

# Jarzynski equality for a gas particle driven out of equilibrium

T G Philbin and J Anders

Department of Physics and Astronomy, University of Exeter, Stocker Road, Exeter EX4 4QL, UK.

E-mail: [t.g.philbin@exeter.ac.uk](mailto:t.g.philbin@exeter.ac.uk)

**Abstract.** One particle in a classical perfect gas is driven out of equilibrium by changing its mass over a short time interval. The work done on the driven particle depends on its collisions with the other particles in the gas. This model thus provides an example of a non-equilibrium process in a system (the driven particle) coupled to an environment (the rest of the gas). We calculate the work done on the driven particle and compare the results to Jarzynski's equality relating a non-equilibrium work process to an equilibrium free-energy difference. The results for this model are generalised to the case of a system that is driven in one degree of freedom while interacting with the environment through other degrees of freedom.

PACS numbers: 05.70.Ln, 05.20.Dd, 51.30.+i

## 1. Introduction

Relations between equilibrium thermodynamic quantities and the work done in non-equilibrium processes are standard textbook material [1]. For example, in a non-equilibrium process the average work done  $\langle w \rangle$  on a thermal ensemble of systems with a fixed volume obeys ([1], § 20)

$$\langle w \rangle \geq \Delta F, \quad (1)$$

where  $\Delta F$  is the change in the Helmholtz free energy of the system after the final state equilibrates at the same temperature as the initial state. This inequality follows from the second law of thermodynamics [1]. The process by which work is done on the system corresponds to a time variation of some of its state parameters and this variation must be identical for each ensemble element. Because of the meaning of  $\Delta F$  in (1), one often envisages the experimental system re-equilibrating after the work is done, so that the final state, as well as the initial state, is in thermal equilibrium. This assumption of a final equilibrium state is not necessary in order to use (1), however: the work can be performed and measured without letting the system re-equilibrate and (1) still gives a prediction regarding the average measured work that can be verified if one can calculate the equilibrium free-energy difference.

In recent times fluctuation theorems have generalised this inequality to a set of non-equilibrium *equalities*. Detailed fluctuation relations show that the probability distributions of stochastically fluctuating quantities for a non-equilibrium process, such as entropy, work and heat, are linked to equilibrium properties and corresponding quantities for the time-reversed process [2, 3, 4]. A well-known integral fluctuation relation is the Jarzynski work relation where the exponentiated work is averaged over its distribution and related to an equilibrium free-energy difference [5, 6]. Jarzynski's relation strengthens (1) by including all moments of the non-equilibrium work resulting in an *equality* from which (1) follows for the first moment. Experiments with biomolecules have used fluctuation theorems to derive, from the measurable work in non-equilibrium pulling experiments, the desired equilibrium free-energy surface of the molecules [7, 8], the latter being impossible to measure directly. In physics, fluctuation theorems have been measured for example, for a defect center in diamond [9], for a torsion pendulum [10], and in an electronic system [11]. Extensions to the quantum regime are reviewed in [12] and the first quantum experiments using nuclear magnetic resonance are reported in [13].

Jarzynski's relation in its general form states

$$\left\langle \exp \left( -\frac{w}{k_B T} \right) \right\rangle = \exp \left( -\frac{\Delta F^*}{k_B T} \right), \quad (2)$$

where  $T$  is the temperature of the initial thermal ensemble and  $\Delta F^*$  denotes the change in a free-energy measure  $F^*$  that differs from the usual Helmholtz free energy by the inclusion of a contribution from the coupling of the system to the thermal environment (heat reservoir) [6]. The final value of  $F^*$  used to calculate  $\Delta F^*$  is that of the final state of the system after it equilibrates at the same temperature ( $T$ ) as the initial state. The usual assumption of quasi-closed systems [1] is based on a system-environment coupling that gives a very small contribution to the system Hamiltonian leading to thermodynamic quantities that are extensive [14]. Quantities such as the Helmholtz free energy  $F$  are then calculated using the free-system Hamiltonian, without taking account of the coupling to the environment [1]. This is often an excellent approximation, but even then the Jarzynski equality, as a strict equality, does not hold for the Helmholtz free-energy difference  $\Delta F$ , but rather for  $\Delta F^*$ . For significant system-environment coupling there will in general be no close relation between  $\Delta F^*$  and the free-energy difference  $\Delta F$  calculated from the Hamiltonian of the uncoupled system.

The free energy  $F^*$  is defined as follows [6]. Let the total Hamiltonian of the system and environment be

$$H(x, X) = H_s(x) + H_e(X) + H_{\text{int}}(x, X), \quad (3)$$

where  $H_s(x)$  is the Hamiltonian of the free system, whose canonical variables are  $x = \{q_i, p_i\}$ ,  $H_e(X)$  is the Hamiltonian of the free environment, whose canonical variables are  $X = \{Q_i, P_i\}$ , and  $H_{\text{int}}(x, X)$  is the system-environment interaction term. If the coupled system and environment are in thermal equilibrium at temperature  $T$  then

it is easy to show that the system is distributed according to a phase space probability density  $\rho(x)$  that can be written as

$$\rho(x) = \frac{1}{Z^*} \exp\left(-\frac{H^*(x)}{k_B T}\right), \quad (4)$$

where  $H^*(x)$  is an effective Hamiltonian for the system [15] known as the Hamiltonian of mean force [16, 17]. The definition of  $H^*(x)$  is

$$H^*(x) = H_s(x) - \frac{1}{k_B T} \ln \left[ \frac{\int dX \exp\{-[H_e(X) + H_{\text{int}}(x, X)]/(k_B T)\}}{\int dX \exp[-H_e(X)/(k_B T)]} \right] \quad (5)$$

and  $Z^*$  is the partition function associated with the distribution (4):

$$Z^* = \int dx \exp\left(-\frac{H^*(x)}{k_B T}\right). \quad (6)$$

The distribution (4) of the system is obtained by “tracing out” the environment; the coupling of the system to the environment means this distribution is not a Boltzmann distribution based on the free-system Hamiltonian  $H_s(x)$ . The choice of  $H^*(x)$  as the effective Hamiltonian in the distribution (4) is not unique: one could add any  $x$ -independent term to (5) (for example one could cancel the denominator in the logarithm) and re-define the partition function (6) accordingly. But  $H^*(x)$  reduces to the free-system Hamiltonian  $H_s(x)$  when the coupling to the environment vanishes, which is a partial justification for its use as an effective Hamiltonian of the system.‡ The free energy  $F^*$  that appears in the Jarzynski equality (2) is the free energy associated with the partition function (6):

$$F^* = -k_B T \ln Z^*. \quad (7)$$

The purpose of this paper is to demonstrate the Jarzynski equality in a model system, an exercise that is nontrivial except in the least interesting case of an isolated system. As Jarzynski remarks in [6], “Exactly solvable models are hard to come by!”, and we aim to provide one. Our choice of model is the perfect gas, one of the classic thermodynamic systems. A single particle in the classical perfect gas is treated in the textbooks as a quasi-closed system, so that the thermodynamic quantities for the particle, and hence for the entire gas, are calculated using the free-particle Hamiltonian [1]. In reality however, the single particle interacts with the other particles through collisions; all the other particles act as a thermal environment to which the single particle is coupled. Our interest in this model arises from the fact that the system we will consider (a single particle) is coupled to the environment (the other particles) whereas the Helmholtz free energy of the system ignores this coupling. The question then arises of the difference between the Helmholtz free energy  $F$  and the free-energy  $F^*$  that appears in Jarzynski’s equality (2) for systems coupled to the environment.

‡ There are other reasons to view  $H^*(x)$  as the effective Hamiltonian of the system [6]. A perhaps surprising validation of the Hamiltonian of mean force occurs in the Casimir effect, where it determines the zero-point and thermal energy density of electromagnetic fields inside materials [18]. Casimir forces calculated from this energy density agree with those deduced from the electromagnetic stress tensor [18].

Anticipating our results, we will find that although  $F^* \neq F$  because of the system-environment coupling,  $\Delta F^*$  will be equal to  $\Delta F$  for the particular non-equilibrium process that we will analyse. The model reveals an interesting general implication from its structure: the Jarzynski equality holds in the form

$$\left\langle \exp \left( -\frac{w}{k_B T} \right) \right\rangle = \exp \left( -\frac{\Delta F}{k_B T} \right), \quad (8)$$

where  $F$  is the Helmholtz free energy calculated from the free-system Hamiltonian, in the general case of a system that is driven in one degree of freedom while interacting with the environment through other degrees of freedom. This is valid *even when the coupling to the environment is arbitrarily large*.

## 2. Model

In the case of a perfect gas the total Hamiltonian (3) of the system (a single particle) plus the environment (all the other particles) is given by

$$H_s(\mathbf{p}) = \frac{p^2}{2m}, \quad H_e(\{\mathbf{P}_i\}) = \sum_i \frac{P_i^2}{2m}, \quad H_{\text{int}}(\mathbf{q}, \{\mathbf{Q}_i\}) = V(\mathbf{q}, \{\mathbf{Q}_i\}). \quad (9)$$

The free-system and free-environment Hamiltonians ( $H_s$  and  $H_e$ , respectively) depend on the canonical momenta of the particles while the interaction potential depends on the particle positions. We consider a gas of finite volume so the interaction potential  $V(\mathbf{q}, \{\mathbf{Q}_i\})$  describes not just the collisions between the particles but also the collisions with the bounding walls (the walls are thus also part of the environment to which the system is coupled). Collisions between the particles in our classical model will have to be calculated exactly so we take them to be impenetrable spheres of radius  $R$ . The interaction potential  $V(\mathbf{q}, \{\mathbf{Q}_i\})$  thus consists of a set of potential barriers created by the particles and the walls; the former barriers are functions of the distances between pairs of particles, while the latter barriers depend on the distances between the particles and the walls.

The gas is initially in thermal equilibrium at temperature  $T$ . In the usual approximation where each particle is considered as a quasi-closed system [1], the coupling between the particles is ignored and the particle forming our system of interest occupies an infinitesimal volume of its phase space centred on  $(\mathbf{q}, \mathbf{p})$  with a probability

$$\rho(\mathbf{q}, \mathbf{p}) d^3q d^3p = Z^{-1} \exp \left( -\frac{p^2}{2mk_B T} \right) d^3q d^3p, \quad Z = V (2\pi mk_B T)^{3/2}, \quad (10)$$

where  $Z$  is the partition function and  $V$  is the volume of the gas. The probability distribution (10) is the Boltzmann-Gibbs distribution resulting from the free-system Hamiltonian  $H_s(\mathbf{p})$  in (9). Note that we do not use the semi-classical phase-space volume element  $d^3q d^3p / (2\pi\hbar)^3$ , as our model is purely classical. The Helmholtz free energy of the particle is thus

$$F = -k_B T \ln Z = -k_B T \ln \left[ V (2\pi mk_B T)^{3/2} \right]. \quad (11)$$

Because of the coupling to the environment ( $H_{\text{int}}$  in (9)), the free energy (11) is not the quantity  $F^*$  that appears in the Jarzynski equality (2). The Hamiltonian of mean force (5) has a contribution from the interaction term  $H_{\text{int}}$  in (9), and this gives a contribution from the interaction to  $Z^*$  and  $F^*$  (equations (6) and (7)). We do not attempt to calculate  $F^*$  here, but return later to its relevance for the Jarzynski equality.

To explore the Jarzynski equality in this model, we consider a thought experiment. We drive the particle of interest out of equilibrium with the rest of the gas by continuously changing its mass over a time interval  $0 \leq t \leq \tau$  according to

$$m_t = m + \frac{t}{\tau}(m_\tau - m). \quad (12)$$

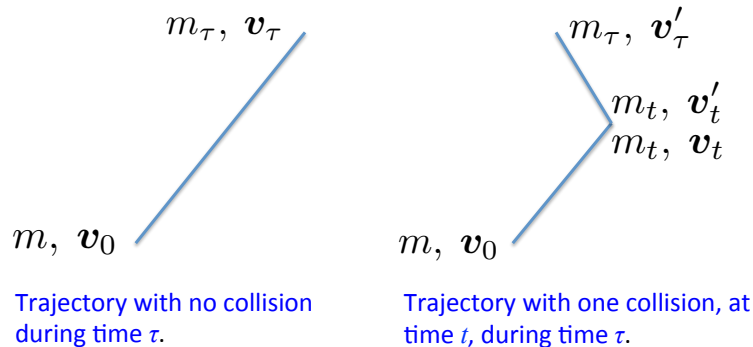
This non-equilibrium process is covered by the Jarzynski equality since it corresponds to a continuous change of a parameter ( $m$ ) in the system Hamiltonian  $H_s$  that does not appear in the interaction  $H_{\text{int}}$  with the environment [6]. The work done on the particle in this process would be trivial if the particle were a free system undergoing Hamiltonian evolution determined by  $H_s$ —in that case the work done would be just the change in energy of the particle. Collisions with the other particles during the process described by (12) make the calculation of the work more interesting.

### 3. Work done along trajectories with and without a collision

If the particle of interest moves freely between collisions (without its mass changing), its momentum  $\mathbf{p}$  is constant and its energy is given by  $H_s = p^2/(2m)$ . The momentum between collisions is constant even when the mass of the particle is changing continuously in time according to (12) because we have defined this process as a change of the mass in the system *Hamiltonian* (as opposed to, say, the Lagrangian). With the changing mass parameter the system Hamiltonian is  $H_s = p^2/(2m_t)$ , and so  $\dot{\mathbf{p}} = -\partial H_s/\partial \mathbf{q} = 0$ . The Jarzynski equality only applies to parameter changes in the Hamiltonian; thus a process in which the particle Lagrangian is  $L_s = m_t v^2/2$ , with  $m_t$  given by (12), would change the momentum of the driven particle and give a work relation that differs from the Jarzynski equality.

A collision changes the momentum of the driven particle discontinuously to a new value that is again conserved until the next collision. During the time interval  $\tau$  in which the mass of the particle is given by (12), the Maxwell distribution of velocities implies a non-zero probability of any number of collisions occurring. Larger numbers of collisions are more unlikely than smaller numbers, however, and to make the problem tractable we will choose  $\tau$  small enough so that it is highly probable there will be no collisions during this time interval, with a very small probability of one collision. The probabilities of two or more collisions during the time  $\tau$  will be neglected as negligible. This approximation will not prevent us from checking the Jarzynski relation (2) as a strict equality because the inclusion of higher numbers of collisions leads to terms with higher powers of  $\tau$  and the equality (2) must separately hold for all orders of  $\tau$ . We

thus consider only two types of trajectories for the driven particle in the time interval  $\tau$ : trajectories with no collisions and trajectories with one collision (see Fig. 1).



**Figure 1.** Trajectories of the driven particle in the time interval  $0 \leq t \leq \tau$  with no collision (left) and with one collision at time  $t$  (right). As the momentum of the particle is conserved between collisions, while its mass  $m_t$  changes, its velocity  $\mathbf{v}_t$  changes with time. A collision at time  $t$  (right) changes the velocity discontinuously from  $\mathbf{v}_t$  to a new value  $\mathbf{v}'_t$ .

Let the velocity of the particle of interest at  $t = 0$  be  $\mathbf{v}_0$ . The momentum of the particle is conserved for  $t > 0$  as long as there is no collision, while its mass  $m_t$  changes; hence its velocity  $\mathbf{v}_t$  at time  $t > 0$  satisfies

$$m\mathbf{v}_0 = m_t\mathbf{v}_t \quad (13)$$

if there is no collision. For the first type of trajectory, which has no collision for the entire interval  $\tau$ , the work done on the particle throughout the change of its mass is the resulting change of energy (with constant momentum  $m\mathbf{v}_0$ ):

$$w_{\text{free}} = \frac{1}{2}m^2v_0^2 \left( \frac{1}{m_\tau} - \frac{1}{m} \right). \quad (14)$$

The second type of trajectory has a collision at time  $t$  during the interval  $\tau$ , in which the velocity of the particle is changed instantly from  $\mathbf{v}_t$  to a new value  $\mathbf{v}'_t$ . The work done on the particle in this trajectory [19, 20] is the sum of the change in its energy in the interval 0 to  $t$  before the collision (with constant momentum  $m\mathbf{v}_0$ ) and the change in its energy in the interval  $t$  to  $\tau$  after the collision (with constant momentum  $m_t\mathbf{v}'_t$ ):

$$w_{\text{col}} = \frac{1}{2}m^2v_0^2 \left( \frac{1}{m_t} - \frac{1}{m} \right) + \frac{1}{2}m_t^2v_t'^2 \left( \frac{1}{m_\tau} - \frac{1}{m_t} \right). \quad (15)$$

To calculate averages involving the work done on the driven particle, we compute separately the contributions of the two types of trajectories. For trajectories of the first type we must weight the average with the probability of no collision occurring during the interval  $\tau$ , and also average over the initial velocity  $\mathbf{v}_0$  of the particle. For trajectories of the second type we must weight averages with the probability of a collision at time  $t$  giving the particle a velocity  $\mathbf{v}'_t$ , average over the possible post-collision velocities  $\mathbf{v}'_t$ , sum (integrate) over all collision times  $t$  in the interval  $0 \leq t \leq \tau$ , and average over the initial velocity  $\mathbf{v}_0$ .

#### 4. Collision statistics

The statistics of a collision is the only challenging aspect of the model. While the collision *rate* in a perfect gas is not difficult to obtain [1], we require something much more detailed for the calculation of the average work and average exponentiated work for the mass variation process. Specifically, we need the statistics of the final velocity of a colliding particle that has been driven out of equilibrium with the rest of the gas (the work (15) contains the post-collision speed  $v'_t$  of the driven particle).

The probability of the driven particle colliding with another particle in unit time is proportional to the flux of the other particles in the reference frame of the driven particle. As the other particles are in thermal equilibrium at temperature  $T$ , their velocities are Maxwell distributed, with each particle having a probability  $\rho(v)d^3v$  of moving at a velocity between  $\mathbf{v}$  and  $\mathbf{v} + d\mathbf{v}$ , where [1]

$$\rho(v)d^3v = \left(\frac{m}{2\pi k_B T}\right)^{3/2} \exp\left(-\frac{mv^2}{2k_B T}\right) d^3v. \quad (16)$$

The thermal distributions of the momenta of the particles are unaffected by the interaction terms in the Hamiltonian (9) because the latter do not contain the momenta. The joint velocity distribution of the particles is therefore just the product of Maxwell distributions (16). (The joint position distributions of the particles will, however, depend on the interaction; in the simple case of potential barriers the joint distribution will rule out spatial overlaps of the particles.) In the rest frame of the driven particle (Fig. 2) a particle with velocity  $\mathbf{v}$  in the laboratory frame moves with the relative velocity

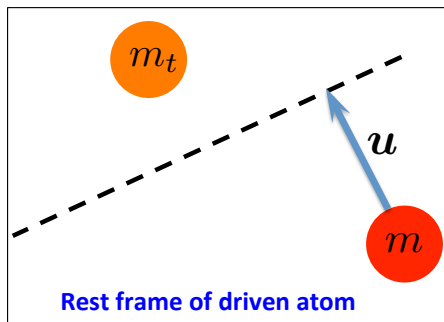
$$\mathbf{u} = \mathbf{v} - \mathbf{v}_t, \quad (17)$$

where  $\mathbf{v}_t$  is the velocity of the driven particle. The other particle moves a distance  $u = |\mathbf{u}|$  per unit time in the frame of the driven particle. If there are  $N$  other particles ( $N+1$  particles in total) then the flux of other particles with relative velocity between  $\mathbf{u}$  and  $\mathbf{u} + d\mathbf{u}$  incident on the driven particle is  $(N/V)u\rho(v)d^3v$ , where we must substitute for  $v$  in terms of  $u$  using (17). Defining  $\theta$  as the angle between the relative velocity  $\mathbf{u}$  and the velocity  $\mathbf{v}_t$  of the driven particle, this flux is, from (16) and (17),

$$\frac{N}{V} \left(\frac{m}{2\pi k_B T}\right)^{3/2} \exp\left[-\frac{m}{2k_B T} (u^2 + 2uv_t \cos\theta + v_t^2)\right] u^3 2\pi \sin\theta d\theta du. \quad (18)$$

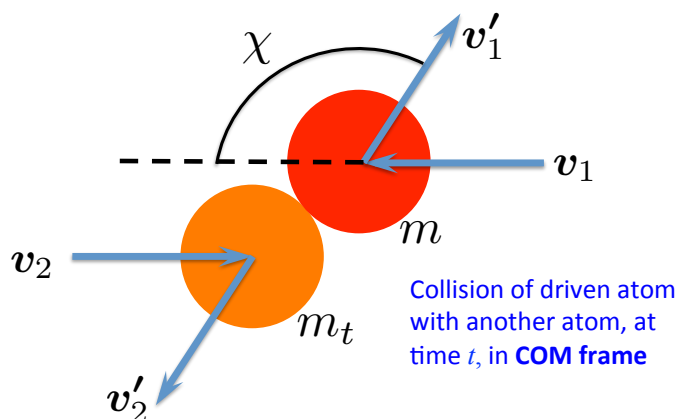
The probability per unit time of the driven particle colliding with a particle having relative velocity between  $\mathbf{u}$  and  $\mathbf{u} + d\mathbf{u}$  is proportional to the flux (18). Denoting this probability per unit time by  $P(u, \theta, v_t)d\theta du$ , the probability for such a collision to occur between times  $t$  and  $t + dt$  is  $P(u, \theta, v_t)d\theta du dt$ , and the probability for such a collision *not* to occur between times  $t$  and  $t + dt$  is  $1 - P(u, \theta, v_t)d\theta du dt$ . The probability of one collision occurring between times  $t$  and  $t + dt$  during the time interval  $\tau$  is thus

$$\left(1 - \int_0^t dt' P(u', \theta', v_{t'})d\theta' du'\right) P(u, \theta, v_t)d\theta du dt \times \left(1 - \int_t^\tau dt'' P(u'', \theta'', v_{t''})d\theta'' du''\right), \quad (19)$$



**Figure 2.** In the rest frame of the driven particle (with mass  $m_t$ ), another particle (with mass  $m$ ) moves at the relative velocity  $\mathbf{u}$ , covering a distance  $u = |\mathbf{u}|$  per unit time.

and we recall that the collision is with a particle having relative velocity between  $\mathbf{u}$  and  $\mathbf{u} + d\mathbf{u}$ . Trajectories with one collision during the interval  $\tau$  have a probability that contains (19), averaged over the time  $t$  of the collision. As described in the last section, we take the time interval  $\tau$  to be so short that the probability of a collision during this time is very small. It follows that we may use  $P(u, \theta, v_t) d\theta du dt$ , instead of (19), in treating trajectories with one collision, which implies our results for these trajectories will be first order in  $\tau$ . Analysis of trajectories with more than one collision will necessarily lead to higher-order terms in  $\tau$  and the Jarzynski equality must hold separately for each order of  $\tau$ .



**Figure 3.** Collision of the driven particle (mass  $m_t$ ) with another particle (mass  $m$ ) at time  $t$  in the centre of mass (COM) frame of the two particles.

The probability per unit time of the driven particle colliding is the flux of incident particles times the scattering cross section. We will write the differential scattering cross section in terms of the scattering angle  $\chi$  in the centre of mass (COM) frame (Fig. 3). In the COM frame [21], the velocities of the two particles before the collision are in opposite directions, as are the velocities after the collision, and the collision only



changes the direction of each velocity, not their magnitudes. We denote, in the COM frame, the pre-collision velocities of the driven particle and the particle with which it collides by  $\mathbf{v}_2$  and  $\mathbf{v}_1$ , respectively (Fig. 3). The corresponding post-collision velocities are denoted  $\mathbf{v}'_2$  and  $\mathbf{v}'_1$ . We then have [21]

$$\mathbf{v}_1 = \frac{m_t}{m + m_t} \mathbf{u}, \quad \mathbf{v}_2 = -\frac{m}{m + m_t} \mathbf{u}, \quad (20)$$

$$|\mathbf{v}'_1| = |\mathbf{v}_1|, \quad |\mathbf{v}'_2| = |\mathbf{v}_2|, \quad \mathbf{v}_1 \cdot \mathbf{v}'_1 = v_1^2 \cos \chi, \quad (21)$$

where  $\mathbf{u}$  is the pre-collision relative velocity (17) in the laboratory frame (which is also the pre-collision relative velocity in the COM frame:  $\mathbf{u} = \mathbf{v}_1 - \mathbf{v}_2$ ) and  $\chi$  is the scattering angle in the COM frame. We have already denoted the post-collision velocity of the driven particle in the laboratory frame by  $\mathbf{v}'_t$ ; the post-collision velocity of the other particle in the laboratory frame is denoted  $\mathbf{v}'$ . In terms of the pre-collision laboratory frame velocities ( $\mathbf{v}$  and  $\mathbf{v}_t$ ) and the post-collision COM-frame velocity  $\mathbf{v}'_1$  of the non-driven particle,  $\mathbf{v}'$  and  $\mathbf{v}'_t$  can be written [21]

$$\mathbf{v}' = \mathbf{v}'_1 + \frac{m\mathbf{v} + m_t\mathbf{v}_t}{m + m_t}, \quad \mathbf{v}'_t = -\frac{m}{m_t}\mathbf{v}'_1 + \frac{m\mathbf{v} + m_t\mathbf{v}_t}{m + m_t}. \quad (22)$$

Recalling that the particles are spheres of radius  $R$ , the impact parameter is  $2R \cos(\chi/2)$  and the differential cross section  $d\sigma$  is [21]

$$d\sigma = \frac{1}{4}(2R)^2 \sin \chi \, d\chi \, d\phi, \quad (23)$$

where  $\phi$  is the azimuthal angle in the COM frame corresponding to rotating the post-collision velocities for fixed scattering angle  $\chi$  around the pre-collision velocities.

The probability of a particle with relative velocity between  $\mathbf{u}$  and  $\mathbf{u} + d\mathbf{u}$  colliding with the driven particle at a time between  $t$  and  $t + dt$  (where  $t$  is less than the very short time interval  $\tau$ ) and being scattered into the solid angle element  $d\chi \, d\phi$  centred on  $(\chi, \phi)$  is equal to the flux (18) times  $dt$  times the differential cross section (23):

$$\frac{2\pi R^2 N}{V} \left( \frac{m}{2\pi k_B T} \right)^{3/2} \exp \left[ -\frac{m}{2k_B T} (u^2 + 2uv_t \cos \theta + v_t^2) \right] u^3 \sin \theta \sin \chi \, du \, d\theta \, d\chi \, d\phi \, dt. \quad (24)$$

This is the (differential) probability for a single trajectory of the driven particle, with initial velocity  $\mathbf{v}_0$ , that includes one collision. The (very small) total probability of the driven particle with initial velocity  $\mathbf{v}_0$  following a trajectory with one collision during the (very short) time interval  $\tau$  is obtained by integrating (24) over all relative speeds  $u$ , all angles  $\theta$  between  $\mathbf{u}$  and  $\mathbf{v}_t$ , all scattering angles  $(\chi, \phi)$ , and all possible collision times  $t$  in the interval  $0 \leq t \leq \tau$ . For averages involving the work done on the driven particle in trajectories with a collision, we require the differential probability (24) of each trajectory with a collision since the work depends on the details of the trajectory (relative velocity, scattering angle and time of collision).

The work done (15) on the driven particle in a trajectory with a collision features its post-collision speed  $v'_t$ ; in order to calculate work averages it is necessary to express this speed in terms of the variables  $u$ ,  $\theta$ ,  $\chi$  and  $\phi$  appearing in the differential probability

(24) for a trajectory with a collision. From the second of (22) we obtain the following expression for  $v_t'^2$ :

$$v_t'^2 = \frac{2m^2}{(m+m_t)^2}u^2 + \frac{2m}{m+m_t}uv_t \cos \theta + v_t^2 - \frac{2m^2}{m_t(m+m_t)}\mathbf{v}'_1 \cdot \mathbf{u} - \frac{2m}{m_t}\mathbf{v}'_1 \cdot \mathbf{v}_t, \quad (25)$$

where we have used the definition of  $\theta$ , introduced before equation (18), as the angle between  $\mathbf{u}$  and  $\mathbf{v}_t$ . The relative velocity  $\mathbf{u}$  is in the direction of  $\mathbf{v}_1$  (since  $\mathbf{u} = \mathbf{v}_1 - \mathbf{v}_2$ ), hence (see Fig. 3) the angle between  $\mathbf{u}$  and  $\mathbf{v}'_1$  is the scattering angle  $\chi$  and we have

$$\mathbf{v}'_1 \cdot \mathbf{u} = v'_1 u \cos \chi = v_1 u \cos \chi = \frac{m_t}{m+m_t}u^2 \cos \chi, \quad (26)$$

where we have used (21) and (20). To simplify the dot product  $\mathbf{v}'_1 \cdot \mathbf{v}_t$  in (25), we refer to Fig. 4. Recall that  $\theta$  is the angle between  $\mathbf{u}$  and  $\mathbf{v}_t$  (left of Fig. 4) and introduce a unit vector  $\mathbf{n}_1$  orthogonal to  $\mathbf{u}$ , lying in the plane of  $\mathbf{u}$  and  $\mathbf{v}_t$ . We can then expand  $\mathbf{v}_t$  as

$$\mathbf{v}_t = \frac{\mathbf{u}}{u}v_t \cos \theta + \mathbf{n}_1 v_t \sin \theta. \quad (27)$$

As noted above equation (26),  $\chi$  is the angle between  $\mathbf{u}$  and  $\mathbf{v}'_1$  (right of Fig. 4). Introducing a unit vector  $\mathbf{n}_2$  orthogonal to  $\mathbf{u}$ , lying in the plane of  $\mathbf{u}$  and  $\mathbf{v}'_1$ , we can expand  $\mathbf{v}'_1$  as

$$\begin{aligned} \mathbf{v}'_1 &= \frac{\mathbf{u}}{u}v'_1 \cos \chi + \mathbf{n}_2 v'_1 \sin \chi = \frac{\mathbf{u}}{u}v_1 \cos \chi + \mathbf{n}_2 v_1 \sin \chi \\ &= \frac{m_t}{m+m_t}u \left( \frac{\mathbf{u}}{u} \cos \chi + \mathbf{n}_2 \sin \chi \right), \end{aligned} \quad (28)$$

where we used (21) and (20). The unit vectors  $\mathbf{n}_1$  and  $\mathbf{n}_2$  lie in a plane orthogonal to the relative velocity  $\mathbf{u}(= \mathbf{v}_1 - \mathbf{v}_2)$ , and hence they lie in a plane orthogonal to the pre-collision velocities  $\mathbf{v}_1$  and  $\mathbf{v}_2$  in the COM frame (see Fig. 3). We can therefore define the angle  $\phi$  in the differential cross section (23), which measures a rotation of the post-collision velocities about the line of the pre-collision velocities in the COM frame, as the angle between the unit vectors  $\mathbf{n}_1$  and  $\mathbf{n}_2$ . The expansions (27) and (28) then give the dot product

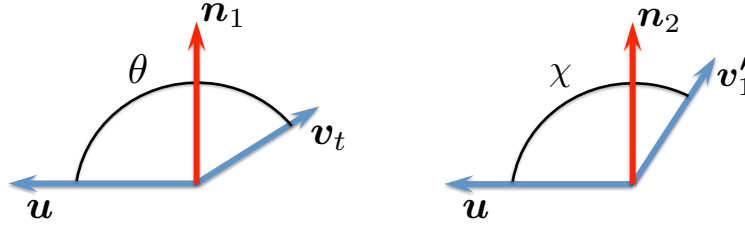
$$\mathbf{v}'_1 \cdot \mathbf{v}_t = \frac{m_t}{m+m_t}uv_t (\cos \chi \cos \theta + \sin \chi \sin \theta \cos \phi). \quad (29)$$

Inserting (26) and (29) in (25), we find the following expression for the square of the post-collision speed of the driven particle in the laboratory frame:

$$v_t'^2 = v_t^2 + \frac{2m^2}{(m+m_t)^2}u^2 (1 - \cos \chi) + \frac{2m}{m+m_t}uv_t [\cos \theta (1 - \cos \chi) - \sin \chi \sin \theta \cos \phi]. \quad (30)$$

## 5. Work averages

We now have the ingredients to compute work averages for the driven particle. In our approximation of a very short driving time  $\tau$ , these averages have contributions from trajectories with no collision and from trajectories with one collision. In order to analytically evaluate the integrals involved we must make an expansion in terms of



**Figure 4.** Left: The plane of the relative velocity  $\mathbf{u}$  of the colliding particles and the pre-collision velocity  $\mathbf{v}_t$  of the driven particle. Right: The plane of the relative velocity  $\mathbf{u}$  and the post-collision velocity  $\mathbf{v}'_1$  in the COM frame of the non-driven particle.

$m_\tau - m$ , the total change in mass of the driven particle during the driving time  $\tau$ . While it is in principle possible to expand the results to arbitrary order in  $m_\tau - m$ , we will only do so up to order  $(m_\tau - m)^2$ . As with expansions in the driving time  $\tau$ , any strict equalities involving work averages must hold separately for all orders of  $m_\tau - m$ . The quantities  $m_t$  and  $v_t$  that appear in the integrands must be written in terms of  $m$  and  $m_\tau$  in order to expand in powers of  $m_\tau - m$ ; we express  $m_t$  in terms of  $m$  and  $m_\tau$  using (12) and  $v_t$  in terms of  $v_0$ ,  $m$  and  $m_\tau$  using (13) and (12).

### 5.1. Work averages for trajectories with no collision

The work done on the driven particle, with initial velocity  $\mathbf{v}_0$ , in a trajectory with no collision is given by (14). Averages over such trajectories of functions of the work (14) must be weighted with the probability of no collision occurring during the driving time  $\tau$ . As noted after (24), the probability of a collision occurring during the driving time  $\tau$ , for initial velocity  $\mathbf{v}_0$  of the driven particle, is given by integrating (24) over all trajectories with a collision; we denote this probability by  $P_{\text{col}}(\tau, v_0)$ . Expanding (24) to order  $(m_\tau - m)^2$  and integrating over  $\phi$ ,  $\chi$ ,  $\theta$ ,  $u$  (from 0 to  $\infty$ ), and  $t$  (from 0 to  $\tau$ ) gives

$$\begin{aligned}
P_{\text{col}}(\tau, v_0) &= \frac{4R^2\tau}{mv_0} \left[ \sqrt{2\pi mk_B T} v_0 \exp\left(-\frac{mv_0^2}{2k_B T}\right) + \pi (k_B T + mv_0^2) \operatorname{erf}\left(v_0 \sqrt{\frac{m}{2k_B T}}\right) \right] \\
&\quad - \frac{2R^2\tau}{m^2 v_0} \left[ \sqrt{2\pi mk_B T} v_0 \exp\left(-\frac{mv_0^2}{2k_B T}\right) - \pi (k_B T - mv_0^2) \operatorname{erf}\left(v_0 \sqrt{\frac{m}{2k_B T}}\right) \right] (m_\tau - m) \\
&\quad + \frac{4\pi R^2 v_0 \tau}{3m^2} \operatorname{erf}\left(v_0 \sqrt{\frac{m}{2k_B T}}\right) (m_\tau - m)^2 + O(m_\tau - m)^3, \tag{31}
\end{aligned}$$

where  $\operatorname{erf}(x) = (2/\sqrt{\pi}) \int_0^x dt e^{-t^2}$  is the error function. This is the probability of the driven particle, with initial velocity  $\mathbf{v}_0$ , experiencing a collision during the driving time  $\tau$ . The probability of no collision, for initial velocity  $\mathbf{v}_0$ , during the time  $\tau$  is one minus (31).

Consider the exponentiated work average  $\langle \exp[-w/(k_B T)] \rangle$  that appears in the Jarzynski equality (2). The contribution to  $\langle \exp[-w/(k_B T)] \rangle$  from trajectories without a collision is obtained by taking  $\exp[-w_{\text{free}}/(k_B T)]$ , with  $w_{\text{free}}$  given by (14), multiplying

by  $(1 - P_{\text{col}}(\tau, v_0))$  and averaging over the initial velocity  $\mathbf{v}_0$  of the driven particle. We denote this contribution to  $\langle \exp[-w/(k_B T)] \rangle$  by  $\langle \exp[-w_{\text{free}}/(k_B T)] \rangle_{\text{free}}$ . At  $t = 0$  the driven particle is still in thermal equilibrium with the rest of the gas so  $\mathbf{v}_0$  is Maxwell distributed with probability distribution (16). Performing the average over  $\mathbf{v}_0$ , in which we expand the integrand to second order in  $m_\tau - m$ , we obtain

$$\begin{aligned} \left\langle \exp\left(-\frac{w_{\text{free}}}{k_B T}\right) \right\rangle_{\text{free}} &= 1 - 16R^2\tau \frac{N}{V} \sqrt{\frac{\pi k_B T}{m}} + \left( \frac{3}{2} - 24R^2\tau \frac{N}{V} \sqrt{\frac{\pi k_B T}{m}} \right) \frac{m_\tau - m}{m} \\ &+ \left( \frac{9}{24} - \frac{35}{6} R^2\tau \frac{N}{V} \sqrt{\frac{\pi k_B T}{m}} \right) \frac{(m_\tau - m)^2}{m^2} + O(m_\tau - m)^3. \end{aligned} \quad (32)$$

For consideration of the inequality (1), we require the average work  $\langle w \rangle$ . The contribution to  $\langle w \rangle$  from trajectories without a collision, which we denote by  $\langle w_{\text{free}} \rangle_{\text{free}}$ , is computed as described in the last paragraph, with  $w_{\text{free}}$  replacing  $\exp[-w_{\text{free}}/(k_B T)]$ . The result is

$$\begin{aligned} \langle w_{\text{free}} \rangle_{\text{free}} &= - \left[ \frac{3k_B T}{2m} - 28\sqrt{\pi} R^2\tau \frac{N}{V} \left( \frac{k_B T}{m} \right)^{3/2} \right] (m_\tau - m) \\ &+ \left[ \frac{3k_B T}{2m} - 37\sqrt{\pi} R^2\tau \frac{N}{V} \left( \frac{k_B T}{m} \right)^{3/2} \right] \frac{(m_\tau - m)^2}{m} + O(m_\tau - m)^3. \end{aligned} \quad (33)$$

## 5.2. Work averages for trajectories with a collision

The work done on the driven particle, with initial velocity  $\mathbf{v}_0$ , in a trajectory with a collision is given by (15). Averages over such trajectories of functions of the work (15) must be weighted with the probability of each trajectory, and these probabilities are given by (24) with the values of  $u$ ,  $\theta$ ,  $\chi$ ,  $\phi$  and  $t$  (collision time) that characterise the trajectory. The work (15) also depends on the trajectory through the same set of parameters, as can be seen by substituting (30) into (15).

The contribution of trajectories with a collision to the average  $\langle \exp[-w/(k_B T)] \rangle$  is denoted by  $\langle \exp[-w_{\text{col}}/(k_B T)] \rangle_{\text{col}}$ . We compute this by weighting  $\exp[-w_{\text{col}}/(k_B T)]$  with the probability (24) of the trajectory corresponding to  $w_{\text{col}}$  and integrating over all trajectories with a collision. Before performing the integrations we expand the integrand to second order in  $m_\tau - m$ . The integrations are over  $u$ ,  $\theta$ ,  $\chi$ ,  $\phi$  and  $t$  and the result is

$$\begin{aligned} \left\langle \exp\left(-\frac{w_{\text{col}}}{k_B T}\right) \right\rangle_{\text{col}} &= R^2\tau \frac{N}{V} \sqrt{\frac{\pi k_B T}{m}} \left[ 16 + \frac{24}{m} (m_\tau - m) + \frac{35}{6m^2} (m_\tau - m)^2 \right] \\ &+ O(m_\tau - m)^3. \end{aligned} \quad (34)$$

The contribution of trajectories with a collision to the average work  $\langle w \rangle$  is denoted by  $\langle w_{\text{col}} \rangle_{\text{col}}$ . This is calculated in the same manner as described in the last paragraph and the result is

$$\langle w_{\text{col}} \rangle_{\text{col}} = \sqrt{\pi} R^2\tau \frac{N}{V} \left( \frac{k_B T}{m} \right)^{3/2} \left[ -28(m_\tau - m) + \frac{103}{3m} (m_\tau - m)^2 \right] + O(m_\tau - m)^3. \quad (35)$$

## 6. Jarzynski equality and the second law

By adding the two contributions (32) and (34), we find the average  $\langle \exp[-w/(k_B T)] \rangle$  of the exponentiated work done on the driven particle, to order  $(m_\tau - m)^2$ :

$$\left\langle \exp\left(-\frac{w}{k_B T}\right) \right\rangle = 1 + \frac{3}{2m}(m_\tau - m) + \frac{3}{8m^2}(m_\tau - m)^2 + O(m_\tau - m)^3. \quad (36)$$

Note that the effect of the collisions, which shows up in terms containing (among other quantities) the particle radius  $R$ , has entirely cancelled out in (36). The *equilibrium* Helmholtz free-energy difference  $\Delta F$  of the driven particle between the equilibrium states corresponding to the final system parameter (mass  $m_\tau$ ) and the initial system parameter (mass  $m$ ) is shown by (11) to be  $\Delta F = -(3/2)k_B T \ln(m_\tau/m)$ , which gives

$$\exp\left(-\frac{\Delta F}{k_B T}\right) = \left(\frac{m_\tau}{m}\right)^{3/2} = 1 + \frac{3}{2m}(m_\tau - m) + \frac{3}{8m^2}(m_\tau - m)^2 + O(m_\tau - m)^3. \quad (37)$$

The results (36) and (37) indicate that the Jarzynski equality holds for this process in the form (8), despite the fact that the system is coupled to the environment. In fact there is no reason why (8) should hold only up to the accuracy of our calculation and cease to hold for higher orders in  $m_\tau - m$  and  $\tau$ . In the next section we show that the process we have considered is such that  $\Delta F^* = \Delta F$ , so we have found a case where the Helmholtz free-energy  $F$  of the system gives the exact work relation (8) even though  $F$  neglects the coupling to the environment.

The formulation (1) of the second law also follows from the Jarzynski equality in the form (8) [5]. For systems coupled to the environment, however, the Jarzynski equality (in general) takes the form (2), leading to the inequality  $\langle w \rangle \geq \Delta F^*$ . In our model the Jarzynski equality takes the form (8), despite the coupling to the environment (see next section for a complete proof), and thus the inequality  $\langle w \rangle \geq \Delta F$  must strictly hold. Adding (33) and (35), we obtain the average work

$$\langle w \rangle = -\frac{3k_B T}{2m}(m_\tau - m) + \left[ \frac{3k_B T}{2m} - \frac{8}{3}\sqrt{\pi}R^2\tau\frac{N}{V}\left(\frac{k_B T}{m}\right)^{3/2} \right] \frac{(m_\tau - m)^2}{m} + O(m_\tau - m)^3. \quad (38)$$

In contrast to (36), the average work (38) does depend on the collisions (coupling to the environment). The equilibrium free-energy difference  $\Delta F = -(3/2)k_B T \ln(m_\tau/m)$  gives

$$\Delta F = k_B T \left[ -\frac{3}{2m}(m_\tau - m) + \frac{3}{4m^2}(m_\tau - m)^2 \right] + O(m_\tau - m)^3. \quad (39)$$

Comparing (38) and (39), we find the inequality  $\langle w \rangle \geq \Delta F$  requires

$$\frac{3}{16} \geq R^2\tau\frac{N}{V}\sqrt{\frac{\pi k_B T}{m}} + O(m_\tau - m)^3. \quad (40)$$

Our calculation is only valid for very small  $\tau$  such that we can neglect terms of order  $\tau^2$  and higher; the bound on  $\tau$  given by (40) can thus be met by our approximations. The results of the next section show that  $\langle w \rangle \geq \Delta F$  will also hold for our model if the exact average work  $\langle w \rangle$  is calculated.

## 7. General condition for cancellation of environment coupling in the Jarzynski equality

For a system coupled to the environment, the Jarzynski equality is (8), where the work is performed by a change in a parameter of the system Hamiltonian  $H_s(x)$ . Denote the work parameter (which was the particle mass  $m$  in our model) by  $\lambda$ . We can then include the work parameter in the functional dependence of the system Hamiltonian, so that it is denoted  $H_s(x, \lambda)$ . Now suppose  $H_s(x, \lambda)$  can be written

$$H_s(x, \lambda) = H_s^{(1)}(x_1, \lambda) + H_s^{(2)}(x_2), \quad (41)$$

where  $x_1$  and  $x_2$  are disjoint subsets of the set of canonical variables  $x = \{q_i, p_i\}$  of the system such that  $x_1 \cup x_2 = x$ . The decomposition (41) applies when there are canonical variables ( $x_2$ ) of the system that do not appear in any of the terms of  $H_s(x, \lambda)$  containing the work parameter  $\lambda$ . Suppose further that the interaction term  $H_{\text{int}}(x, X)$  in the full system-environment Hamiltonian (3) only contains the subset  $x_2$  of the system canonical variables, i.e.

$$H_{\text{int}}(x, X) = H_{\text{int}}(x_2, X). \quad (42)$$

Now consider the partition function (6), which we here denote by  $Z^*(\lambda)$  to record its dependence on  $\lambda$ . Using the definition (5) of the Hamiltonian of mean force,  $Z^*(\lambda)$  for the special case (41) and (42) takes the form

$$\begin{aligned} Z^*(\lambda) &= \int dx_1 \exp \left[ -H_s^{(1)}(x_1, \lambda)/(k_B T) \right] \\ &\times \frac{\int dx_2 \int dX \exp \left\{ - \left[ H_s^{(2)}(x_2) + H_e(X) + H_{\text{int}}(x_2, X) \right] / (k_B T) \right\}}{\int dX \exp \left[ -H_e(X)/(k_B T) \right]}. \end{aligned} \quad (43)$$

It is clear from (43) that the ratio of the partition function  $Z^*(\lambda)$  for two different values  $\lambda_A$  and  $\lambda_B$  of the work parameter  $\lambda$  is independent of the coupling to the environment, i.e. independent of  $H_{\text{int}}$ :

$$\frac{Z^*(\lambda_B)}{Z^*(\lambda_A)} = \frac{\int dx_1 \exp \left[ -H_s^{(1)}(x_1, \lambda_B)/(k_B T) \right]}{\int dx_1 \exp \left[ -H_s^{(1)}(x_1, \lambda_A)/(k_B T) \right]}. \quad (44)$$

The difference  $\Delta F^* = F^*(\lambda_B) - F^*(\lambda_A)$  between the two values of the free energy (7) corresponding to  $\lambda_A$  and  $\lambda_B$  is determined by the ratio (44). Hence in this case the free-energy difference  $\Delta F^*$  is equal to the Helmholtz free-energy difference  $\Delta F$  that ignores the coupling to the environment. This result establishes the following theorem regarding the Jarzynski equality: *If the terms in the system Hamiltonian that contain the work parameter are independent of the system canonical variables that appear in the interaction with the environment, then the Jarzynski equality holds exactly with the uncoupled system free-energy difference  $\Delta F$ .* Note that this theorem is independent of the strength of the coupling to the environment.

We see from (9) that the above theorem applies to our model, and this is consistent with our results for the Jarzynski equality in the previous section.

## 8. Conclusions

Jarzynski's recent results [5, 6] strengthen standard thermodynamics by providing an equality for a wide range of non-equilibrium processes. Our goal here has been to provide a model that is solvable and that includes coupling to an environment, in the belief that such models will help assess the significance of Jarzynski's results [5, 6]. We have explored how Jarzynski's work equality can be applied to one of the most basic thermodynamic systems—the perfect gas. The work done on a gas particle undergoing a change of mass was calculated taking into account the collisions of the particle with the gas. While the effect of collisions may be incorporated into the free energy of the particle, we have shown that its free energy difference is identical to an isolated particle and the standard Jarzynski equation holds. This is despite the coupling to the bath of other particles. Our model provides an example of an interesting general situation where (arbitrarily strong) coupling to the environment has no effect on Jarzynski's work relation. Given that the particular calculations presented here reveal a general theorem, the further study of simple models with coupling to the environment seems justified.

## References

- [1] Landau L D and Lifshitz E M 1980 *Statistical Physics Part 1* 3rd ed (Oxford: Butterworth-Heinemann)
- [2] Evans D J, Cohen E G D and Morriss G P 1993 *Phys. Rev. Lett.* **71** 2401
- [3] Crooks G E 1999 *Phys. Rev. E* **60** 2721
- [4] Kawai R, Parrondo J M R and Van den Broeck C 2007 *Phys. Rev. Lett.* **98** 1
- [5] Jarzynski C 1997 *Phys. Rev. Lett.* **78** 2690
- [6] Jarzynski C 2004 *J. Stat. Mech.: Theor. Exp.* P09005
- [7] Liphardt J, Dumont S, Smith S B, Tinoco Jr I and Bustamente C 2002 *Science* **296** 1832
- [8] Collin D, Ritort F, Jarzynski C, Smith S B, Tinoco Jr I and Bustamente C 2005 *Nature* **437** 231
- [9] Schuler S, Speck T, Tietz C, Wrachtrup J and Seifert U 2005 *Phys. Rev. Lett* **94** 180602
- [10] Douarche F, Joubaud S, Garnier N B, Petrosyan A and Ciliberto S 2006 *Phys. Rev. Lett* **97** 140603
- [11] Saira O P, Yoon Y, Tanttu T, Möttönen M, Averin D V and Pekola J P 2012 *Phys. Rev. Lett* **109** 180601
- [12] Esposito E, Harbola U and Mukamel S 2009 *Rev. Mod. Phys.* **81** 1665
- [13] Batalhão T, *et al.*, arXiv:1308.3241v1 (2013)
- [14] Abe S 1999 *Phys. Lett. A* **263** 424
- [15] Gelin M F and Thoss M 2009 *Phys. Rev. E* **79** 051121
- [16] Kirkwood J G 1935 *J. Chem. Phys.* **3** 300
- [17] Campisi M, Talkner P and Hänggi P 2009 *Phys. Rev. Lett.* **102** 210401
- [18] Philbin T G 2011 *New J. Phys.* **13** 063026; Philbin T G and Horsley S A R 2013 arXiv:1304.0977[quant-ph]
- [19] Crooks G E 1998 *J. Stat. Phys.* **90** 1481
- [20] Sagawa T and Ueda M 2010 *Phys. Rev. Lett.* **104** 090602
- [21] Landau L D and Lifshitz E M 1976 *Mechanics* 3rd ed (Oxford: Butterworth-Heinemann)