Journal of Statistical Mechanics: Theory and Experiment
An IOP and SISSA journal

Work and heat fluctuations in two-state systems: a trajectory thermodynamics formalism

F Ritort

Departament de Fisica Fonamental, Facultat de Física, Universitat de Barcelona, Diagonal 647, 08028 Barcelona, Spain
E-mail: ritort@ffn.ub.es
URL: http://www.ffn.ub.es/ritort

Received 7 July 2004
Accepted 18 October 2004
Published 29 October 2004
Online at stacks.iop.org/JSTAT/2004/P10016
doi:10.1088/1742-5468/2004/10/P10016

Abstract. Two-state models provide phenomenological descriptions of many different systems, ranging from physics to chemistry and biology. We investigate work fluctuations in an ensemble of two-state systems driven out of equilibrium under the action of an external perturbation. We calculate the probability density $P_N(W)$ that work equal to W is exerted upon the system (of size N) along a given non-equilibrium trajectory and introduce a trajectory thermodynamics formalism to quantify work fluctuations in the large-N limit. We then define a *trajectory entropy* $S_N(W)$ that counts the number of non-equilibrium trajectories $P_N(W) = \exp(S_N(W)/k_{\rm B}T)$ with work equal to W and characterizes fluctuations of work trajectories around the most probable value $W^{\rm mp}$. A *trajectory free energy* $\mathcal{F}_N(W)$ can also be defined, which has a minimum at $W = W^\dagger$, this being the value of the work that has to be efficiently sampled to quantitatively test the Jarzynski equality. Within this formalism a Lagrange multiplier is also introduced, the inverse of which plays the role of a *trajectory temperature*. Our general solution for $P_N(W)$ exactly satisfies the fluctuation theorem by Crooks and allows us to investigate heat fluctuations for a protocol that is invariant under time reversal. The heat distribution is then characterized by a Gaussian component (describing small and frequent heat exchange events) and exponential tails (describing the statistics of large deviations and rare events). For the latter, the width of the exponential tails is related to the aforementioned *trajectory temperature*. Finite-size effects to the large-N theory and the recovery of work distributions for finite N are also discussed. Finally, we pay particular attention

PII: S1742-5468(04)87940-8 1742-5468/04/P10016+33$30.00

to the case of magnetic nanoparticle systems under the action of a magnetic field H where work and heat fluctuations are predicted to be observable in ramping experiments in micro-SQUIDs.

Keywords: intermittency

Contents

1. Introduction

There has been recent interest in the experimental measure of work fluctuations and the test of the so-called fluctuation theorems. Initially proposed in the context of sheared systems in a steady state [1], several versions of such theorems have been derived [2]. In particular, specific identities have been obtained in the context of stochastic systems that show how it is possible to recover the equilibrium free-energy change in a reversible transformation by exponential averaging over many non-equilibrium trajectories that start at equilibrium [3]–[5]. Let us consider a system initially in equilibrium in contact with a thermal bath (at temperature T) that is submitted to an isothermal perturbation according to a given protocol. Work fluctuations (WF) refer to the fact that the work W exerted upon the system depends on the particular non-equilibrium trajectory followed by the system. As the initial configuration or the trajectory is stochastic, the value of the work W changes among different trajectories, all generated with the same perturbation protocol. Transient violations (TV) of the second law refer to the fact that, among all possible WF, a fraction of them absorb heat from the bath that is transformed into work. Taken individually, these rare trajectories violate the Clausius inequality, $Q \leq T\Delta S$, where Q is the heat supplied from the bath to the system and ΔS is the change in the entropy, a state function defined through the transformation. In a transformation cycle $\Delta S = 0$ these TV satisfy $Q > 0$; i.e., they can absorb a net amount of heat from the bath during the cycle. In terms of the dissipated work $W_{\rm dis}$, the Clausius relation can be expressed in the following form:

$$W_{\rm dis} = W - W_{\rm rev} = T\Delta S - Q \geq 0. \tag{1}$$

In this expression W is the total work exerted upon the system. According to the first law of thermodynamics (conservation of the energy) W is given by $W = \Delta E - Q$, where ΔE is the change in the internal energy, and $W_{\rm rev}$ is the reversible work (identical to the free-energy change $\Delta F = \Delta E - T\Delta S$). Both the heat Q and $W_{\rm dis}$ (or W) are trajectory dependent; however, ΔS and $W_{\rm rev}$ are both trajectory independent as they are state functions, only dependent on the initial and final states. The Clausius inequality (1) has to be understood as a result that is valid after averaging the fluctuating quantities Q and W over an infinite number of trajectories (in what follows we will denote this average by $\overline{(..)}$). The second law reads $\overline{W_{\rm dis}} \geq 0$ and TV of the second law refer to the existence of trajectories where $W_{\rm dis} < 0$. From this point of view, TV are just WF characterized by the fact that $W_{\rm dis} < 0$. The interest in studying TV is that these describe large deviations of the work that have to be sampled in order to recover equilibrium free-energy differences from non-equilibrium measurements [6].

The steadily increasing development of nanotechnologies during the last decade has made WF experimentally accessible. Recent experiments on single RNA hairpins unfolded under the action of mechanical force [7] and micro-sized beads trapped by laser tweezers and moved through a solvent [8] have provided a first quantitative estimate of WF and TV. Related measurements include the experimental verification of the Gallavotti–Cohen fluctuation theorem in Rayleigh–Bernard convection [9] and turbulent flows [10]. This research is potentially very interesting as it leads to new insights about the physical processes occurring at the nanoscale, a frontier that marks the onset of complex organization of matter [11]. A characteristic of WF is that they are quickly suppressed as the system size or the time window of the measurement increases.

The central quantity describing WF is the work probability distribution $P_N(W)$ (N stands for the system size), $P_N(W)\mathrm{d}W$ being the fraction of non-equilibrium trajectories with work between W and $W + \mathrm{d}W$. The knowledge of this quantity is important for what it tells us about the mathematical form of the tails of the distribution, relevant to understanding the importance of large deviations of work values with respect to the average value. A precise knowledge of the form of the tails in that distribution gives us hints about how many experiments need to be done in order to recover equilibrium quantities from non-equilibrium experiments. In this work we investigate an ensemble of two-level systems as an explicit example where $P_N(W)$ can be analytically computed in the large-N approach using a path integral method. This approach allows us to exactly derive several exact results describing work and heat fluctuations in the system in the large-N limit but also for finite N. The most important result in the paper is the introduction of a trajectory thermodynamics formalism, the key quantity being the *trajectory entropy* $S_N(W)$. This allows us to infer several quantities such as the *trajectory free energy* $\mathcal{F}_N(W)$ and the *trajectory temperature* $\lambda(W)$, the latter being a Lagrange multiplier that plays the role of the inverse of a temperature, an intensive variable related to the statistics of large deviations or tails in the work and heat distributions. Two-state models represent a broad category of systems where WF and TV can be predicted to be experimentally observable, making the present calculations relevant as they might allow a detailed comparison between theory and experiments. In particular, we propose magnetic nanoparticles as excellent candidate systems to experimentally test the present theory.

The plan of the paper is as follows. In section 2 we describe the model and the large-N approach. In section 3 we develop the trajectory thermodynamics formalism that allows us to reconstruct the work distribution, and define a *trajectory entropy* $S_N(W) = \log(P_N(W))$ and a *trajectory free energy* $\mathcal{F}_N(W)$. In section 4 we show how the saddle-point equations derived in section 2 can be numerically solved. The dependence of the main parameters of the theory (most probable work W^{mp}, transient violations work $W^{\dagger}$ and fluctuation-dissipation ratio R) on the field protocol are discussed in section 4.1. Within the formalism it is then possible to show, section 5, that the entropy per particle $s(w)$ (w being the work W per particle) exactly satisfies the fluctuation theorem by Crooks. Moreover, it is possible to infer the shape of the tails in the work distribution from the sole knowledge of the Lagrange multiplier conjugated to the *trajectory entropy*, $\lambda(w)$, that plays the role of the inverse of a temperature (which we call the *trajectory temperature*) in the formalism. In section 6 we study heat fluctuations in the model. We show the existence of two sectors in the heat distribution that are described by a Gaussian central part (corresponding to small and most probable deviations) and two exponential tails (corresponding to large and rare deviations) showing the presence of intermittent heat fluctuations in the theory. In section 7 we discuss finite-size corrections to the large-N theory and how $P_N(W)$ for finite N can be reconstructed using the results from the large-N approach. Particular emphasis is finally placed in section 8 in the case of magnetic nanoparticle systems where WF are predicted to be experimentally observable and described by the present theory. Section 9 presents the conclusions.

2. Ensemble of two-state systems: the large-N approach

A broad category of systems can be modelled by an ensemble or collection of independent two-state systems. These offer realistic descriptions of electronic and optical devices that can function in two different configurations: atoms in their ground and excited states, magnetic particles whose magnetic moment can point in two directions, or biomolecules in their native and unfolded states, among others. Throughout the paper, and in view of the possible experimental implications, we will adopt the nomenclature of magnetic systems. A particle i in the ensemble ($1 \leq i \leq N$) has magnetic moment μ and can point in two directions according to the sign of the spin $\sigma_i = \pm 1$. A given configuration in the ensemble is specified by a string of spin values $\mathcal{C} \equiv \{\sigma_1, \sigma_2, \ldots, \sigma_N\}$. In the presence of an external field H, the energy of a configuration $\mathcal{C}$ is given by

$$E(\mathcal{C}) = -\mu H M(\mathcal{C}) = -\mu H \sum_{i=1}^{N} \sigma_i, \tag{2}$$

$M(\mathcal{C}) = \sum_{i=1}^{N} \sigma_i$ being the total magnetization of the system. The transition rates for individual particles will be denoted as $p^{\mathrm{up}}(H), p^{\mathrm{down}}(H)$ to indicate the transitions $\sigma = -1 \to \sigma' = 1$ and $\sigma = 1 \to \sigma' = -1$ respectively. These rates satisfy detailed balance, therefore $p^{\mathrm{up}}(H)/p^{\mathrm{down}}(H) = \exp(-2\beta\mu H)$, where $\beta = 1/k_{\mathrm{B}}T$, T being the bath temperature and $k_{\rm B}$ the Boltzmann constant. The overall transition rate is given by $p^{\rm tot}(H) = p^{\rm up}(H) + p^{\rm down}(H)$. Although it is possible to introduce structural disorder in the ensemble (e.g. by allowing μ or $p^{\rm up}(H)$ to be a random quenched variable), in the following analysis we will restrict ourselves to the non-disordered or mono-disperse case.

Let the system be prepared at $t = 0$ in an equilibrium state at an initial value of the field $H_0 = H_{\rm i}$ and let us consider an external isothermal perturbation that changes the field H according to a protocol function $H(t)$. Throughout this paper we will denote this non-equilibrium process a *ramping* experiment. If the variation is slow enough then the process is quasi-static and the system goes through a sequence of equilibrium states. However, if the rate $\dot{H}$ is large compared to the relaxation time of the particle then the magnetization $M = \sum_{i=1}^{N} \sigma_i$ does not follow the equilibrium curve $M_{\rm eq}(H) = N \tanh(\beta\mu H)$. To specify a trajectory it is then convenient to discretize time in N_s time-steps of duration Δt each and take the continuous-time limit $\Delta t \to 0, N_s \to \infty$ (with the total time $t = N_s \Delta t$ fixed) at the end. The perturbation protocol is specified by the sequence of values $\{H_k; 1 \leq k \leq N_s\}$, and a trajectory $\mathcal{T}$ is defined by the sequence of configurations $\mathcal{T} = (\mathcal{C}_k; 1 \leq k \leq N_s)$, where $\mathcal{C}_k = \{\sigma_i^k; 1 \leq i \leq N\}$ is the configuration at time $t = k\Delta t$. The total work exerted upon the system along a given trajectory is given by [5]

$$W(\mathcal{T}) = -\mu \sum_{k=0}^{N_s-1} M_{k+1}(H_{k+1} - H_k), \tag{3}$$

$M_k = \sum_{i=1}^{N} \sigma_i^k$ being the magnetization at time-step k. The dissipated work for a given trajectory is the difference between the total work and the reversible one, $W_{\rm dis} = W - W_{\rm rev}$, where $W_{\rm rev} = \Delta F$ is the change in equilibrium free energy between the initial and final values of the field. The free energy is given by $F(H) = -Nk_{\rm B}T \log(2\cosh(\beta\mu H))$. To quantify WF we have to compute the probability distribution for the total work measured over all possible non-equilibrium trajectories,

$$P_N(W) = \sum_{\mathcal{T}} p(\mathcal{T})\delta(W - W(\mathcal{T})) = \sum_{\{\sigma_i^k\}} p(\mathcal{T})\delta\left(W + \mu \sum_{k=0}^{N_s-1} M_{k+1}(H_{k+1} - H_k)\right), \tag{4}$$

where $p(\mathcal{T})$ denotes the probability of a given trajectory. The subindex N in $P_N(W)$ is written to emphasize the dependence of the distribution on the size of the system. $P_N(W)$ is computed using the Bayes formula $p(\mathcal{T}) = \prod_{k=0}^{N_s-1} q_k(\{\sigma_i^{k+1}\}|\{\sigma_i^k\}) p_0(\{\sigma_i^0\})$, where $q_k(\{\sigma'\}|\{\sigma\})$ denotes the transition probability to go from $\{\sigma\}$ to $\{\sigma'\}$ at time-step k, and $p_0(\{\sigma_i^0\})$ is the initially equilibrated (i.e. Boltzmann–Gibbs) distribution. Evaluation of the integral (4) requires the following steps: (1) trace out spins in the sum; (2) insert the factorized expression for $p(\mathcal{T})$; (3) use the integral representation for the delta function $\delta(x) = (1/2\pi)\int_{-\infty}^{\infty} {\rm d}\lambda\, \exp({\rm i}\lambda x)$, and (4) insert the following factor,

$$1 = \prod_{k=0}^{N_s-1} \frac{1}{2\pi} \int_{-\infty}^{\infty} {\rm d}\gamma_k\, {\rm d}M_k\, \exp\left({\rm i}\gamma_k\left(M_k - \sum_{i=1}^{N} \sigma_i^k\right)\right). \tag{5}$$

After some manipulations this leads to the following expression for the work probability distribution (up to some unimportant multiplicative terms):

$$P_N(W) \propto \int {\rm d}\lambda \prod_{k=0}^{N_s-1} ({\rm d}\gamma_k\, {\rm d}m_k) \exp(A(w, \lambda, \{\gamma_k\}, \{m_k\})), \tag{6}$$

where A is the saddle-point function, $w = W/N, m_k = M_k/N$ (throughout the paper we will use lower-case letters to refer to intensive quantities). The function $a = A/N$ is given by

$$
\begin{aligned}
a(w, \lambda, \{\gamma_k\}, \{m_k\}) = {} & -\lambda\left(w + \mu \sum_{k=0}^{N_s-1} m_{k+1}(H_{k+1} - H_k)\right) - \sum_{k=0}^{N_s} \gamma_k m_k \\
& + \sum_{k=0}^{N_s-1}\left(\frac{m_k+1}{2}\log(u_{k+1}) + \frac{1-m_k}{2}\log(v_{k+1})\right) \\
& + \log(\mathrm{e}^{\gamma_0} p^{\mathrm{up}}(H_{\mathrm{i}}) + \mathrm{e}^{-\gamma_0} p^{\mathrm{down}}(H_{\mathrm{i}})).
\end{aligned}
\tag{7}
$$

The terms u_k, v_k are given by

$$
u_{k+1} = \exp(\gamma_{k+1})(1 - p_k^{\mathrm{down}}) + \exp(-\gamma_{k+1})p_k^{\mathrm{down}} \tag{8}
$$

$$
v_{k+1} = \exp(\gamma_{k+1})p_k^{\mathrm{up}} + \exp(-\gamma_{k+1})(1 - p_k^{\mathrm{up}}) \tag{9}
$$

with the boundary condition $\gamma_{N_s} = 0$. The quantities $p^{\mathrm{up}}(H_{\mathrm{i}}), p^{\mathrm{down}}(H_{\mathrm{i}})$ are the transition rates at time $s = 0$, and we are assuming that at the initial condition the system is in thermal equilibrium. In the continuous-time limit (6) becomes a path integral over the variable λ and the functions $\gamma(t), m(t)$ with

$$
\begin{aligned}
a(w, \lambda, \gamma(s), m(s)) = {} & -\lambda\left(w + \mu\int_0^t m(s)\dot{H}(s)\,\mathrm{d}s\right) \\
& + \tfrac{1}{2}\int_0^t (m(s)(2\dot{\gamma}(s) + c(s)) + d(s))\,\mathrm{d}s + \log(\mathrm{e}^{\gamma(0)}p^{\mathrm{up}}(H_{\mathrm{i}}) + \mathrm{e}^{-\gamma(0)}p^{\mathrm{down}}(H_{\mathrm{i}})),
\end{aligned}
\tag{10}
$$

where

$$
c(s) = p^{\mathrm{down}}(s)(\exp(-2\gamma(s)) - 1) - p^{\mathrm{up}}(s)(\exp(2\gamma(s)) - 1) \tag{11}
$$

$$
d(s) = p^{\mathrm{down}}(s)(\exp(-2\gamma(s)) - 1) + p^{\mathrm{up}}(s)(\exp(2\gamma(s)) - 1). \tag{12}
$$

As we are interested in the crossover to the large-N regime we can estimate the integral (6) by using the saddle-point method. For each value of the work trajectory w the dominant contribution is given by the solution of the functional equations

$$
\frac{\delta a}{\delta\lambda} = w + \mu\int_0^t m(s)\dot{H}(s)\,\mathrm{d}s = 0 \tag{13}
$$

$$
\frac{\delta a}{\delta\gamma(s)} = \dot{m}(s) + m(s)p^{\mathrm{tot}}(s) - (p^{\mathrm{up}}(s) - p^{\mathrm{down}}(s)) + m(s)d(s) + c(s) = 0 \tag{14}
$$

$$
\frac{\delta a}{\delta m(s)} = \dot{\gamma}(s) - \lambda\mu\dot{H}(s) + \frac{1}{2}c(s) = 0 \tag{15}
$$

with the boundary conditions

$$
\gamma(t) = 0; \qquad m(0) = \tanh(\gamma(0) + \beta\mu H_{\mathrm{i}}). \tag{16}
$$

Note that the boundary conditions are a bit special as causality is broken. The function $\gamma(s)$ has the boundary condition located at the final time $s = t$ while the boundary condition for $m(s)$ is located at the initial time $s = 0$. These equations can be numerically solved in general and analytically solved only partially and for some particular cases (e.g. in the case where the rate $\dot{H}$ is constant). Before presenting detailed numerical solutions to these equations we should point out several general aspects of such solutions. At first we note how, for a given value of λ, equation (15) together with the boundary condition $\gamma(t) = 0$ can be solved giving the solution $\gamma_\lambda(s)$, the subindex λ emphasizing the dependence of this solution on the parameter λ. Inserting this result in (14) and using the boundary condition (16) we get the solution $m_\lambda(s)$. Finally, insertion of $m_\lambda(s)$ in (13) gives a value for the work $w(\lambda)$. This last relation can then be inverted[1] to give $\lambda(w)$, and from it, the solutions $\gamma_\lambda(s), m_\lambda(s)$ will also depend on the value of w. To better emphasize this dependence we will denote by $\lambda(w), \gamma_w(s), m_w(s)$ the solutions of (13)–(15) for a given value of w and

$$s(w) = a(w, \lambda(w), \gamma_w(s), m_w(s)) \tag{17}$$

the corresponding extremal value of a. We will also make explicit the w-dependence in the time-dependent quantities $c(s), d(s)$ in (11), (12) and denote them by $c_w(s), d_w(s)$ respectively. Furthermore, we can define the trajectory entropy $S_N(W)$:

$$P_N(W) = \exp(S_N(W)). \tag{18}$$

In the large-N limit, from (6), (17) we have

$$s(w) = \lim_{N\to\infty} \frac{S_N(W)}{N} \qquad \text{with } W = Nw, \tag{19}$$

the function $s(w)$ playing the role of a trajectory entropy per particle that counts the density of trajectories per particle with work equal to w. This means that, for N finite, $\Phi_N(w)\,\mathrm{d}w = \exp(Ns(w))\,\mathrm{d}w$ is approximately proportional to the fraction of trajectories with work between w and $w + \mathrm{d}w$. From (18), (19) an approximate expression for the work probability distribution can be written:

$$P_N(w) = \frac{\Phi_N(w)}{\int_{w_{\min}}^{w_{\max}} \Phi_N(w')\,\mathrm{d}w'} = \frac{\exp(Ns(w))}{\int_{w_{\min}}^{w_{\max}} \exp(Ns(w'))\,\mathrm{d}w'} \tag{20}$$

where $w_{\min}$ and $w_{\max}$ are the minimum and maximum possible values of the work. Clearly, from (3) these values are given by $w_{\max} = -w_{\min} = \mu(H_\mathrm{f} - H_\mathrm{i})$, where $H_\mathrm{f} = H(t)$ is the final value of the magnetic field. The subindex N in $P_N(w)$ and $\Phi_N(w)$ emphasizes the dependence of these quantities on the size of the system. Finally, we note that, albeit the solutions (13)–(15) have been obtained using the saddle-point approximation (only valid for large N), the final result (20) can be very accurate for small values of N. This result, that at first glance may appear striking, is just a consequence of the non-interacting character of the Hamiltonian (2). This point is discussed in more detail in section 7. There we show that, albeit (20) is only approximate for finite N, the cumulants that we can extract from $s(w)$ are exact for any N. This allows us to exactly reconstruct the finite N distribution from the sole knowledge of $s(w)$.

[1] For that we are assuming that $w(\lambda)$ is a monotonic function, a result that we have not proven in full generality, yet it is in accordance with all the cases we analysed.

The action $A = Na$ in (10) could be used (employing Monte Carlo algorithms) to generate trajectories according to their probability $P_N(w)$.[2] Inserting (15) in (10) we get

$$s(w) = -\lambda w + \tfrac{1}{2}\int_0^t d_w(s)\,\mathrm{d}s + \log(\mathrm{e}^{\gamma(0)} p^{\mathrm{up}}(H_{\mathrm{i}}) + \mathrm{e}^{-\gamma(0)} p^{\mathrm{down}}(H_{\mathrm{i}})). \tag{21}$$

The value w for which $s(w)$ is maximum yields the most probable work ($w = w^{\mathrm{mp}}$) among all trajectories. This can be evaluated using the equation

$$s'(w) = \frac{\partial a}{\partial w} = -\lambda(w), \tag{22}$$

where we have used the chain rule together with the extremum conditions (13)–(15) as well as (10). We will see later in section 6 that the Lagrange multiplier $\lambda(w)$ is related to the inverse of a new energy scale or temperature that describes the tails of the work distribution. This quantity is of much current interest as it describes the statistics of rare events and large deviations of work values from the average which are observable in small systems. The extremum solution of (22) can then be written as $\lambda(w^{\mathrm{mp}}) = 0$,

$$\left.\frac{\partial s(w)}{\partial w}\right|_{w=w^{\mathrm{mp}}} = 0 \qquad \text{or} \qquad \lambda(w^{\mathrm{mp}}) = 0. \tag{23}$$

This solution solves (13)–(15) giving $\gamma_{w^{\mathrm{mp}}}(s) = c_{w^{\mathrm{mp}}}(s) = d_{w^{\mathrm{mp}}}(s) = 0$. Equations (13), (14) then give the solution for the most probable trajectory (usually derived using standard statistical methods),

$$\dot{m}(s) = -m(s)p^{\mathrm{tot}}(s) + (p^{\mathrm{up}}(s) - p^{\mathrm{down}}(s)). \tag{24}$$

The reversible process is a special case (only valid for slow enough perturbation protocols) and corresponds to $\dot{m}(s) = 0$ or

$$m(s) = (p^{\mathrm{up}}(s) - p^{\mathrm{down}}(s))/p^{\mathrm{tot}}(s) = \tanh(\beta\mu H(s)). \tag{25}$$

3. Trajectory thermodynamics formalism

From the trajectory entropy $s(w)$ we can construct a trajectory free energy $\mathcal{F}(w)$ useful to predict under which conditions TV are properly sampled and fluctuation theorems can be quantitatively verified. For this we consider the Jarzynski equality [3],

$$\overline{\exp\left(-\frac{W}{k_{\mathrm{B}}T}\right)} = \exp\left(-\frac{\Delta F}{k_{\mathrm{B}}T}\right), \tag{26}$$

that we can write as

$$\begin{aligned}\exp\left(-\frac{\Delta F}{k_{\mathrm{B}}T}\right) &= \int \mathrm{d}W\, P_N(W) \exp\left(-\frac{W}{k_{\mathrm{B}}T}\right) \\ &= \int \mathrm{d}W\, \exp\left(-\frac{W}{k_{\mathrm{B}}T} + S_N(W)\right) = \int \mathrm{d}W \exp\left(-\frac{\mathcal{F}_N(W)}{k_{\mathrm{B}}T}\right),\end{aligned} \tag{27}$$

[2] The easiest procedure then would be to start from an initial trajectory $\gamma(s), m(s)$ (satisfying the boundary conditions $m(0) = \tanh(\gamma(0) + \beta\mu H_{\mathrm{i}}); \gamma(t) = 0$) and perform successive 'local' updates along the trajectory and accepting the moves according to the change in the action A (by using an algorithm that satisfies detailed balance, as defined by the action A, and respects the boundary conditions).

where we used (18) and we have defined the trajectory free energy,

$$\mathcal{F}_N(W) = W - k_\mathrm{B} T S_N(W). \tag{28}$$

In the large-N limit, using (19), we can write

$$\begin{aligned}\exp\left(-\frac{\Delta F}{k_\mathrm{B}T}\right) &= \int \mathrm{d}w\, \exp\left(-\frac{N}{k_\mathrm{B}T}(w - k_\mathrm{B}Ts(w))\right) \\ &= \int \mathrm{d}w\, \exp\left(-\frac{N\mathcal{F}(w)}{k_\mathrm{B}T}\right) \equiv \exp\left(-\frac{N\mathcal{F}(w^\dagger)}{k_\mathrm{B}T}\right),\end{aligned} \tag{29}$$

where

$$\mathcal{F}(w) = w - k_\mathrm{B}Ts(w) \tag{30}$$

is a trajectory free energy (per particle) that depends on the particular value of the work w. Evaluating the integral (29) by the steepest descent method and using (21) we obtain the *thermodynamic* relations

$$\frac{1}{k_\mathrm{B}T} = \frac{\partial s(w)}{\partial w}\bigg|_{w=w^\dagger} = -\lambda(w^\dagger) \tag{31}$$

$$\mathcal{F}(w^\dagger) = \Delta F/N = w_\mathrm{rev} = w^\dagger - k_\mathrm{B}Ts(w^\dagger) = -\frac{k_\mathrm{B}T}{2}\int_0^t d_{w^\dagger}(s)\,\mathrm{d}s. \tag{32}$$

Using the definition (30) together with (23), (31) we have the relations

$$\frac{\partial\mathcal{F}(w)}{\partial w}\bigg|_{w=w^\mathrm{mp}} = 1 \tag{33}$$

$$\frac{\partial\mathcal{F}(w)}{\partial w}\bigg|_{w=w^\dagger} = 0 \tag{34}$$

i.e. the entropy has a maximum at $w = w^\mathrm{mp}$ and the free energy has a minimum at $w = w^\dagger$. These relations bear similarity to those considered in thermodynamics but now apply to work trajectory values. For the case of the canonical ensemble the quantities $s(w), \mathcal{F}(w), w$ play the role of the standard entropy, free energy and internal energy, while $\lambda(w)$ is the intensive variable corresponding to the inverse of a temperature. Actually, from (22), (31) we can define the *trajectory temperature*,

$$\hat{T}(w) = -\frac{1}{k_\mathrm{B}\lambda(w)}, \tag{35}$$

which can be positive or negative depending on the sign of $\lambda(w)$. The trajectory temperature satisfies the equality $\hat{T}(w^\dagger) = T$ and, for now, it is just a Lagrange multiplier devoid of any particular physical meaning. We will see below in section 6 that, under some conditions, it is possible to endow the trajectory temperature with a physical interpretation.

A graphical construction of the relations (31), (32) is shown in figure 1. This figure illustrates how the most important quantities $w_\mathrm{rev}, w^\mathrm{mp}, w^\dagger, \overline{w}$ are related to each other. In particular, $\overline{w} = \int \mathrm{d}w\, wP_N(w)$ is expected to differ from w^mp, albeit that difference can be small for highly symmetric distributions.

The difference between $w^\dagger$ and w^mp indicates that the average (29) is properly weighed whenever trajectories with work values around $w^\dagger$ are sampled. This result indicates that

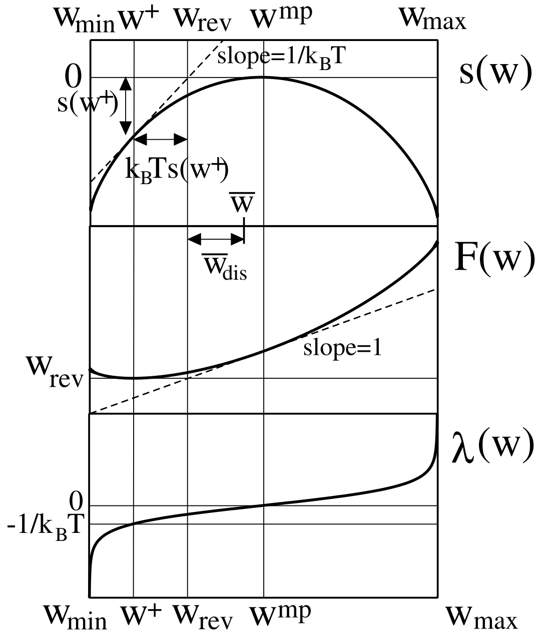

Figure 1. Diagrams showing the different relevant quantities in the trajectory thermodynamics formalism. Upper panel: trajectory entropy $s(w)$ related by (20) to the density of trajectories with work equal to w. Middle panel: trajectory free energy $\mathcal{F}(w) = w - k_\mathrm{B}Ts(w)$. Lower panel: Lagrange multiplier $\lambda(w)$. There are six most relevant work values: w_max and w_min for the maximum and minimum values of the work; w^mp, the most probable work value given by $s'(w^\mathrm{mp}) = \lambda(w^\mathrm{mp}) = 0$ or $\mathcal{F}'(w^\mathrm{mp}) = 1$; $w^\dagger$, the value of the work that has to be sampled to recover free energies from non-equilibrium work values using the Jarzynski equality (26) (this is given by $s'(w^\dagger) = -\lambda(w^\dagger) = 1/k_\mathrm{B}T$ or $\mathcal{F}'(w^\dagger) = 0$); $w_\mathrm{rev} = F(H_\mathrm{f}) - F(H_\mathrm{i})$, the reversible work; and $\overline{w_\mathrm{dis}} = \int_{w_\mathrm{min}}^{w_\mathrm{max}} (w - w_\mathrm{rev}) P_N(w)\,\mathrm{d}w$, the average dissipated work. They are related by $w_\mathrm{min} < w^\dagger < w^\mathrm{mp} < w_\mathrm{max}$ while the second law of thermodynamics imposes $w_\mathrm{dis} \geq 0$.

proper sampling of non-equilibrium work values around $w^\dagger$ is required to derive equilibrium free energies from non-equilibrium measurements by using the Jarzynski equality. A proper sampling of work values around $w^\dagger$ can be guaranteed when, out of the total number of trajectories, a finite fraction of work trajectory values in the vicinity of $w^\dagger$ is observed. From a practical point of view this means that the histogram of work values must extend down to $w^\dagger$. If this is not achieved, then the exponential average performed over a finite number of non-equilibrium experiments has a bias that can be estimated in some cases [14, 15]. Equations (31), (32) are readily solved at the Gaussian level (i.e. assuming that $P_N(w)$ is exactly a Gaussian or $s(w)$ a quadratic function) giving $w^\dagger = w^\mathrm{mp} - \sigma_w^2/k_\mathrm{B}T$ (σ_w^2 being the variance of the Gaussian work distribution). For quasi-reversible processes in the linear response regime [15], the fluctuation-dissipation theorem implies $\sigma_w^2 \simeq 2k_\mathrm{B}Tw_\mathrm{dis}^\mathrm{mp}$, giving $w_\mathrm{dis}^\dagger \simeq -w_\mathrm{dis}^\mathrm{mp}$, i.e. trajectories with negative values of the dissipated work that are of the order (in absolute value) of the average dissipated work must be sampled to quantitatively verify the validity of the Jarzynski equality. An example of such a quasi-reversible process, where $P_N(w_{\rm dis})$ is exactly a Gaussian, is the case of a Brownian particle subjected to an harmonic potential and dragged in a fluid [8, 16, 17].

4. Numerical solution of the equations

Equations (13)–(15) can be numerically solved in general. We assume Glauber transition rates given by

$$p^{\rm up}(H) = p^{\rm tot}(H)q(H); \qquad p^{\rm down}(H) = p^{\rm tot}(H)(1-q(H)) \tag{36}$$

with $q(H) = (1+\tanh(\beta\mu H))/2$ and $p^{\rm tot}(H) = 1/\tau_{\rm relax}(H) = \alpha(H)$ corresponding to the inverse of the relaxation time. In this case,

$$p^{\rm up}(s) = \alpha(H(s))\frac{\exp(\beta\mu H(s))}{2\cosh(\beta\mu H(s))} \tag{37}$$

$$p^{\rm down}(s) = \alpha(H(s))\frac{\exp(-\beta\mu H(s))}{2\cosh(\beta\mu H(s))}. \tag{38}$$

Inserting these expressions in (11), (12) we obtain

$$c(s) = -\alpha(H(s))\frac{\sinh(2\gamma(s)+\beta\mu H(s))}{\cosh(\beta\mu H(s))} + \alpha(H(s))\tanh(\beta\mu H(s)) \tag{39}$$

$$d(s) = \alpha(H(s))\frac{\cosh(2\gamma(s)+\beta\mu H(s))}{\cosh(\beta\mu H(s))} - \alpha(H(s)). \tag{40}$$

The solution of the equations consists of the following steps:

(1) *Solution of* $\gamma_\lambda(s)$. With the boundary condition at the final time $s = t$, $\gamma_\lambda(t) = 0$, equation (15) has to be numerically integrated backwards in time. Inserting (37), (38) in (15) we obtain

$$\dot{\gamma}(s) = \lambda\mu\dot{H}(s) + \alpha(H(s))\sinh(\gamma(s))(\cosh(\gamma(s)) + \sinh(\gamma(s))\tanh(\beta\mu H(s))). \tag{41}$$

However, a direct numerical integration of this equation leads to divergences and numerical instabilities. It is then convenient to express (41) in terms of a new variable $\epsilon(s) = 1/\cosh(\gamma(s))$ which displays smooth behaviour. Equation (41) becomes

$$\dot{\epsilon}(s) = -\frac{\tanh(\gamma(s))}{\cosh(s)}\Bigg(\lambda\mu\dot{H}(s) + \alpha(H(s))\sinh(\gamma(s))\left(\frac{1}{\epsilon(s)} + \sinh(\gamma(s))\tanh(\beta\mu H(s))\right)\Bigg) \tag{42}$$

with the boundary condition $\epsilon(t) = 1$. This equation can then be easily numerically integrated to give $\gamma_\lambda(s)$ for a given value of λ.

(2) *Solution of* $m_\lambda(s)$. Once the solution of (42) for a given value of λ, $\gamma_\lambda(s)$, is found, then it is possible to integrate (14) to find $m_\lambda(s)$. Because (14) is linear its solution can be explicitly written:

$$m_\lambda(s) = m_\lambda(0)\exp\left(\int_0^s A_1(u)\,{\rm d}u\right) + \int_0^s {\rm d}u\,A_2(u)\exp\left(\int_u^s A_1(v)\,{\rm d}v\right) \tag{43}$$

with the definitions

$$A_1(s) = \alpha(H(s))\frac{\sinh(2\gamma_\lambda(s) + \beta\mu H(s))}{\cosh(\beta\mu H(s))} \tag{44}$$

$$A_2(s) = -\alpha(H(s))\frac{\cosh(2\gamma_\lambda(s) + \beta\mu H(s))}{\cosh(\beta\mu H(s))} \tag{45}$$

with the initial condition $m_\lambda(0) = \tanh(\gamma_\lambda(0) + \beta\mu H_{\rm i})$.

(3) *Evaluation of* $w, s(w), \mathcal{F}(w)$. Once $\gamma_\lambda(s), m_\lambda(s)$ are known then we can evaluate w using (13), the entropy $s(w)$ using (21) and the free energy $\mathcal{F}(w) = w - k_{\rm B}Ts(w)$.

(4) *Dependence of the numerical algorithm on the sign of* λ. We must emphasize that the solution of the equations previously described only works in a sector of values of λ of a given sign, $\lambda < 0$, and leads to numerical instabilities in the other sector, $\lambda > 0$, indicative that the transformation $\epsilon(s) = 1/\cosh(\gamma(s))$ is inappropriate for $\lambda > 0$. We have found a simple way out of this problem. It can be easily proven that the solution of (15) for a given value of $\lambda > 0$ is equivalent to the solution of that equation with the value of λ with its sign reversed $(-\lambda < 0)$ and for the reversed field protocol $H_r(s) = -H(s)$ (the subindex r stands for reversed). Equation (15) can then be solved and the resulting reversed solutions $m_r(s), \gamma_r(s), c_r(s), d_r(s)$ give the final solutions for the original value of $\lambda > 0$: $m(s) = -m_r(s), \gamma(s) = -\gamma_r(s), c(s) = -c_r(s), d(s) = d_r(s)$ (all change sign except $d(s)$). At first glance, this symmetry property might seem to be related to the content of the fluctuation theorem. However, this relation is only apparent because the reversed process in this case does not correspond to the time-reversal protocol which should be instead $H_r(s) = H(t-s)$ (see the discussion below in section 5).

For the present numerical analysis, and for the sake of simplicity, we will consider a particular example where the ramping field $H(s)$ changes from $H(0) = H_{\rm i}$ to $H(t) = H_{\rm f}$ at a constant rate $r = \dot{H}$:

$$H(s) = H_{\rm i} + rs, \qquad r = \dot{H} = \frac{H_{\rm f} - H_{\rm i}}{t}. \tag{46}$$

We will also consider $p^{\rm tot}(H) = \alpha$ independent of the field. This is tantamount to assuming that $p^{\rm tot}(H)$ corresponds to a microscopic attempt frequency or, rather, that the activation barrier is field independent. We numerically solved the equations in natural units $\mu = k_{\rm B}T = 1$ and we have chosen $\alpha = 1$ as the characteristic relaxation timescale of the system. Results for different values of $H_{\rm i}, H_{\rm f}$ have been obtained by doing ramping experiments at different values of the ramping speed r. In figure 2 we show, in the particular example $H_{\rm i} = 0$, $H_{\rm f} = 1$, the results for the magnetization trajectory solutions $m_\lambda(s)$ and the Lagrange multiplier $\gamma_\lambda(s)$. These are plotted as a function of the time-dependent field $H(s)$ for different values of λ and for a given value of the ramping speed. In figures 3, 4 we show several trajectory thermodynamics quantities at different ramping speeds $(r = 0.01, 0.1, 1, 10)$. In the left panel of figure 3 we plot the magnetization for the most probable trajectories $m_{\lambda=0}(s)$ as a function of $H(s)$. In the right panel of figure 3 and in 4 we show the different trajectory thermodynamics quantities as a function of $w_{\rm dis}$: the inverse temperature $\lambda(w_{\rm dis})$, the trajectory entropy $s(w_{\rm dis})$ and the trajectory free energy $\mathcal{F}(w_{\rm dis})$.

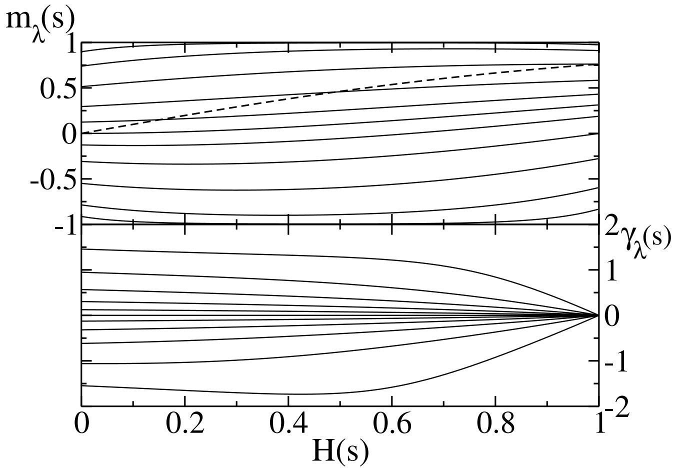

Figure 2. Protocol with $H_{\rm i} = 0, H_{\rm f} = 1$ and ramping speed $r = 1$. We consider natural units $\mu = k_{\rm B}T = 1$. Curves correspond to different values of λ ($\lambda = -5, -2, -1, -0.5, -0.2, 0., 0.2, 0.5, 1, 2, 5$ from top to bottom in the upper and lower panel). Upper panel: magnetization $m_\lambda(s)$ obtained from (43). The dashed curve is the equilibrium magnetization $m_{\rm eq}(H) = \tanh(H)$ corresponding to the reversible ramping experiment $r = 0$. Lower panel: Lagrange multiplier $\gamma_\lambda(s)$ obtained from (42). Note the boundary condition $\gamma_\lambda(t) = 0$ and the presence of the most probable trajectory $\gamma_{\lambda=0}(s) = 0, \forall s$.

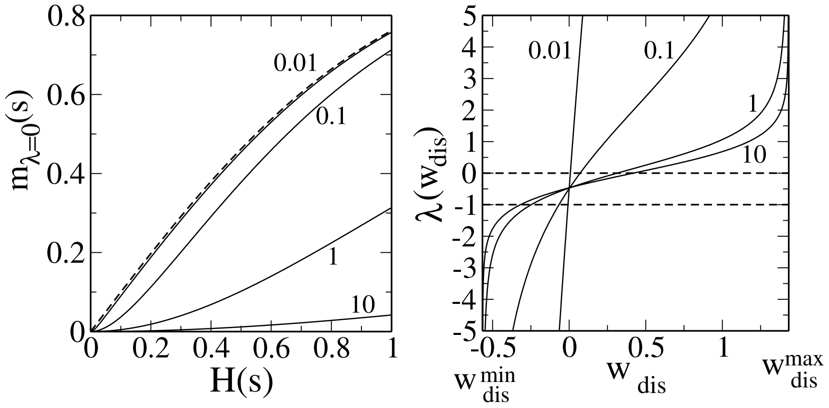

Figure 3. Protocol with $H_{\rm i} = 0, H_{\rm f} = 1$ and different ramping speeds $r = 0.01, 0.1, 1, 10$ (indicated by numbers along the continuous curves in both panels). We consider natural units $\mu = k_{\rm B}T = 1$. The reversible work is $w_{\rm rev} = -0.433\,781$ and $w_{\rm max} = 1, w^{\rm max}_{\rm dis} = w_{\rm max} - w_{\rm rev} = 1.433\,781, w^{\rm min}_{\rm dis} = w_{\rm min} - w_{\rm rev} = -0.566\,219$. Left panel: magnetization evolution for the most probable trajectories. The dashed line corresponds to the reversible trajectory for $r \to 0$. Right panel: Lagrange multiplier $\lambda(w_{\rm dis})$ for different ramping speeds. The intersection of the different curves with the dashed line $\lambda = 0$ gives $w^{\rm mp}$ while the intersection with $\lambda = -1$ gives $w^\dagger$. The intersection of all lines at different speeds around $\lambda = -0.5$ is only accidental (looking at a larger resolution or considering other parameters for the protocol such common crossing does not exist).

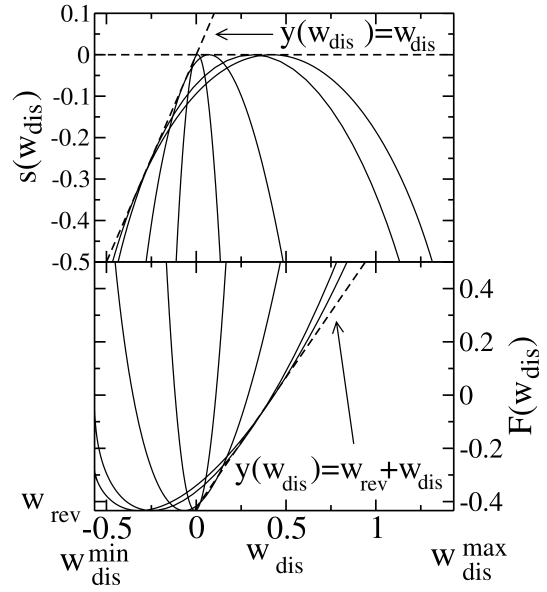

Figure 4. The same parameters and ramping speeds as in figure 3. Narrower curves correspond to lower ramping speeds. Upper panel: dynamical entropies $s(w)$ plotted as functions of $w_{\rm dis} = w - w_{\rm rev}$. According to the trajectory thermodynamics relations (31), (32) the straight line $y(w_{\rm dis}) = w_{\rm dis}/k_{\rm B}T$ (we take $k_{\rm B}T = 1$) is tangent to the curve $s(w_{\rm dis})$ at $w_{\rm dis} = w^{\dagger}_{\rm dis} = w^{\dagger} - w_{\rm rev}$ and crosses the $w_{\rm dis}$-axis at $w_{\rm dis} = 0, s(w_{\rm dis}) = 0$. All entropies vanish at $w_{\rm dis} = w^{\rm mp}_{\rm dis} = w^{\rm mp} - w_{\rm rev}$. Lower panel: trajectory free energy $\mathcal{F}(w_{\rm dis})$. It is identical to the equilibrium free-energy change $\Delta F = w_{\rm rev}$ at $w_{\rm dis} = w^{\dagger}_{\rm dis}$. According to the same relations (31), (32) the straight line $y(w_{\rm dis}) = w_{\rm rev} + w_{\rm dis}$ is tangent to the curve $\mathcal{F}(w_{\rm dis})$ at $w_{\rm dis} = w^{\rm mp}_{\rm dis} = w^{\rm mp} - w_{\rm rev}$ and crosses the $w_{\rm dis}$-axis at $w_{\rm dis} = 0, \mathcal{F}(w_{\rm dis}) = w_{\rm rev}$.

4.1. Average and variance of the work distribution

As has been schematically depicted in figure 1, there are different work quantities that can be of relevance to characterize work fluctuations. We have already defined the most probable work $w^{\rm mp}$ and the work $w^{\dagger}$. Another important quantity is the average work $\overline{w}$,

$$\overline{w} = \int_{w_{\rm min}}^{w_{\rm max}} w P_N(w)\, {\rm d}w, \tag{47}$$

where $P_N(w)$ was defined in (4) or in approximate form in (20). In most cases (for instance, when the work distribution has asymmetric tails) the average work $\overline{w}$ is different from the most probable work $w^{\rm mp}$. $\overline{w}$ can be lower or higher than the most probable work $w^{\rm mp}$. However, in our large-N theory, $w^{\rm mp} = \overline{w}$ and we will use indistinctly both quantities in this section. We defer the discussion about finite-size effects in these quantities until section 7. Another important quantity that characterizes the work distribution is its variance,

$$\sigma_w^2 = \overline{w^2} - (\overline{w})^2 = \overline{w^2_{\rm dis}} - (\overline{w_{\rm dis}})^2. \tag{48}$$

The average work $\overline{w}$ (or w^{mp}) is the most relevant physical quantity that connects with classical thermodynamics. The second law of thermodynamics establishes that it cannot be lower than the reversible work, $\overline{w} \geq w_{\mathrm{rev}}$. However, it is clear that there can be WF such that $w < w_{\mathrm{rev}}$. These have been called transient violations (TV) of the second law. The relevant work value characterizing this sector of trajectories is given by $w^{\dagger}$. Clearly, $\overline{w}$ is always higher than $w^{\dagger}$. In figure 5 (left panel) we show the dependence of $w_{\mathrm{dis}}, w^{\dagger}$ with the ramping speed when the field is ramped from $H_{\mathrm{i}} = 0$ to H_{f} for different values of H_{f}. It is possible to write down explicit analytic expressions for the cumulants of the distribution $P_N(w)$ in the large-N limit. Interestingly, and due to the non-interacting character of the model (2), the cumulants derived from the large-N approach are exact at all values of N; see section 7. In particular, the first moment is given by

$$\overline{w_{\mathrm{dis}}} = w^{\mathrm{mp}} - w_{\mathrm{rev}} = 2\mu \int_0^t \mathrm{d}s\, \dot{H}(s) \int_0^s \mathrm{d}u\, \dot{H}(u) \frac{\partial q(H)}{\partial H}\bigg|_{H=H(u)} \exp\left(- \int_u^s \mathrm{d}v\, \alpha(H(v)) \right). \tag{49}$$

The expression for the second cumulant or variance can be obtained by expanding the function $s(w)$ (21) up to second order with $\lambda(w)$ as the small parameter. Using the result $s'(w^{\mathrm{mp}}) = 0$ we get

$$s(w) = s(w^{\mathrm{mp}}) + \frac{1}{2}\left(\frac{\partial^2 s(w)}{\partial w^2} \right)(w - w^{\mathrm{mp}})^2 + \mathcal{O}[(w - w^{\mathrm{mp}})^3]. \tag{50}$$

From (20) and (22) we obtain the relation

$$\sigma_w^2 = \frac{1}{N}\left[\frac{\partial^2 s(w)}{\partial w^2}\bigg|_{w=w^{\mathrm{mp}}} \right]^{-1} = \frac{1}{N}\frac{\mathrm{d}w(\lambda)}{\mathrm{d}\lambda}\bigg|_{\lambda=0}. \tag{51}$$

We do not reproduce the details of this lengthy calculation here; the same results have been already obtained in a slightly different context in a previous work and in the limit of large free-energy changes ΔF as compared to $k_{\mathrm{B}}T$ [12].

Another interesting aspect of the present theory is that it is possible to expand the cumulants around the limits $r \to 0$ or ∞. The former is particularly interesting because it corresponds to the so-called linear-response regime. In [12] this regime was considered by expanding the average dissipated work up to linear order in the perturbation speed. By using dimensional arguments and direct comparison with the equivalent expression derived in the context of mechanical force [12], we can derive the following result:

$$\overline{w_{\mathrm{dis}}} = \frac{\mu \Delta M \tau_{\mathrm{relax}}(H_{\mathrm{c}} = 0)}{N} r + \mathcal{O}(r^2), \tag{52}$$

where $\Delta M = M_{\mathrm{eq}}(H_{\mathrm{f}}) - M_{\mathrm{eq}}(H_{\mathrm{i}})$ is the difference between the equilibrium total magnetizations at the initial and final values of the field whereas $\tau_{\mathrm{relax}}(H_{\mathrm{c}} = 0) = 1/p_{\mathrm{tot}}(H_{\mathrm{c}} = 0)$ is the relaxation time at the critical value of the field where the configurations $\sigma = +1$ and -1 are equiprobable (i.e. at $H_{\mathrm{c}} = 0$). Equation (52) indicates that dissipated work is large when the relaxation time is larger as we can expect. The linear response regime breaks down for large ramping speeds when $\overline{w_{\mathrm{dis}}} \gg w_{\mathrm{rev}}$ and the dissipated work starts to saturate. An interesting quantity quantifying deviations from the

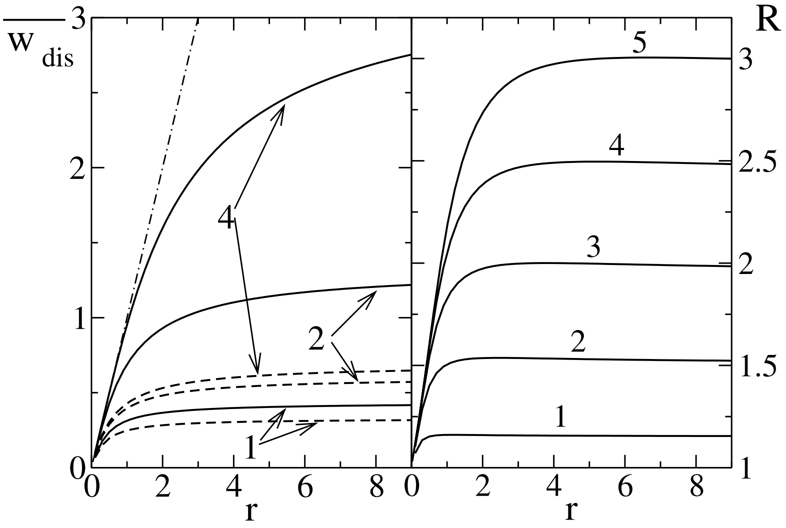

Figure 5. Ramping experiments with $H_\mathrm{i} = 0$ and different values of H_f as a function of the ramping speed r. We consider natural units $\mu = k_\mathrm{B}T = 1$. Left panel: w_dis (continuous curves) and $-w^\dagger_\mathrm{dis}$ (dashed curves) for different values of H_f indicated in the figure. The dash–dotted straight line corresponds to the linear response behaviour $w_\mathrm{dis} = r$ given by (52). Note that $w^\dagger_\mathrm{dis}$ is negative so we changed its sign in order to compare it with w_dis. Right panel: fluctuation-dissipation ratio R as a function of r evaluated at different values of H_f (from bottom to top, $H_\mathrm{f} = 1, 2, 3, 4, 5$).

linear-response regime is the fluctuation-dissipation ratio R defined by

$$R = \frac{\sigma^2_W}{2k_\mathrm{B}T\overline{W_\mathrm{dis}}} = \frac{N\sigma^2_w}{2k_\mathrm{B}T\overline{w_\mathrm{dis}}}. \qquad (53)$$

In the limit of small $r \to 0$, when $\overline{w_\mathrm{dis}} \propto r$, then R converges to 1 (in agreement with the fluctuation-dissipation theorem, a result that in the context of steady-state systems has been proven in [13]) but deviates from 1 as r increases. In the right panel of figure 5 we show $R(r)$ when the field is ramped from $H_\mathrm{i} = 0$ to H_f at different values of H_f. In this case the behaviour of both $\overline{w_\mathrm{dis}}$ and R is monotonic with r. In figure 6 we show the same ramping experiments but comparing, for a given ramping speed, the results for $w_\mathrm{dis}, -w^\dagger, R$ as a function of H_f. For values of H_f small enough the ramping process is well described by the linear response-approximation discussed in section 3 where $w^\dagger_\mathrm{dis} \simeq -w^\mathrm{mp}_\mathrm{dis}$.

Let us finish this section by emphasizing that the dependence $R(r)$ can be quite complicated and even non-monotonic in some cases. Such behaviour is observed in the case where the ramping protocol is given by $H(s) = H_\mathrm{A}(1 - 2s/t)$, i.e. the field starts at a given value $H_\mathrm{i} = -H_\mathrm{A}$ (H_A denotes the field amplitude) and increases until its reversed value $H_\mathrm{f} = H_\mathrm{A}$ is reached. This case is of much interest regarding heat fluctuations and is discussed in detail in section 6. In figure 7 we show the behaviour of the average work $\overline{w}$ (equal to the average dissipated work as $w_\mathrm{rev} = 0$ due to the independence of the free energy on the sign of H_A) and R as a function of the ramping speed for different values of H_A.

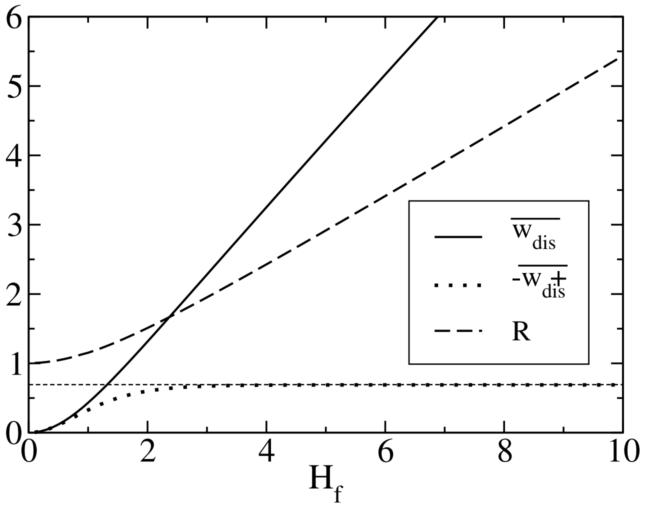

Figure 6. Ramping experiment with $H_{\rm i}=0$ and $r=100$ as a function of $H_{\rm f}$. We consider natural units $\mu=k_{\rm B}T=1$. The average dissipated work $w_{\rm dis}^{\rm mp}$ (continuous curve) and R (dashed curve) increase with the field but $w_{\rm dis}^{\dagger}$ (dotted line) saturates to a finite value equal to $-\log(2)$ (we represent $-w_{\rm dis}^{\dagger}$ in order to compare with $w_{\rm dis}^{\rm mp}$).

5. The fluctuation theorem

The saddle-point equations (13)–(15) were derived in the large-N limit. Indeed (20) is not exact for finite N but has corrections. However, the results obtained for $s(w)$ exactly satisfy the fluctuation theorem of Crooks [5]. This theorem states the following. Let us consider a process where the system is perturbed according to a protocol $H_{\rm F}(s)$ during the time interval $[0,t]$, the system initially being in equilibrium at the value of the field $H_{\rm i}=H_{\rm F}(0)$. We will call this the forward (F) process. Let us now consider the reverse process defined as the time-reversed of the forward process: in this process the system starts in equilibrium at time $s=0$ at the value of the field $H_{\rm i}=H_{\rm F}(t)$ and the field is changed according to the protocol $H_{\rm R}(s)=H_{\rm F}(t-s)$. Let the distribution of works generated in this way be $P_{\rm F}(W),P_{\rm R}(W)$ for the forward (F) and reverse (R) processes respectively. The two distributions satisfy the following relation [5]:

$$\frac{P_{\rm F}(W)}{P_{\rm R}(-W)}=\exp\left(\frac{W-\Delta F}{k_{\rm B}T}\right)=\exp\left(\frac{W_{\rm dis}}{k_{\rm B}T}\right), \tag{54}$$

where $\Delta F=F(H_{\rm f})-F(H_{\rm i})$ is the change in the equilibrium free energy. By rewriting this identity as $P_{\rm R}(-W)=P_{\rm F}(W)\exp(\frac{-W+\Delta F}{k_{\rm B}T})$ and integrating it between $W=-\infty$ and ∞ we obtain the Jarzynski equality $\langle\exp(-W_{\rm dis}/k_{\rm B}T)\rangle_{\rm F}=1$, where $\langle\cdots\rangle_{\rm F}$ stands for a dynamical average over work values obtained along the forward process.

If we now substitute (20) into the relation (54) we obtain

$$s_{\rm F}(w_{\rm dis})-s_{\rm R}(-w_{\rm dis})=\frac{w_{\rm dis}}{k_{\rm B}T}, \tag{55}$$

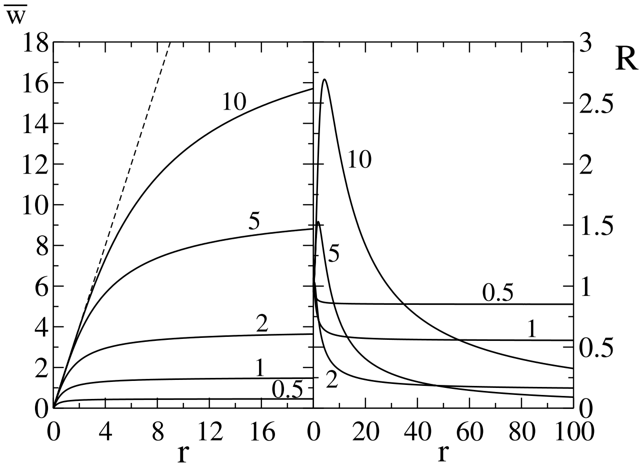

Figure 7. Average work w (left) and fluctuation-dissipation ratio R (right) for the case $H_{\rm f} = -H_{\rm i} = H_{\rm A}$ as a function of the ramping speed. We consider natural units $\mu = k_{\rm B}T = 1$. The different curves correspond to different values of the amplitude field $H_{\rm A}$. These are indicated in the plot beside each curve. The straight dashed line in the left panel corresponds to the linear-response relation $\overline{w} = 2r$ in (52). As explained in the last paragraph of section 5, and for this particular protocol, the relation $w^\dagger = -w^{\rm mp}$ is exact for all values of r and $H_{\rm A}$. Moreover, for large values of $H_{\rm A}$ the work coincides with the heat exchanged; see section 6.

where we have taken $P_{\rm F,R}(W) \propto \exp(N s_{\rm F,R}(w))$ and we have disregarded the normalization constant in the distribution (20) as unimportant. Because the quantity $s(w)$ used in (20) is only exact in the large-N limit one might be tempted to think that (55) does not hold. To prove the validity of (55) we rewrite (54) in the following way:

$$\frac{1}{N}\log(P_{\rm F}(W)) - \frac{1}{N}\log(P_{\rm R}(-W)) = \frac{W_{\rm dis}}{Nk_{\rm B}T}. \tag{56}$$

In the large-N limit the distributions (20) satisfy

$$\lim_{N\to\infty}\frac{1}{N}\log(P_{\rm F,R}(W)) = s_{\rm F,R}(W) \tag{57}$$

and therefore (55) is exact with $w_{\rm rev} = \lim_{N\to\infty}\Delta F/N$. The present approach seems quite general so the trajectory entropy derived in a large-N theory in any statistical model (interacting or non-interacting) must verify the relation (55). Another interesting relation that can be obtained from (55) relates the values of $w^{\rm mp}, w^\dagger$ for the forward and reverse processes. Differentiating (55) respect to w we obtain

$$s'_{\rm F}(w) + s'_{\rm R}(-w) = \lambda_{\rm R}(-w) - \lambda_{\rm F}(w) = \frac{1}{k_{\rm B}T}, \tag{58}$$

where we used (22). Therefore, the identity $s'_{\rm F}(w^{\rm mp}) = \lambda_{\rm F}(w^{\rm mp}) = 0$ (23) implies $s'_{\rm R}(-w^{\rm mp}) = \lambda_{\rm R}(-w^{\rm mp}) = 1/k_{\rm B}T$. From (31) we then infer that $w^\dagger$ for the reverse process is identical to $-w^{\rm mp}$ for the forward process, and vice versa. This relation is quite interesting because it suggests that in order to estimate (e.g. in experiments) the value of $w^{\dagger}$ for a given non-equilibrium process it is enough to determine w^{mp} for the reversed process, a quantity that is experimentally accessible.

An interesting case of (54) occurs whenever the forward and reverse processes are symmetrical mirror images, $H_{\mathrm{F}}(s) = -H_{\mathrm{R}}(s) = -H_{\mathrm{F}}(t-s)$. This can be accomplished when $H_{\mathrm{A}} = H_{\mathrm{f}} = -H_{\mathrm{i}}$ along the forward process and the protocol satisfies $H(s) = -H(t-s)$. In this case the forward and the reverse work distributions are identical, $W_{\mathrm{rev}} = \Delta F = 0$ (or $w = w_{\mathrm{dis}}$), and (55) reads

$$s(w) - s(-w) = \frac{w}{k_{\mathrm{B}}T}, \tag{59}$$

where we have used $s(w) = s_{\mathrm{F}}(w) = s_{\mathrm{R}}(w)$. The validity of (59) can be further demonstrated by close inspection of equations (13)–(15). Let $s(w)$ be the value of the dynamical entropy for a given value of the work w associated with the value of the Lagrange multiplier λ and the magnetization $m(s)$. Then, for the reversed value of the work $-w$, it is possible to show that the corresponding solutions are $-\lambda - 1/k_{\mathrm{B}}T$ for the Lagrange multiplier and $-m(t-s)$ for the magnetization solution. The resulting entropy is then $s(-w) = s(w) - w/k_{\mathrm{B}}T$, as given in (59). A remarkable consequence of this special case is the aforementioned fact that $w^{\dagger} = -w^{\mathrm{mp}}$ at any ramping speed and for any value of the amplitude of the field H_{A}. This case was already shown in figure 7. The present symmetric case is especially interesting because the work done upon the system can be identified with the heat exchanged between the system and the bath. The conditions required for such identification are discussed next.

6. Heat fluctuations and tails

Until now we have focused our efforts on investigating work fluctuations. However, in the same way as the work fluctuates, the heat exchanged between the system and the bath also does. The validity of the mechanical equivalence of heat (the content of the first law of thermodynamics) suggests that there should not be an important difference between heat and work. Heat is more difficult to measure experimentally than work, and this is the reason why we tend to be more interested in the latter.

A motivation to investigate heat fluctuations has recently arisen in the context of steady state and ageing systems. In the first case, heat fluctuations were investigated for the simple model of a bead dragged through a viscous fluid [17]. In the second case they were studied for the case of a spin-glass model characterized by slow dynamics and ageing [18]–[20]. In both cases, a Gaussian component was identified in the heat distributions, together with some exponential tails. For the steady state system these exponential tails were a consequence of the validity of an asymptotic fluctuation-theorem for the heat, while in the ageing system the tails were the signature of intermittency effects that have been experimentally observed in glasses and colloids [21, 22].

The heat along a given trajectory can be inferred using energy conservation, $-Q + W = \Delta E$. To extract the heat we just need to know the change in energy ΔE between the final and initial configurations as well as the work W. Here we adopt the sign convention (contrary to that adopted in section 1) that positive heat corresponds to net heat delivered by the system on the surroundings. A particular case where work and heat fluctuations are identical is the case described in the preceding section, where $H_{\mathrm{f}} = -H_{\mathrm{i}} = H_{\mathrm{A}}$. Due to time reversal symmetry $W_{\mathrm{rev}} = \Delta F = 0$. Now, if the field amplitude H_{A} is large enough, the difference in energy $\Delta E = -\mu\Delta(MH)$ is practically zero, so $Q = W$. For example, if $H_{\mathrm{A}} = 5$, then $\tanh(5) = 0.999\,91$ (as always we take $\beta = \mu = 1$), so the initial equilibrium magnetization is $M_{\mathrm{eq}}(H_{\mathrm{A}}) = -N$. The final magnetization after ramping the field to H_{A} is again of order N and therefore the fluctuations from trajectory to trajectory in ΔE are negligible as compared to the total work. In figure 8 we show the trajectory entropy and free energy for the case $H_{\mathrm{A}} = 10$. We have chosen to represent variables in terms of heat per particle $q = Q/N$ rather than work, to give a view of what general shape we can expect from heat distributions. In terms of the heat we expect that the same mathematical definitions and relations that we defined in the case of work are also valid. For instance, the heat entropy and the heat free energy are defined in the same way as we did for their work counterparts just replacing w by q; in particular, $\mathcal{F}(q) = q - k_{\mathrm{B}}Ts(q)$. Also the equivalent of (22) holds,

$$\frac{\partial s(q)}{\partial q} = -\lambda(q). \tag{60}$$

The most probable heat q^{mp} ($\lambda(q^{\mathrm{mp}}) = 0$) and the quantity $q^{\dagger}$ ($\lambda(q^{\mathrm{mp}}) = -1/k_{\mathrm{B}}T$) can also be defined. Moreover, a relation equivalent to (59) is also expected to hold for large enough values of H_{A},

$$s(q) - s(-q) = \frac{q}{k_{\mathrm{B}}T}. \tag{61}$$

An interesting aspect of the heat entropy $s(q)$ shown in the left panel of figure 8 is the presence of quadratic behaviour for small values of q ($q \simeq 0$) together with a linear behaviour in the tails ($|q| \gg 1$). These characteristic features of the heat entropy $s(q)$ can be inferred by looking at $\lambda(q)$, shown in figure 9. That figure shows that $\lambda(q)$ is linear with q for $q \simeq 0$, giving a quadratic behaviour for $s(q)$ at small values of q. This linear shape in $\lambda(q)$ corresponds to a Gaussian distribution for $P(q) = \exp(s(q))$. It also shows that for a wide range of $|q|$ values there are two plateaus at $\lambda(q) \sim \lambda_+, -\lambda_-$ for positive and negative values of q respectively. These plateaus correspond to the exponential tails in the distribution. This behaviour is quite generic at all ramping speeds; the distinction in $\lambda(q)$ between both plateaus and the linear behaviour at small q becomes more clear as the ramping speed decreases, i.e. in the low dissipation regime. In such conditions, q^{mp} is not very large and the linear response approximation holds. The Gaussian sector describes the statistics of small and most probable fluctuations; the exponential tails describe rare events and large deviations. In what follows we analyse the Gaussian and exponential tails in more detail.

In the region where both q, q^{mp} are not too large we have

$$s(q) = -\frac{1}{N\sigma_q^2}(q - q^{\mathrm{mp}})^2 \qquad q, q^{\mathrm{mp}} \ll 1. \tag{62}$$

Substituting this relation in (61) we get

$$s(q) - s(-q) = \frac{2q^{\mathrm{mp}}}{N\sigma_q^2}q = \frac{q}{k_{\mathrm{B}}T}, \tag{63}$$

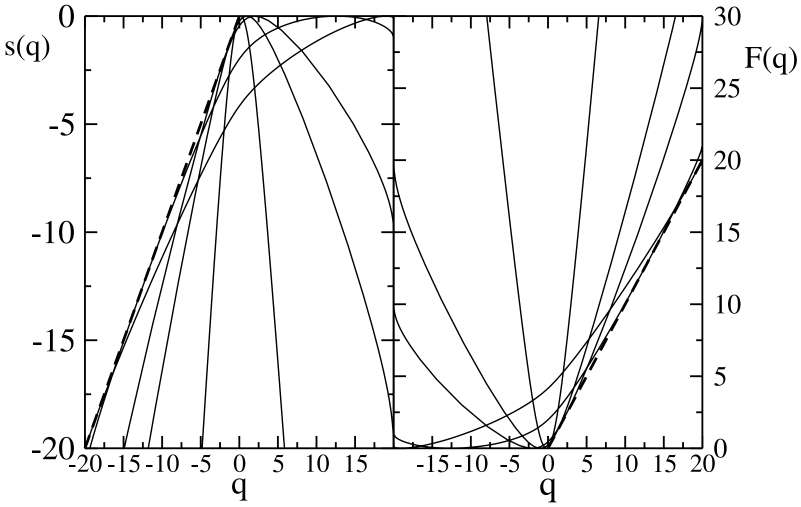

Figure 8. Trajectory entropy $s(q)$ and trajectory free energy $\mathcal{F}(q)$ for the case $H_{\rm f} = -H_{\rm i} = 10$ at ramping speeds $r = 0.1, 0.5, 1, 10, 100$ (from the most narrow to the most wide distributions). The dashed line in the left panel is $y(q) = q/k_{\rm B}T$ (we take $k_{\rm B}T = 1$) and is tangent to $s(q)$ at a value $q^\dagger$. The dashed line in the right panel corresponds to $y(q) = q$ and is tangent to $\mathcal{F}(q)$ at the value $q^{\rm mp}$. Both $q^{\rm mp}, q^\dagger$ depend on the ramping speed.

implying

$$\frac{N\sigma_q^2}{2q^{\rm mp}k_{\rm B}T} = 1. \tag{64}$$

This result shows that the fluctuation-dissipation ratio (53) is equal to 1 if heat fluctuations are restricted to the sector of q small. Small fluctuations are a key assumption of linear-response theory which also leads to (64).

This quadratic behaviour then goes over straight lines in the most negative and positive sectors of q,

$$s(q) = C - \lambda_+ q \qquad q \gg 1 \tag{65}$$

$$s(q) = C + \lambda_- q \qquad q \ll -1 \tag{66}$$

where C is a constant and λ_+, λ_- correspond to the values of $\lambda(q)$ in the plateaus shown in figure 9. Note that the constant C in (65), (66) is the same in both sectors. In fact, the relation (61) imposes such a constraint between the positive and negative tails in the probability distributions. Substituting (65), (66) into (61) we obtain

$$\lambda_- - \lambda_+ = \frac{1}{k_{\rm B}T}, \tag{67}$$

meaning that the width of the tails is larger for negative values of q than for positive values. This can be interpreted by saying that, despite the fact that the average heat q is positive, rare fluctuation events occur as often for $q < 0$ (i.e. when the system adsorbs heat from the surroundings) as they do for $q > 0$ (when the system delivers heat to the

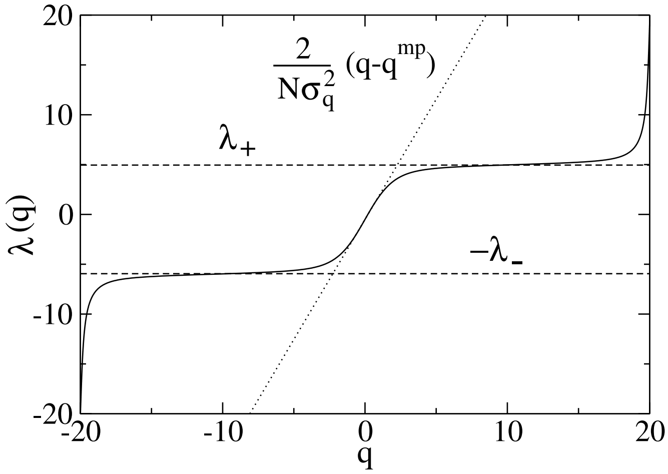

Figure 9. $\lambda(q)$ for the same parameters as in figure 8 for the case $r = 0.1$. We note the presence of a linear behaviour for $\lambda(q)$ for small values of q, $\lambda(q) = (2/N\sigma_q^2)(q - q^{\mathrm{mp}})$ ($q^{\mathrm{mp}} = 0.207, N\sigma_q^2 = 0.83$) and two plateaus for $q \gg 1$ and $q \ll -1$ ($\lambda_+ = 4.95/k_{\mathrm{B}}T, \lambda_- = 5.95/k_{\mathrm{B}}T$). The former gives rise to the Gaussian component in the heat distribution describing the statistics of most probable values. The latter gives rise to two exponential tails for the distribution describing the statistics of rare events.

surroundings). The shape of the heat distribution $P(q) = \exp(s(q))$ is then dominated by a central Gaussian distribution with exponential tails at its extremes. These features are illustrated in figure 10. In the quasi-reversible limit, $r \to 0$, the height of the two plateaus grows (to the leading order) like $1/2rk_{\mathrm{B}}T$; the difference between them is given by (67).

What is the physical interpretation of the plateaus observed in figure 9? As we already mentioned in (35), the values of the two plateaus, λ_- and λ_+, allow us to characterize the non-equilibrium process by two temperatures,

$$T_- = \frac{1}{k\lambda_-}; \qquad T_+ = -\frac{1}{k\lambda_+} \tag{68}$$

with $T_- > 0, T_+ < 0$. The fact that these two temperatures emerge in the quasi-reversible limit suggests a possible physical interpretation. As we already emphasized, the plateaus observed are related to the presence of exponential tails in the work (heat) distributions. Interestingly, a link between the presence of exponential tails in heat distributions and effective temperatures (describing violations of the fluctuation-dissipation theorem; see [37] for reviews) has been recently found in the context of glassy systems in the ageing regime [18]. In that context, the value of the effective temperature has been related to the width of the exponential tail in the heat distribution, and interpreted in terms of a microcanonical *protomeasure* that is self-generated by the dynamics. This suggests that λ_- and λ_+ could be the fingerprint of the emergence of a dynamical measure in the quasi-reversible limit. This conclusion is also consistent with the fact that ageing systems are characterized by a low-entropy production [38], reinforcing the idea that the concept of the effective temperature can be partially rescued in low-dissipation non-equilibrium regimes. Given the extreme simplicity of the model under study one may wonder whether this

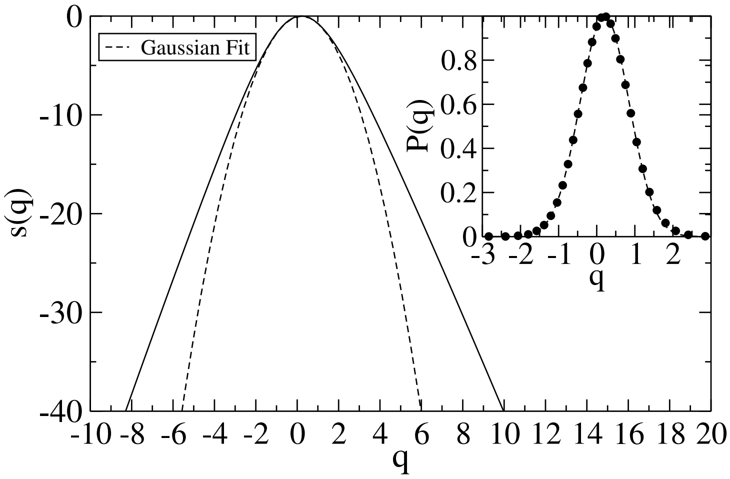

Figure 10. Heat entropy $s(q)$ for the case $H_{\mathrm{f}} = -H_{\mathrm{i}} = 10$ at ramping speed $r = 0.1$. Main figure: the sector of small or most probable fluctuations $q \sim 0$ can be well fitted to the Gaussian (62) (dashed curve) with parameters $q^{\mathrm{mp}} = 0.207, N\sigma_q^2 = 0.83$ satisfying (64). The tails extend beyond the Gaussian central part and are of exponential type as described in (65), (66) with $\lambda_- = 5.95/k_{\mathrm{B}}T, \lambda_+ = 4.95/k_{\mathrm{B}}T, C = 9.16$. These exponential tails describe the statistics of large deviations or rare events. Inset: heat-distribution $P(q) = \exp(s(q))$ (dots) and the Gaussian fit (the dashed curve of the main figure) showing that the small q sector of fluctuations (those that are frequently observable) is very well fitted by a Gaussian despite the fact that rare-event tails are big and observable only when plotting $s(q)$ or the distribution $P(q)$ in logarithmic scale.

conclusion is valid in general for non-equilibrium systems in asymptotically low dissipation regimes.

Finally we want to mention that if the amplitude field H_{A} is not large enough, then there may be contributions to the heat distribution coming out from the fluctuations in the difference in energy between the initial and final configurations. The effect of the value of H_{A} on the value of the average work and the fluctuation-dissipation ratio have been already shown in figure 7; in particular, non-monotonic behaviour is observed for R.

7. Finite-size effects

The method we have developed in this paper allowed us to calculate $P_N(w)$ in the large-N limit. However, due to the non-interacting character of the model, all cumulants of the distribution obtained in the large-N limit are also exact for finite N. The proof is quite straightforward. Let us define the generation function of all cumulants of the distribution $P_N(W)$ in (4),

$$g_N(x) = \log(\overline{\exp(xW)}) = \log\left(\int \mathrm{d}W\ \exp(xW)P_N(W)\right). \tag{69}$$

Cumulants of $P_N(W)$ are obtained using the following formula:

$$c_N(k) = \frac{\partial^k g_N(x)}{\partial x^k}\bigg|_{x=0}, \tag{70}$$

k being a positive integer. Using the non-interacting character of the model, then we can write

$$W = \sum_{i=1}^{N} w_i \rightarrow g_N(x) = N g_1(x) \rightarrow c_N(k) = N c_1(k) \tag{71}$$

and therefore all cumulants of the distribution are independent of the size of the system (except by a multiplicative constant equal to N). This implies that the expression given for $\overline{w_{\rm dis}}$ in (49) and R in (53) are independent of N. Therefore, the results we obtained in the large-N limit are *exact for any finite value of* N.

However, albeit cumulants do not depend on N, the shape of the distribution $P_N(w)$ in (20) depends on the size N and only in the large-N limit does the approximate distribution (20) become exact. For instance, the value of $w^{\rm mp}$ depends on N and converges to $\overline{w_{\rm dis}}$ for large enough values of N. In practice, already for $N = 5$–10 convergence of the approximate distribution (20) to the exact result is excellent. In order to compare the approximate distributions we obtain from our theory with the exact ones at finite N we have done numerical simulations of the model. The simulation procedure is described below in section 8. In figure 11 we show the distributions we have obtained for $N = 10$ compared to the numerical simulations at different ramping speeds. The agreement between theory (continuous lines) and simulations (symbols) is good, although it worsens progressively as the ramping speed increases and the system is strongly driven out of equilibrium. An important feature of the distributions is observed for large r and small N: the presence of a finite fraction of trajectories that dissipate a maximum amount of work equal to $w_{\rm max} = \mu(H_{\rm f} - H_{\rm i})$. For these trajectories the ramping speed is so high and the size so small that no change in the initial configuration occurs along the trajectory. We will call these *trapped* trajectories. The fraction of trapped trajectories contributes with a term $\delta(w - w_{\rm max})$ to the work distribution,

$$P_N(w) = \tilde{P}_N(w) + \alpha(N)\delta(w - w_{\rm max}), \tag{72}$$

where $\tilde{P}_N(w)$ is a continuous function and $\alpha(N)$ is a size-dependent constant that decreases with N and asymptotically vanishes in the large-N limit. The delta function in (72) is a *small*-N contribution that is not captured by the present large-N theory. Nevertheless, it might be analytically derived using the approach described below in section 7.1.

In figure 12 we show the effect of the size on the distributions at a moderate ramping speed. For $N = 1$ the agreement is not good, although the behaviour of the left tails is reasonably well reproduced. However, already for $N = 5$ the agreement has improved considerably. We conclude that it is between $N = 1$ and 5 that finite-size effects are important. In figure 13 we confirm this strong *small*-N dependence by plotting the most probable work as obtained from the numerical simulations as a function of r for different sizes $N = 1, 2, 5, 20$.

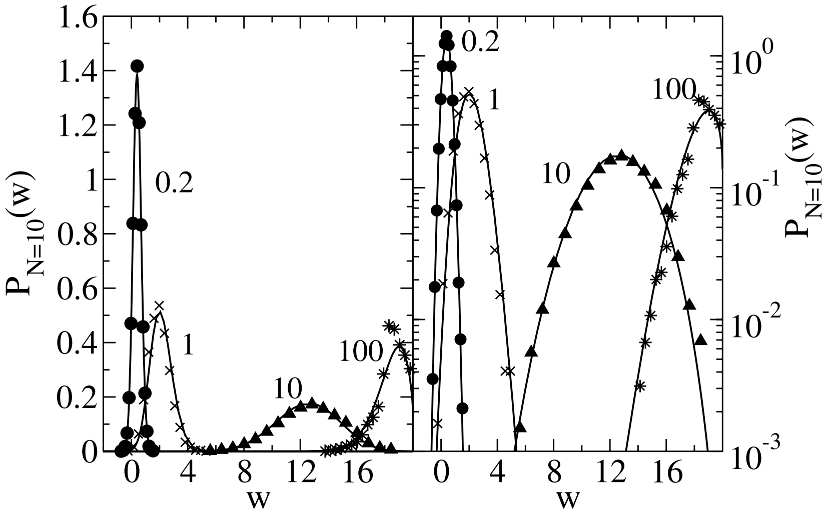

Figure 11. Distributions $P_N(w)$ for the case $H_\mathrm{f} = -H_\mathrm{i} = 10$ and $N = 10$ at different ramping speeds (indicated along each curve). The continuous lines are the results obtained from the present theory using (20). The symbols are results obtained from numerical simulations of the model for 10^4 trajectories. The right panel is the same figure but in logarithmic scale. For the largest ramping speed $r = 100$ there is a finite fraction of trajectories (about 37% of the total number of trajectories) where the spins have no time to relax. These trajectories contribute with a singular term at $w = w_\mathrm{max} = 20$ to the distribution $P_{N=10}(w)$ as described in (72). It cannot be captured by the present large-N theory so we did not include it in the histogram obtained from the numerical simulations.

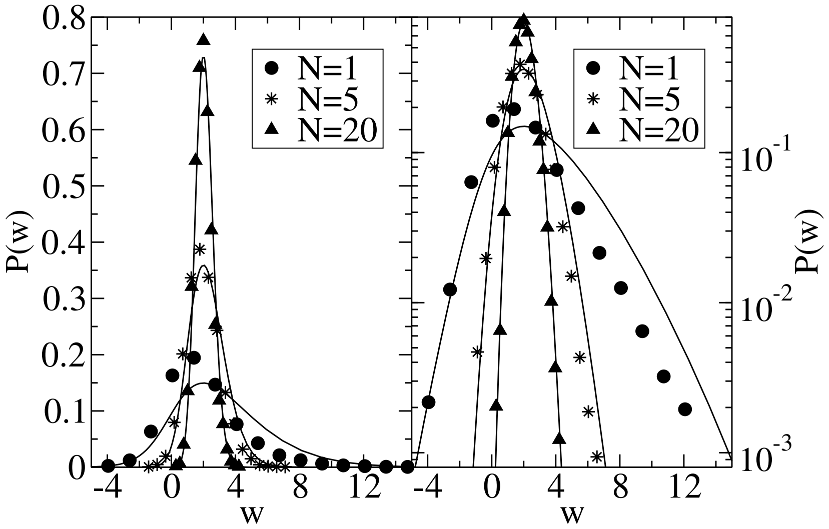

Figure 12. The same as in figure 11 but showing the dependence of $P_N(w)$ with N at $r = 1$. The agreement is not good for $N = 1$ in the central region of the distribution but is reasonably good for the left tail of the distribution (see the right plot in logarithmic scale). The agreement improves noticeably beyond $N \simeq 5$.

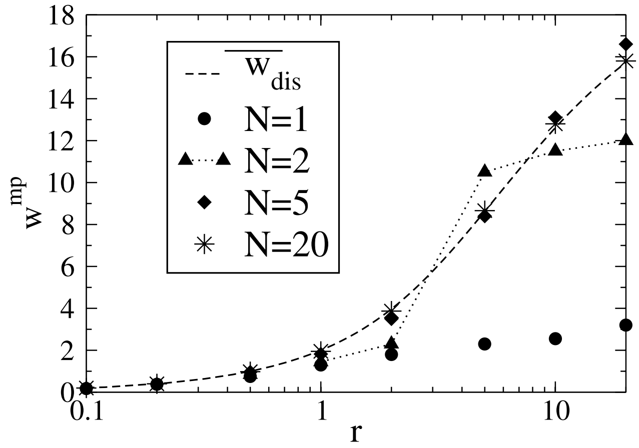

Figure 13. The same parameters as in figure 11 but now showing the dependence of the most probable work value w^{mp} with N. The different symbols in the points correspond to different sizes as indicated in the legend of the figure. The continuous dashed curve is w^{mp} as derived from the theory in the large-N limit (where $w^{\mathrm{mp}} = \overline{w_{\mathrm{dis}}}$, the latter being independent of N). In the linear-response regime, $r \ll 1$, data have converged to the theory for all sizes. Although finite-size effects are large for $N = 1, 2$ and $r \gg 1$, already for $N = 5$ the simulations have converged to the theory at all ramping speeds. Data for $N = 2$ have been connected by a dotted line to emphasize the sharp increase of w^{mp} around $r \simeq 5$. This sharp increase originates from the presence of two separated peaks in the work distribution whose heights coincide at a given value of the ramping speed around $r \simeq 5$.

7.1. Reconstructing $P_1(w)$ from the large-N theory

A crucial aspect of the present model is that it is non-interacting. Therefore, if we were to know the work probability distribution for $N = 1$ (a single spin) then we could reconstruct the general distribution $P_N(w)$. In fact, let $\hat{P}_N(s)$ denote the Laplace transform of $P_N(w)$,

$$\hat{P}_N(s) = \int_0^\infty \exp(-Ws) P_N(W)\, \mathrm{d}W. \tag{73}$$

Using the result $W = \sum_{i=1}^N w_i$ we can write

$$\hat{P}_N(s) = (\hat{P}_1(s))^N, \tag{74}$$

allowing us to reconstruct $P_N(w)$ from the sole knowledge of $P_1(w)$. Although the analytical computation of $P_1(w)$ might be possible by using other approaches, throughout this paper we have considered a *thermodynamic approach* where the large-N theory has been taken as an approximation to finite N. This approach turns out to give exact results for all cumulants of the distribution, thereby suggesting that the reconstruction of $P_1(w)$ from $P_N(w)$ might be possible. One could naively think that this is possible just using (74) together with the knowledge of $P_N(w)$. Unfortunately, this is not the case, as the knowledge of $P_N(w)$ is only approximate, as we showed in the previous section.

There is, however, a possible strategy to reconstruct $P_1(w)$ that is based on the fact that cumulants are exactly known. Let us define the following function:

$$h(x) = \lim_{N\to\infty} \frac{g_N(x)}{N}, \tag{75}$$

where $g_N(x)$ was defined in (69). In the large-N limit we can solve $h(x)$ by applying the saddle-point approximation,

$$\begin{aligned} h(x) &= \lim_{N\to\infty} \frac{1}{N} \log(\overline{\exp(xW)}) \\ &= \lim_{N\to\infty} \frac{1}{N} \log\left(\int \mathrm{d}W\, \exp(xW)\exp(Ns(w))\right) = xw(x) + s(w(x)), \end{aligned} \tag{76}$$

where $w(x)$ is the solution of the equation

$$\left.\frac{\mathrm{d}s(w)}{\mathrm{d}w}\right|_{w=w(x)} = -x. \tag{77}$$

For a given value of w, (22) shows that $x = \lambda(w)$. For instance, for $x = 0, -1/k_{\mathrm{B}}T$ we get $w = w^{\mathrm{mp}}, w^{\dagger}$ respectively. Therefore, we can express $h(x)$ in terms of w rather than x:

$$h(w) = w\lambda(w) + s(w). \tag{78}$$

By inserting (71) in (75) we get $g_1(w) = h(w)$, and therefore

$$\int \mathrm{d}w'\, \exp(\lambda(w)(w'-w))P_1(w') = \exp(s(w)). \tag{79}$$

Formally, this integral equation is closed and provides an exact solution for $P_1(w)$ in terms of the entropy $s(w)$. Unfortunately we have been unable to solve it in full generality (as detailed knowledge of the solution in (77) is required). Yet, for $P_1(w)$ it still holds that there are exponential tails identical to those we already derived for $P_N(w)$ in the large-N limit. To show this result we use (22) and rewrite (79) as follows:

$$\int \mathrm{d}w'\, \exp\left(-s(w) - \frac{\partial s(w)}{\partial w}(w'-w)\right)P_1(w') = 1. \tag{80}$$

Let us now suppose now that $\lambda(\omega)$ is approximately constant (equal to $\hat{\lambda}$) showing a plateau over a given region of work values. From (22) then $s(w') \simeq s(w) + (\partial s(w)/\partial w)(w' - w)$, and

$$P_1(w') \propto \exp(s(w')) = C\exp(\hat{\lambda}w'), \tag{81}$$

where C is a constant. This shows that the width of the exponential tail for $P_1(w)$ (and, by extension, for $P_N(w)$ at any value of N) is equal to $\hat{\lambda}$.

8. The case of magnetic nanoparticles

In this section we discuss a system where the previous theory could be experimentally tested. We focus our attention on thermally activated magnetic nanoparticle systems [23]. These systems have the great advantage that the dynamics is invariant under time-reversal of the magnetic field $H \rightarrow -H$. Also many magnetic field cycles can be experimentally realized in micro-SQUID machines allowing one to experimentally extract the work distribution with good precision. The main experimental limitation to observe WF though is the quite large value of the magnetic moment of the nanoparticle. Transition rates are described by the Brown–Neel formula,

$$\tau_{\text{relax}}(H) = \tau_0 \exp\left(\frac{B(H)}{k_{\text{B}}T}\right), \tag{82}$$

where τ_0 is a microscopic time describing relaxation within a state and $B(H)$ is a field dependent barrier. We consider two cases: (A) paramagnetic molecular clusters where the energy barrier is nearly field independent $B(H) = B_0$ (this case could also describe specific ferro and ferrimagnetic nanoparticles where the anisotropy contribution to the zero-field barrier is negligible; for a discussion see [24]); (B) ferromagnetic nanoparticles with axial anisotropy where $B(H)$ depends on the intensity of the external field projected on the easy magnetization axis as described by the Stoner–Wohlfarth expression $B(H) = B_0(1-|H/H_{\text{c}}|)^{\alpha}$, where H_{c} is the field required to suppress the barrier and α is an exponent in the range 1.5–2. Recent experiments have demonstrated how the height of the barrier B_0 can be considerably reduced by applying a transverse field, making it possible to observe magnetization reversible transitions (also called telegraph noise measurements) in single Co nanoparticles at low temperatures [25, 26].

As we already discussed in section 5, in a magnetic system a time-reversal invariant protocol can be accomplished by switching the magnetic field H from $-H_{\text{A}}$ to H_{A} (H_{A} being the amplitude of the field), the free energy and the rates being an even function of H. Under such conditions work and heat are equivalent if H_{A} induces a magnetization close to its saturation value. From the experimental point of view, it is relevant to understand under which conditions large deviations from the most probable work are observable. By large deviations we understand work (heat) fluctuations corresponding to work (heat) values around $w^{\dagger}$ ($q^{\dagger}$). A useful quantity that tell us how difficult it is to sample that region of work values is the ratio Ω describing the fraction of trajectories that transiently violate the second law, $w \leq 0$. This fraction is given by the integrated fluctuation theorem [2, 5]. This is obtained by rewriting (54):

$$P_N(-W) = P_N(W) \exp\left(-\frac{W - \Delta F}{k_{\text{B}}T}\right) = \exp\left(-\frac{W_{\text{dis}}}{k_{\text{B}}T}\right), \tag{83}$$

where we have taken $P_N(W) = P_{\text{F}}(W) = P_{\text{R}}(W)$. Integrating this expression from $W = 0$ up to $W = \infty$, we obtain

$$\Omega = \frac{\mathcal{N}(w<0)}{\mathcal{N}(w>0)} = \left\langle \exp\left(-\frac{Nw}{k_{\text{B}}T}\right)\right\rangle_{w>0}, \tag{84}$$

where $\mathcal{N}(w<0), \mathcal{N}(w>0)$ are the fraction of trajectories for which the total work is negative and positive respectively,

$$\mathcal{N}(w<0) = \int_{-\infty}^{0} \mathrm{d}W\, P(W); \qquad \mathcal{N}(w>0) = \int_{0}^{\infty} \mathrm{d}W\, P(W), \tag{85}$$

and the average on the rhs of (84) is restricted to the subset of trajectories for which $w>0$. Quite generally, we expect that Ω is a non-universal function dependent on all cumulants of $s(w)$, yet its exponential dependence in N assures that, in the regime where TV are observable, Ω is approximately described by the value of the average total work divided by the bath temperature $\overline{W}/k_{\mathrm{B}}T$, which is approximately given by $Nw^{\mathrm{mp}}/k_{\mathrm{B}}T$. In ramping experiments one can measure magnetization curves (i.e. M versus H curves). The experimental measure of Ω would be straightforward by repeatedly measuring the magnetization curves of individual nanoparticles over many cycles. According to (84), to evaluate Ω we just require to count the fraction of trajectories where the area under the curve of magnetization is positive (or negative).

We choose Glauber rates as these have been experimentally demonstrated to describe very well the relaxation of single magnetic moments [27, 28]. These are given by (36), where $\tau_{\mathrm{relax}}(H)$ is given by (82). We consider ramping experiments [29] where N particles are subject to the action of a field that is switched from $H=-H_{\mathrm{A}}$ up to $H=H_{\mathrm{A}}$ at a constant speed $\dot{H}$. We generate individual trajectories according to the Glauber rates by starting from initial configurations with $M=M_{\mathrm{eq}}(H_{\mathrm{A}})$ and repeating the ramping protocol many times; each time the total work (3) is computed, $W=-\mu\int_{-H_{\mathrm{A}}}^{H_{\mathrm{A}}} M(H)\,\mathrm{d}H$. If H_{sw} is the field at which the magnetization of a given particle switches for the first time then, for a given trajectory, some of the particles will switch state at a value of the field $H_{\mathrm{sw}}<0$, while others will switch at $H_{\mathrm{sw}}>0$. For fast ramping speeds the dynamics is well described by a first-order Markov process [30] and the dissipated work for that trajectory will be identical to the value $2H_{\mathrm{sw}}$ averaged over all particles. In general, for lower ramping speeds, the relation between the dissipated work and the value of H_{sw} is more complicated. To estimate Ω we generate trajectories and evaluate the fraction of them with $W>0$ and $W<0$. We chose to do numerical simulations rather than applying the large-N theory to give a more clear picture about which results can we expect from a finite number of ramping experiments (around 10 000). In the main panel of figure 14 we plot the value of Ω (84) as obtained for different ramping protocols in cases (A) and (B). All points scatter around a generic (but non-universal) curve useful to predict in which regime TV are expected to be observable. An important advantage of the time-reversal symmetry property $H\rightarrow -H$ of magnetic nanoparticle systems, as compared to other systems [7, 8], is the feasibility of performing many ramping cycles in a single experiment making TV observable for Ω values as low as 10^{-4}. According to figure 14, TV should be observable for work values as large as $20k_{\mathrm{B}}T$.

9. Conclusions

Two-state systems provide a simple conceptual framework to analyse work fluctuations (WF) and transient violations (TV) of the second law. These non-equilibrium effects are expected to be relevant and observable for nanosized objects when the energies involved are several times $k_{\mathrm{B}}T$, k_{B} being the Boltzmann constant and T the temperature of

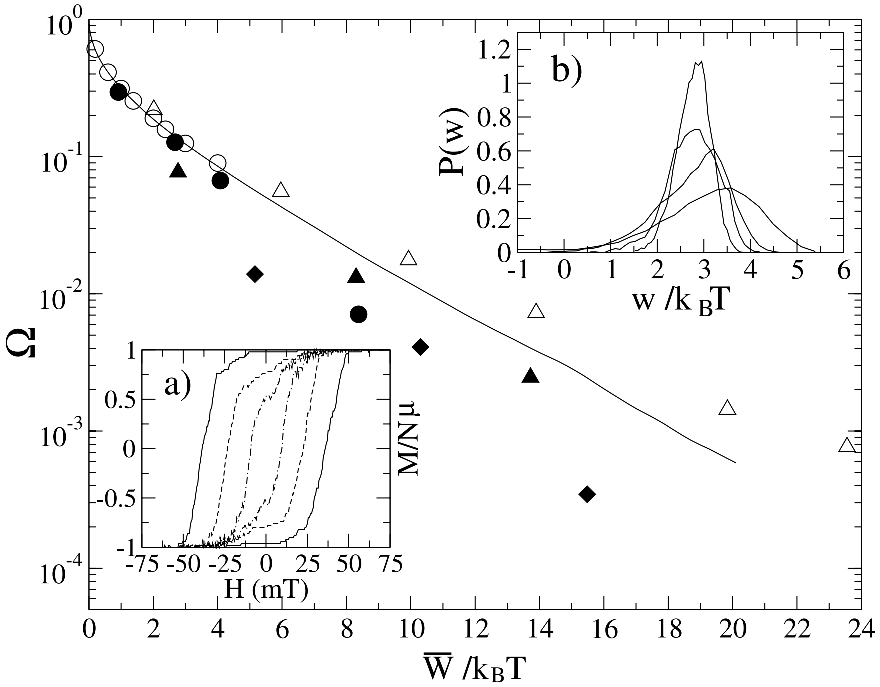

Figure 14. Application of the theory to magnetic nanoparticles for a case where $w_{\rm rev} = 0$ and $w \simeq q$. Main panel: Ω (84) as a function of $\overline{W}/k_{\rm B}T$ for cases (A) and (B). Number of particles range from $N = 1$ up to 20 for both cases. Case (A) corresponds to nanoparticles (open symbols) with $\mu = 300\ \mu_{\rm B}, T = 200$ K, $H_{\rm A} = 2$ T, $\tau_0 = 10^{-7}$ s, $B_0 = 2300$ K at ramping speeds $r = 1$ mT s^{-1} (circles), 10 mT s^{-1} (triangles up). Case (B) corresponds to ferromagnetic particles (filled symbols) with $\mu = 500\ \mu_{\rm B}, T = 40$ K, $H_{\rm A} = 238$ mT, $H_{\rm c} = 119$ mT, $B_0 = 500$ K, $\tau_0 = 10^{-5}$ s, $\alpha = 1.5$ and ramping speeds of 0.01 (circles), 0.1 (triangles up), 1 (diamonds) T s^{-1} (continuous, dashed and dotted curves respectively). The continuous curve is the prediction for a Gaussian work distribution (see the discussion at the end of section 3) in the linear-response regime $\sigma_w^2 = 2k_{\rm B}T\overline{w_{\rm dis}}$ ($R = 1$ in (53)). Insets: both are for ferromagnetic nanoparticles (case (B)) with the same parameters as in the main panel. Inset (a) shows hysteresis cycles for $N = 100$ ferromagnetic nanoparticles at the three ramping speeds (larger hysteresis for higher speeds). Inset (b) shows dissipated work distributions at ramping speeds 0.1 T s^{-1} for $N = 1, 5, 10, 20$ particles (larger sizes correspond to narrower distributions).

the bath. These have been already observed in the unfolding of small RNA hairpins [7] as well as in polystyrene beads dragged through a solvent [8]. Related measurements include the experimental test of the Gallavotti–Cohen fluctuation theorem in Rayleigh–Bernard convection [9] and turbulent flows [10]. Other experiments include the observation of gravitatory potential energy fluctuations in driven granular media [31]. The scientific discipline behind all such rich phenomenology deserves to be called *thermodynamics of small systems*. It deals with the thermal behaviour of non-equilibrium small systems where the typical energies are few times $k_{\rm B}T$. The statistics of energy exchange processes between the system and the thermal environment is described by frequent Gaussian distributed events plus rare events corresponding to large statistical deviations from the average value. The theoretical and experimental study of these fluctuations could be of relevance to understanding issues related to the organization and function of biological matter in the nanoscale [32].

In this paper we studied WF in two-state systems. This case is particularly relevant for two reasons. First, it is the simplest example of a non-linear system describing thermally activated processes. This inherent simplicity has indeed motivated several studies of this model in the context of glassy dynamics (e.g. in [33]). Second, the non-interacting character of the energy function makes the model exactly solvable. The model being non-interacting, one may wonder how relevant this model is for describing the non-equilibrium behaviour observed in real systems (e.g. in [9, 10]) characterized by spatial correlations. Our answer risks being somewhat superficial; however, we believe that by studying the complex behaviour of such systems we can learn under which conditions the observed non-equilibrium behaviour must be interpreted as a necessary result of a complicated underlying structure. Even more important, this study paves the way to consider this phenomenology in more complex interacting systems.

The main results of our study can be summarized as follows. We have introduced a trajectory thermodynamics formalism with the specific aim to quantify WF in such model. We have shown how to define a trajectory entropy $s(w)$ that characterizes WF around the most probable value w^{mp}, and a trajectory free energy $\mathcal{F}(w)$ whose minimum value at $w = w^{\dagger}$ specifies the value of the work that needs to be efficiently sampled to quantitatively test the Jarzynski equality. The theory requires the introduction of a Lagrange multiplier $\lambda(w)$, its inverse playing the role of a temperature in the trajectory thermodynamics formalism. Analytical expressions for the trajectory potentials $s(w), \mathcal{F}(w)$ have been also derived. In general, both values w^{mp} and $w^{\dagger}$ are of the same magnitude but opposite sign, meaning that large deviations of WF need to be sampled to recover equilibrium free energies from non-equilibrium measurements, e.g. by using the Jarzynski equality.

We have then carried out a systematic study of WF in the framework of the large-N theory. Several results are worth mentioning. First of all, we have found an analytical expression for the trajectory entropy that satisfies the fluctuation theorem by Crooks [5] that relates forward and reverse processes. An important result is that the value of the work $w^{\dagger}$ that has to be sampled in order to test the Jarzynski equality is equal to the most probable value of the work (with a minus sign) for the reverse process. Intuitively this means that the forward and reverse distributions must overlap each other in order to get good estimates of the work using the Jarzynski equality, a result that was emphasized a long time ago by Bennett [34]. Furthermore, if both forward and reverse processes are symmetric mirror images then $w^{\mathrm{rev}} = 0$ and $w^{\dagger} = -w^{\mathrm{mp}}$ independently of how far the system is driven out of equilibrium. This last case is particularly interesting as the total work practically coincides with the heat. The fluctuation theorem by Crooks is then also applicable to the heat in that limit, a result that is quite reminiscent of a heat fluctuation theorem recently derived [17, 35]. For the heat distribution, we find that it is described by a central Gaussian distribution describing *local equilibrium*, i.e. with $R = 1$, and long exponential tails with widths described by the Lagrange multiplier $\lambda(w)$, which plays the role of the inverse of a temperature. Strictly speaking, because the temperature must be a positive quantity, only the tails in the negative sector $q \ll -1$ where λ is negative admit such an interpretation (i.e. in the sector of WF dominated by TV). It has not escaped our attention that this temperature could be related to other non-equilibrium temperatures that have been defined in other contexts [36, 37], such as steady-state [39] or ageing systems [18].

Our study raises the following question: to what extent are work and heat fluctuations equivalent? We already emphasized in section 6 that work and heat should be equivalent; at least this is the underlying content of the first law of thermodynamics. However, from the perspective provided by the present analysis, some important differences can be underlined. Exponential tails are more often observed in the heat rather than in the work. Such a result has been explicitly shown in the case of a bead dragged through a fluid [17] where the work is clearly Gaussian distributed while the heat displays exponential tails. However, in that case the origin of this difference lies in the fact that the motion for the bead is described by a stochastic linear equation which in general might not be the case. The difference between heat and work has its root at the true microscopic definition of these quantities. Heat is identical to work when the final energy of the system is constrained be identical to the initial value (i.e. $Q = W$ if $\Delta E = 0$; for the heat we adopt the sign convention of section 6). The simplest interpretation is that exponential tails in the work distribution are always present if the model is non-linear by definition (which is not the case for the aforementioned case of the bead dragged through the fluid). However, work distributions always tend to be masked by a Gaussian contribution arising from the Gaussian fluctuations that characterize the initial equilibrium state. Therefore, only when thermal fluctuations in the initial and final states are negligible as compared to the total amount of work along the trajectory are the measured work distributions paralleled by the heat distributions, and tails can be observed. This explains the qualitative difference observed between the functions $\lambda(w)$ in figure 9 and the right panel in figure 3. In the latter, Gaussian fluctuations in the energy of the initial and final configurations tend to mask the presence of the exponential tails.

We also studied finite-size effects to test how good the large-N theory is and provided a strategy to re-derive the finite-N work distribution from the large-N result. An important conclusion is that the large-N theory accounts for the existence of exponential tails also at finite N, the value of the widths λ_+, λ_- (corresponding to the plateaus in $\lambda(w)$) being independent of N. In addition, we applied the theory to magnetic nanoparticle systems which provide an experimental realization of two-state systems. We studied under which conditions the theory can be experimentally tested. Our results suggest that WF and TV should be observable whenever average work values are not much larger than $20k_{\mathrm{B}}T$. It is realistic to say that we are currently at the limit of the resolution of current micro-SQUID devices for the detection of single small magnetic moments (around a few hundreds of μ_{B}). Surely, we will see developments in the near future and experimental measurements of WF in magnetic systems, as well as the test of the present theory, will become possible.

Acknowledgments

The author acknowledges the warm hospitality of Bustamante and Tinoco labs at UC Berkeley where this work was done. He is grateful to C Bustamante, J Gore, C Jarzynski, A Labarta, M Palassini and I N Tinoco for useful discussions. This work has been supported by the David and Lucile Packard Foundation, the Spanish Ministerio de Ciencia y Tecnología Grant BFM2001-3525 and Generalitat de Catalunya.

References

[1] Evans D J, Cohen E G D and Morriss G P, 1993 *Phys. Rev. Lett.* **71** 2401

[2] Subsequent work has been reviewed in Evans D and Searles D, 2002 *Adv. Phys.* **51** 1529

[3] Jarzynski C, 1997 *Phys. Rev. Lett.* **78** 2690

[4] Kurchan J, 1998 *J. Phys. A: Math. Gen.* **31** 3719

[5] Crooks G E, 1998 *J. Stat. Phys.* **90** 1481
Crooks G E, 2000 *Phys. Rev.* E **61** 2361

[6] Hummer G and Szabo A, 2001 *Proc. Nat. Acad. Sci.* **98** 3658

[7] Liphardt J, Dumont S, Smith S B, Tinoco I Jr and Bustamante C, 2002 *Science* **296** 1832

[8] Wang G M, Sevick E M, Mittag E, Searles D J and Evans D J, 2002 *Phys. Rev. Lett.* **89** 050601

[9] Ciliberto S and Laroche C, 1998 *J. Physique* IV **8** 215

[10] Ciliberto S, Garnier N, Hernandez S, Lacpatia C, Pinton J-F and Ruiz Chavarria G, 2003 *Preprint* nlin.CD/0311037v2

[11] Laughlin R B, Pines D, Schmalian J, Stojkovic B P and Wolynes P, 2000 *Proc. Nat. Acad. Sci.* **97** 32

[12] Ritort F, Bustamante C and Tinoco I Jr, 2002 *Proc. Nat. Acad. Sci.* **99** 13544

[13] Gallavotti G and Cohen E G D, 1995 *J. Stat. Phys.* **80** 931

[14] Zuckerman D M and Woolf T B, 2002 *Chem. Phys. Lett.* **351** 445
Zuckerman D M and Woolf T B, 2002 *Phys. Rev. Lett.* **89** 180602

[15] Gore J, Ritort F and Bustamante C, 2003 *Proc. Nat. Acad. Sci.* **100** 12564

[16] Mazonka O and Jarzynski C, 1999 *Preprint* cond-mat/9912121

[17] Van Zon R and Cohen E G D, 2003 *Phys. Rev. Lett.* **91** 110601
Van Zon R and Cohen E G D, 2003 *Phys. Rev.* E **67** 046102

[18] Crisanti A and Ritort F, 2004 *Europhys. Lett.* **66** 253

[19] Ritort F, 2004 *Proc. Workshop 'Unifying Concepts in Granular Media and Glasses'* ed A Coniglio, A Fierro, H J Hermann and M Nicodemi (Amsterdam: Elsevier) [cond-mat/0311370]

[20] Ritort F, 2004 *J. Phys. Chem.* B **108** 6893

[21] Buisson L, Bellon L and Ciliberto S, 2003 *J. Phys.: Condens. Matter* **15** S1163

[22] Cipelletti L *et al*, 2003 *J. Phys.: Condens. Matter* **15** S257

[23] Wernsdorfer W, 2001 *Adv. Chem. Phys.* **118** 99

[24] Gonzalez-Miranda J M and Tejada J, 1994 *Phys. Rev.* B **49** 3867

[25] Wernsdorfer W, Bonet-Orozco E, Hasselbach K, Benoit A, Barbara B, Demoncy N, Loiseau A, Pascard H and Mailly D, 1997 *Phys. Rev. Lett.* **78** 1791

[26] Jamet M, Wernsdorfer W, Thirion C, Dupuis V, Mélinon P and Pérez A, 2001 *Phys. Rev. Lett.* **86** 4676

[27] Caneschi A *et al*, 2001 *Preprint* cond-mat/0106224

[28] Coulon C *et al*, 2004 *Preprint* cond-mat/0404620

[29] Kurkïjarvi J, 1972 *Phys. Rev.* B **6** 832
Gunther L and Barbara B, 1994 *Phys. Rev.* B **49** 3926
Garg A, 1995 *Phys. Rev.* B **51** 15592

[30] Hanggi P, Talkner P and Borkovec M, 1990 *Rev. Mod. Phys.* **62** 251

[31] Feitosa K and Menon N, 2004 *Phys. Rev. Lett.* **92** 164301

[32] Ritort F, 2003 *Seminaire Poincaré* **2** 193
Ritort F, 2004 *Preprint* cond-mat/0401311

[33] Perez-Madrid A, Reguera D and Rubí J M, 2003 *Physica* A **329** 357

[34] Bennett C H, 1976 *J. Comput. Phys.* **22** 245

[35] Van Zon R, Ciliberto S and Cohen E G D, 2003 *Preprint* cond-mat/0311629

[36] For a review, Casas-Vazquez J and Jou D, 2003 *Rep. Prog. Phys.* **66** 1937

[37] For reviews, Cugliandolo L F, 2002 *Preprint* cond-mat/0210312
Crisanti A and Ritort F, 2003 *J. Phys. A: Math. Gen.* **36** R181

[38] Cugliandolo L F, Dean D S and Kurchan J, 1997 *Phys. Rev. Lett.* **79** 2168

[39] Zamponi F, Ruocco G and Angelani L, 2004 *Preprint* cond-mat/0403579