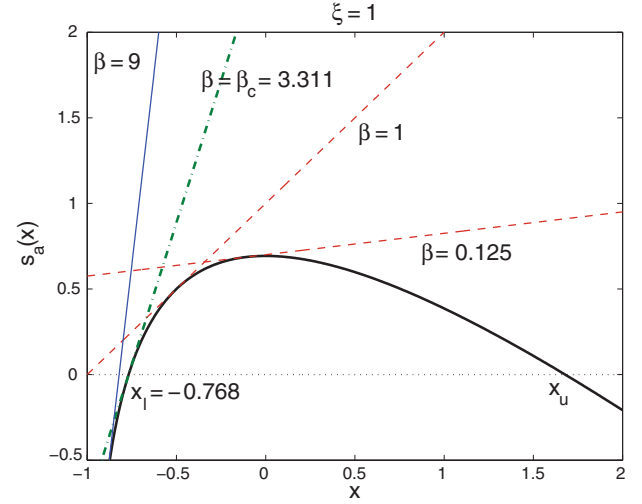$$s_a(x) = \xi \ln 2 - x^2/2, \quad x_l \leqslant x \leqslant x_u, \tag{B8}$$

FIG. 5. (Color online) Transition between the low- and high-temperature regimes in the random energy model with energy levels sampled from a gamma distribution (see the text for details).

where $x_l \approx -\sqrt{2\xi \ln 2}$ and $x_u \approx \sqrt{2\xi \ln 2}$. In this limit the results are similar to the Gaussian REM. In the opposite limit, when $\xi \to \infty$ is

$$s_a(x) = \xi \ln 2 + \ln(1+x), \quad x \approx x_l, \tag{B9}$$

where $x_l \approx -1 + 2^{-\xi}$. The entropy function is defined as

$$s(\epsilon) = \begin{cases} s_a(x) = \xi \ln 2 + \ln(1+x) - x, & x_l \leqslant x \leqslant x_u \\ -\infty, & x < x_l, x > x_u. \end{cases} \tag{B10}$$

It can be shown that for any pair ϵ, δ, with probability one,

$$\lim_{K \to \infty} \frac{1}{K} \ln \mathcal{N}(\epsilon, \epsilon + \delta) = \sup_{[\epsilon,\epsilon+\delta]} s(x). \tag{B11}$$

The canonical partition function for a particular realization of the M energy levels at temperature β^{-1} is

$$Z_M(\beta) = \sum_{i=1}^{2^{\xi K}} e^{-\beta E_i}. \tag{B12}$$

In the limit $K \to \infty$,

$$Z_M(\beta) \doteq \int_{x_l}^{x_u} \exp\{K[s_a(x) - \beta x]\} dx$$

$$\doteq \exp\left[K \max_{x \in [x_l, x_u]} (s_a(x) - \beta x)\right] \tag{B13}$$

to exponential accuracy. Depending on the temperature, the maximum is either in the interval $(x_l, 0)$ (high temperature) or at x_l (low temperature),

$$\arg\max_{x \in [x_l, x_u]}[s_a(x) - \beta x] = \begin{cases} -\frac{\beta}{1+\beta}, & \beta \leqslant \beta_c \\ x_l, & \beta > \beta_c, \end{cases} \tag{B14}$$

where $\beta_c = -x_l/(1 + x_l)$. A graphical construction that illustrates this transition is presented in Fig. 5 for $\xi = 1$. The curves displayed are $s_a(x)$ and straight lines with slope β that are tangent to $s_a(x)$ at the points that are local maxima of $s_a(x) - \beta x$. For high temperatures $\beta \leqslant \beta_c$ the local maximum

is within the interval $[x_l, x_u]$ and is therefore the solution of Eq. (B14). At low temperatures $\beta > \beta_c$, the solution to Eq. (B14) is x_l, which in this regime is a global maximum, but not a local one. Using these relations it can be shown that, in this limit, the system undergoes a second-order phase transition

$$\lim_{K \to \infty} \frac{1}{K} \mathbb{E}[\ln Z_M(\beta)] = \begin{cases} \xi \ln 2 + \beta - \ln(1 + \beta), & \beta \leqslant \beta_c \\ -\beta x_l = \beta \beta_c/(1 + \beta_c), & \beta > \beta_c. \end{cases}$$
(B15)

Continuity at β_c implies that

$$\ln(1 + \beta_c) - \frac{\beta_c}{1 + \beta_c} = \xi \ln 2. \tag{B16}$$

Consider now a sample of $M = 2^{\xi K}$ work values $\{W_m\}_{m=1}^M$ from the distribution

$$p(W) = \frac{1}{|\alpha|\Gamma(K)} \left(\frac{W}{\alpha}\right)^{K-1} e^{-W/\alpha} \theta(\alpha W), \tag{B17}$$

with $K = Nd/2$. The connection with the variant of REM analyzed in this section is achieved by making the change of variable

$$W_m = \langle W \rangle + \alpha E_m, \quad m = 1, \ldots, M, \tag{B18}$$

where $\langle W \rangle = N\frac{d}{2}\alpha$ is the average work and E_m are IIDRVs sampled from the distribution (B1). In terms of these new variables

$$\langle e^{-W} \rangle_M = \frac{1}{M} \sum_{m=1}^M e^{-W_m} = \frac{1}{M} e^{-\langle W \rangle} \sum_{m=1}^M e^{-\alpha E_m}. \tag{B19}$$

Identifying α with β in the random energy model, we conclude that there is an abrupt change of behavior of the Jarzynski estimator of the free energy differences for $M \to \infty, N \to \infty$ with $\ln M/N \to$ const, as a function of $\xi = \frac{2\ln M}{Nd \ln 2}$,

$$\mathbb{E}[\Delta F_M] = -\mathbb{E}[\ln\langle e^{-W} \rangle_M] = \langle W \rangle + \ln M$$
$$- \mathbb{E}\left[\ln \sum_{m=1}^M e^{-\alpha E_m}\right]$$
$$= \begin{cases} N\frac{d}{2}\ln(1 + \alpha), & \xi \geqslant \xi_c \\ N\frac{d}{2}\alpha[1 + x_l(\xi)] + \ln M, & \xi < \xi_c, \end{cases} \tag{B20}$$

where

$$\xi_c = \frac{\ln(1 + \alpha) - \frac{\alpha}{1+\alpha}}{\ln 2} \tag{B21}$$

and $x_l(\xi)$ is the negative solution of the nonlinear equation

$$x_l(\xi) - \ln[1 + x_l(\xi)] = \xi \ln 2, \quad -1 < x_l(\xi) < 0. \tag{B22}$$

At the transition point $\xi = \xi_c$,

$$x_l(\xi_c) = -\frac{\alpha}{1 + \alpha}. \tag{B23}$$

For a fixed number of particles N, the transition takes place when the number of measurements is above the threshold

$$M_c = \exp\left[N\frac{d}{2}\left(\ln(1 + \alpha) - \frac{\alpha}{1 + \alpha}\right)\right]. \tag{B24}$$

Alternatively, for fixed number of experiments M, the transition occurs when the number of particles in the gas

is below

$$N_c = \frac{2\ln M}{d\left[\ln(1 + \alpha) - \frac{\alpha}{1+\alpha}\right]}. \tag{B25}$$

APPENDIX C: THE DISCRETE RANDOM ENERGY MODEL

Consider a system with $M = 2^K$ energy levels $\{E_i\}_{i=1}^M$. These energy levels are independent identically distributed random variables sampled from a binomial distribution

$$p(E) = \frac{1}{2^N}\binom{N}{\frac{1}{2}N + E}, \quad E = -\frac{N}{2}, -\frac{N}{2} + 1, \ldots, \frac{N}{2}. \tag{C1}$$

The canonical partition function at temperature β^{-1} is

$$Z_M(\beta) = \sum_{i=1}^{2^K} e^{-\beta E_i}. \tag{C2}$$

In the limit

$$M \to \infty \quad N \to \infty, \quad \gamma = \frac{K}{N} = \frac{\ln M}{N \ln 2} \to \text{const} \tag{C3}$$

the system undergoes a second-order phase transition [24]

$$\mathbb{E}[\ln Z_M(\beta)] = \mathbb{E}\left[\ln\left(\sum_{i=1}^{2^K} e^{-\beta E_i}\right)\right]$$
$$= \begin{cases} K\ln 2 + N\ln\cosh\frac{\beta}{2}, & \beta \leqslant \beta_c \\ \frac{1}{2}\beta N \tanh\frac{\beta_c}{2}, & \beta > \beta_c, \end{cases} \tag{C4}$$

where

$$\beta_c = \begin{cases} \infty, & \gamma \geqslant 1 \\ \ln[1 - h_2^{-1}(1 - \gamma)] - \ln[h_2^{-1}(1 - \gamma)], & \gamma < 1 \end{cases} \tag{C5}$$

and the function $h_2^{-1}(y) \in [0, 1/2]$ is the inverse of the binary entropy

$$h_2(x) = -x\log_2 x - (1 - x)\log_2(1 - x). \tag{C6}$$

To make the connection with the ideal gas compression experiment in the calculation of

$$\mathbb{E}[\Delta F_M] = -\mathbb{E}[\ln\langle e^{-W} \rangle_M] = -\mathbb{E}\left[\ln\left(\frac{1}{M}\sum_{m=1}^M e^{-n_m \epsilon}\right)\right], \tag{C7}$$

where $n_m \in \{0, 1, \ldots, N\}$ follows a binomial distribution of parameters N, $p(0) = 1/2$, we make the change of variable

$$n_m = \frac{1}{2}N + E_m, \quad E_m \in \left\{-\frac{N}{2}, -\frac{N}{2} + 1, \ldots, \frac{N}{2}\right\}. \tag{C8}$$

Using these new random variables

$$
\mathbb{E}\left[\Delta F_M\right] = -\mathbb{E}\left[\ln\left(e^{-(N\epsilon/2)}\frac{1}{M}\sum_{m=1}^{M}e^{-E_m\epsilon}\right)\right]
$$

$$
= N\frac{\epsilon}{2} + \ln M - \mathbb{E}\left[\ln\left(\sum_{m=1}^{M}e^{-E_m\epsilon}\right)\right]. \quad \text{(C9)}
$$

Finally, identifying ϵ with β in the discrete random energy model and $\epsilon_c = \beta_c$,

$$
\mathbb{E}[\Delta F_M] = \begin{cases} N\left[\frac{\epsilon}{2} - \ln\cosh\frac{\epsilon}{2}\right] = N\ln\frac{2}{1+e^{-\epsilon}}, & \epsilon \leqslant \epsilon_c \\ N\left[\gamma\ln 2 + \frac{\epsilon}{2}\left(1 - \tanh\frac{\epsilon_c}{2}\right)\right], & \epsilon > \epsilon_c, \end{cases}
$$

$$(C10)$$

where we have used that fact that $K = \ln M/\ln 2$ and $\gamma = K/N = \frac{\ln M}{N\ln 2}$.

[1] J. W. Gibbs, *Elementary Principles in Statistical Mechanics: Developed with Especial Reference to the Rational Foundation of Thermodynamics* (Dover, New York, 1902).

[2] A. Khinchin, *Mathematical Foundations of Statistical Mechanics* (Dover, New York, 1949).

[3] R. Zwanzig, *Nonequilibrium Statistical Mechanics* (Oxford University Press, Oxford, 2001).

[4] P. Castiglione, M. Falcioni, A. Lesne, and A. Vulpiani, *Chaos and Coarse Graining in Statistical Mechanics* (Cambridge University Press, Cambridge, 2008).

[5] E. Zermelo, Ann. Phys. (Leipzig) **293**, 485 (1896).

[6] J. Liphardt, S. Dumont, S. B. Smith, I. Tinoco, and C. Bustamante, Science **296**, 1832 (2002).

[7] E. Barkai, Y. Jung, and R. Silbey, Annu. Rev. Phys. Chem. **55**, 457 (2004).

[8] C. Bustamante, J. Liphardt, and F. Ritort, Phys. Today **58**(7), 43 (2005).

[9] A. Ruschhaupt, J. G. Muga, and M. G. Raizen, J. Phys. B **39**, 3833 (2006).

[10] J. J. Thorn, E. A. Schoene, T. Li, and D. A. Steck, Phys. Rev. Lett. **100**, 240407 (2008).

[11] F. Ritort, *Advances in Chemical Physics* (Wiley, New York, 2008), pp. 31–123.

[12] P. Gaspard, *Chaos, Scattering, and Statistical Mechanics* (Cambridge University Press, Cambridge, 1998).

[13] I. Oppenheim and P. Mazur, Physica **23**, 197 (1957).

[14] P. Mazur and I. Oppenheim, Physica **23**, 216 (1957).

[15] J. L. Lebowitz and J. K. Percus, Phys. Rev. **124**, 1673 (1961).

[16] T. Hill, *Thermodynamics of Small Systems* (Benjamin, New York, 1963).

[17] U. Seifert, Eur. Phys. J. B **64**, 423 (2008).

[18] C. Jarzynski, Phys. Rev. E **73**, 046105 (2006).

[19] J. Gore, F. Ritort, and C. Bustamante, Proc. Natl. Acad. Sci. USA **100**, 12564 (2003).

[20] B. Derrida, Phys. Rev. Lett. **45**, 79 (1980).

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

[22] C. Moukarzel and N. Parga, Physica A **177**, 24 (1991).

[23] A. Bovier, I. Kurkova, and M. Löwe, Ann. Probab. **30**, 605 (2002).

[24] K. Ogure and Y. Kabashima, J. Stat. Mech. Theor. Exp. (2009) P03010.

[25] M. Palassini and F. Ritort, in *Book of Abstracts XXV Congreso de Física Estadística*, FisEs08 (Universidad de Salamanca, Salamanca, Spain, 2008), p. 40.

[26] M. Palassini and F. Ritort, [http://meetings.aps.org/link/BAPS.2008.MAR.V17.12] .

[27] M. Palassini and F. Ritort, Phys. Rev. Lett. **107**, 060601 (2011).

[28] G. Düring and J. Kurchan, Europhys. Lett. **92**, 50004 (2010).

[29] G. Ben Arous, L. V. Bogachev, and S. A. Molchanov, Probab. Theory Relat. Fields **132**, 579 (2005).

[30] M. Clusel and E. Bertin, Int. J. Mod. Phys. B **22**, 3311 (2008).

[31] C. Jarzynski, Phys. Rev. Lett. **78**, 2690 (1997).

[32] C. Jarzynski, J. Stat. Mech. Theor. Exp. (2004) P09005.

[33] B. Palmieri and D. Ronis, Phys. Rev. E **75**, 011133 (2007).

[34] J. G. Kirkwood, J. Chem. Phys. **3**, 300 (1935).

[35] R. C. Lua and A. Y. Grosberg, J. Phys. Chem. B **109**, 6805 (2005).

[36] D. M. Zuckerman and T. B. Woolf, Phys. Rev. Lett. **89**, 180602 (2002).

[37] D. M. Zuckerman and T. B. Woolf, J. Stat. Phys. **114**, 1303 (2004).

[38] M. Mézard and A. Montanari, *Information, Physics, and Computation* (Oxford University Press, Oxford, 2009).

[39] C. Moukarzel and N. Parga, Physica A **185**, 305 (1992).

[40] S. Presse and R. Silbey, J. Chem. Phys. **124**, 054117 (2006).

[41] G. E. Crooks and C. Jarzynski, Phys. Rev. E **75**, 021116 (2007).

[42] I. Prigogine and F. Henin, J. Math. Phys. **1**, 349 (1960).

[43] F. Henin, P. Résibois, and F. Andrews, J. Math. Phys. **2**, 68 (1961).

[44] D. Ronis and I. Oppenheim, Physica A **86**, 475 (1977).

[45] I. Oppenheim, Prog. Theor. Phys. Suppl. **99**, 369 (1989).

[46] L. Boltzmann, Sitzungsber. Kaiser. Akad. Wiss. **66**, 275 (1872) [in *Kinetic Theory of Gases: An Anthology of Classic Papers With Historical Commentary,* edited by Nancy S. Hall (Imperial College Press, London, 2003)], pp. 262–349.