



UNIVERSITY OF LEEDS

This is a repository copy of *Validation of the Jarzynski relation for a system with strong thermal coupling: an isothermal ideal gas model*.

White Rose Research Online URL for this paper:
<http://eprints.whiterose.ac.uk/1960/>

Article:

Baule, A., Evans, R.M.L. and Olmsted, P.D. (2006) Validation of the Jarzynski relation for a system with strong thermal coupling: an isothermal ideal gas model. *Physical Review E : Statistical, Nonlinear and Soft Matter Physics*, 74 (6). Art. No. 061117- (10 pages). ISSN 1550-2376

<https://doi.org/10.1103/PhysRevE.74.061117>

Reuse

See Attached

Takedown

If you consider content in White Rose Research Online to be in breach of UK law, please notify us by emailing eprints@whiterose.ac.uk including the URL of the record and the reason for the withdrawal request.



eprints@whiterose.ac.uk
<https://eprints.whiterose.ac.uk/>



White Rose
university consortium
Universities of Leeds, Sheffield & York

White Rose Research Online

<http://eprints.whiterose.ac.uk/>

This is an author produced version of a paper published in **Physical Review E : Statistical, Nonlinear and Soft Matter Physics.**

White Rose Research Online URL for this paper:

<http://eprints.whiterose.ac.uk/1960/>

Published paper

Baule, A., Evans, R.M.L. and Olmsted, P.D. (2006) *Validation of the Jarzynski relation for a system with strong thermal coupling: an isothermal ideal gas model.* Physical Review E : Statistical, Nonlinear and Soft Matter Physics, 74 (6). Art. No. 061117.

Validation of the Jarzynski relation for a system with strong thermal coupling: an isothermal ideal gas model

A. Baule, R.M.L. Evans, P.D. Olmsted

School of Physics and Astronomy, University of Leeds, LS2 9JT, United Kingdom

We revisit the paradigm of an ideal gas under isothermal conditions. A moving piston performs work on an ideal gas in a container that is strongly coupled to a heat reservoir. The thermal coupling is modelled by stochastic scattering at the boundaries. In contrast to recent studies of an adiabatic ideal gas with a piston [R.C. Lua and A.Y. Grosberg, *J. Phys. Chem. B* 109, 6805 (2005); I. Bena et al., *Europhys. Lett.* 71, 879 (2005)], container and piston stay in contact with the heat bath during the work process, such that the heat reservoir as well as the system depend on the work parameter λ . We show that under this condition microscopic reversibility is broken for a moving piston. Our model is thus not included in the class of systems for which the non-equilibrium work theorem has been derived rigorously either by Hamiltonian [C. Jarzynski, *J. Stat. Mech.* P09005 (2004)] or stochastic methods [G.E. Crooks, *J. Stat. Phys.* 90, 1481 (1998)]. Nevertheless the validity of the non-equilibrium work theorem is confirmed both numerically for a wide range of parameter values and analytically in the limit of a very fast moving piston, i.e. in the far non-equilibrium regime.

PACS numbers: 05.70.Ln, 05.20.-y, 82.20.Wt

I. INTRODUCTION

Despite its successful application in numerous experimental and numerical studies (see e.g. the review articles [1, 2]), the validity of the non-equilibrium work theorem or ‘Jarzynski relation’ [3] still remains under discussion. Indeed this ongoing critique is mainly due to the surprising nature of the theorem: it states an exact equality that holds in situations arbitrarily far from equilibrium, under very general assumptions. More precisely, it states that the free energy difference between two equilibrium states can be extracted from work measurements along irreversible trajectories connecting these two states. Therefore one can, in principle, obtain equilibrium information from a non-equilibrium experiment, which is of particular interest in chemical and biophysical applications. For example, the non-equilibrium work theorem has been successfully applied to the stretching of a single protein [4]. In this experiment the work performed by a single RNA molecule tethered between a solid substrate and a controllable cantilever in an aqueous salt solution is measured for slow (reversible) and fast (irreversible) stretching. The free energy difference between its folded and unfolded conformations is obtained from the reversible process using ordinary equilibrium thermodynamics. On the other hand applying Jarzynski’s relation to the work values obtained from the irreversible process also reproduces this result within experimental errors, thus confirming the theorem.

However it has been questioned, in Ref. [5], whether this experiment indeed creates a non-equilibrium situation. It is argued that a slow or fast work process does not necessarily guarantee its reversibility or irreversibility. Rather the work rate has to be compared with the strength of the coupling (rate of heat transfer) between the system and its thermal environment. If the work rate is apparently large, but still smaller than the rate

of heat transfer, the system is essentially maintained in an equilibrium state. This is claimed [5] to be the case in the protein stretching experiment, since the surrounding liquid allows for rapid thermalization. Under such conditions the theorem is expected to hold trivially.

The above discussion highlights the importance of properly assessing the thermostating process between the system and the heat reservoir. The purpose of the present paper is to investigate the Jarzynski relation for the most simple thermodynamic system under isothermal and non-equilibrium conditions. In a gedankenexperiment an ideal gas is isothermally expanded in a heat-conducting container by pulling a piston at different velocities. Work is performed when the gas particles hit the piston during its movement. Similar ideal gas models have been investigated in [6, 7], but under adiabatic conditions, i.e. without considering heat transfer during the work process. The extension to an isothermal situation provides important further insight into Jarzynski’s relation.

The remainder of this paper is organized as follows. In the next section we present a brief outline of the non-equilibrium work theorem for a system with strong thermal coupling. In Sec. III the isothermal ideal gas model is introduced, which allows for an analytical formulation of the non-equilibrium work theorem. We present the results of a numerical study of this model and revisit the adiabatic expansion of the ideal gas in Sec. IV. Finally we conclude with a summary of the main points and a brief outlook.

II. THE NON-EQUILIBRIUM WORK THEOREM

The non-equilibrium work theorem can be formulated in the following way. The system of interest is prepared

in an initial state of equilibrium while in contact with a heat reservoir at temperature T . By changing an external parameter λ — the work parameter — according to a fixed protocol from A to B the system is subjected to a thermodynamic process at the end of which it reaches a final state not necessarily in equilibrium. This process can possibly drive the system arbitrarily far away from equilibrium while performing a certain amount of work, W . During the work process the system may or may not stay in contact with the thermal environment. Upon reaching the final parameter value $\lambda = B$ the system relaxes to equilibrium by exchanging heat with the reservoir but it is assumed that no further work is performed. If we repeat this process following the same protocol infinitely many times we obtain a distribution of work values $p(W)$ due to the stochastic nature of the initial equilibrium state, which is sampled from a canonical distribution. The Jarzynski relation then states a strong constraint on this work distribution [3]:

$$\langle e^{-\beta W} \rangle = \int dW p(W) e^{-\beta W} = e^{-\beta \Delta F}. \quad (1)$$

The average over the exponentiated work values equals the exponential of the free energy difference between the initial and final equilibrium states of the system, whether or not the final equilibrium state is actually realized in the disturbed system. β is the inverse temperature $1/\beta = k_B T$ and ΔF is the ratio of the equilibrium partition functions:

$$\Delta F = F_B - F_A = -\beta^{-1} \ln \frac{Z_B}{Z_A}. \quad (2)$$

It should be noted that F_B corresponds to the actual free energy of the final state only if the work parameter is finally held fixed at $\lambda = B$ until the system has thermalized with the reservoir. Without thermalization the system may well be in a state out of equilibrium such that no free energy can be defined.

The relation, Eq. (1), holds irrespective of the particular character of the work process and is valid beyond the linear response regime. For a fast switching of λ one may perform a non-equilibrium process and still obtain the equilibrium free energy difference by evaluating the exponential work average. Thus the Jarzynski relation is one of the few exact results applicable far from equilibrium. Since $\langle \exp(x) \rangle \geq \exp(\langle x \rangle)$, Eq. (1) implies the second law of thermodynamics formulated for work and free energy, $\langle W \rangle \geq \Delta F$. The equality $\langle W \rangle = \Delta F$ is only true for a reversible (quasistatic) process.

The Jarzynski relation has been derived both for deterministic Hamiltonian dynamics [3, 8] and for stochastic dynamics [9]. Without resorting to the full proof we shall make a few comments on the derivation of Eq. (1).

A. Hamiltonian derivation

In Jarzynski's original derivation [3] almost a decade ago, the specific assumption was made that the coupling between system and heat reservoir is sufficiently small as to neglect the interaction term in the Hamiltonian. Under this condition the work process is effectively assumed to be adiabatic. Only recently a derivation has been presented that does not rely on the weak coupling assumption [8], where the starting point is the Hamiltonian

$$\mathcal{H}(\Gamma; \lambda) = H(x; \lambda) + H_E(y) + h_{int}(x, y), \quad (3)$$

for the system and thermal environment. Here, $H(x; \lambda)$ denotes the Hamiltonian of the system, $H_E(y)$ is the Hamiltonian of the heat reservoir and $h_{int}(x, y)$ the interaction Hamiltonian. $\Gamma = (x, y)$ refers to a point in the combined phase space. It is important to note that the dependence on the work parameter enters only via the Hamiltonian of the system, while the heat reservoir is assumed to be λ -independent. For finite $h_{int}(x, y)$ the equilibrium distribution of the system of interest has to be described by a modified Boltzmann-factor $p_S(x; \lambda) \propto \exp[-\beta H^*(x; \lambda)]$ where $H^*(x; \lambda)$ is a potential of mean force [2, 8]:

$$\begin{aligned} H^*(x; \lambda) &= H(x; \lambda) - \beta^{-1} \ln \frac{\int dy \exp[-\beta(H_E(y) + h_{int}(x, y))]}{\int dy \exp[-\beta H_E(y)]} \end{aligned} \quad (4)$$

With this consideration, the left hand side of Eq. (1) is reduced to the ratio of the partition functions of the system of interest only, resulting in the same result as in the case of weak coupling.

One should note a subtlety that applies to systems with rigid boundaries whose position varies with the control parameter $\lambda(t)$ (which is the case e.g. for the free expansion of an ideal gas in a box, or expansion against a piston). Such boundaries have to be regarded as potentials in the Hamiltonian of the system. If one imposes instead time-dependent constraints, the resulting Hamiltonian evolution does not conserve phase space volume thus leading to an apparent violation of Jarzynski's relation. The correct procedure assumes a potential strength depending on a parameter ϵ which in the limit $\epsilon \rightarrow \infty$ becomes a rigid boundary. As a consequence evaluating the exponential work average in Eq. (1) requires correct ordering of the thermodynamic limit (number of repetitions) before the ϵ limit [15].

B. Derivation for stochastic microscopic reversible dynamics

In [9] the non-equilibrium work theorem was derived under the assumption of stochastic Markovian microscop-

ically reversible dynamics. The crucial condition of microscopic reversibility is formulated as [9]:

$$\frac{P(x(t)|\lambda(t))}{P(\bar{x}(-t)|\bar{\lambda}(-t))} = \exp[-\beta Q(x(t), \lambda(t))] \quad (5)$$

where $P(x(t)|\lambda(t))$ is the probability of following the path $x(t)$ during the work process and $P(\bar{x}(-t)|\bar{\lambda}(-t))$ is the probability of the corresponding time-reversed path during the time-reverse of the process, and Q denotes the heat transferred from the heat reservoir to the system. Equation (5) is claimed to hold arbitrarily far from equilibrium and allows for a simple proof of Jarzynski's relation as discussed in [9, 10]. In the following we discuss the derivation of Eq. (5) more thoroughly.

Microscopic reversibility is usually expressed by the condition of detailed balance for the transitions between states a and b [11]:

$$\frac{\omega(a \rightarrow b)}{\omega(b \rightarrow a)} = e^{-\beta(E_b - E_a)}, \quad (6)$$

where the 'states' a, b refer to discrete volume elements in the phase space of the system and $\omega(a \rightarrow b), \omega(b \rightarrow a)$ are the corresponding transition rates. It is argued in [9] that Eq. (5) follows directly by discretizing the path into single time steps, which each obey detailed balance. More precisely, the evolution of the system $x(t) = \{x_0, \dots, x_t\}$ is considered for a fixed sequence of the work parameter $\lambda(t) = \{\lambda_1, \dots, \lambda_t\}$ such that the single time steps can be decoupled into two substeps. First the control parameter is changed, $\lambda_i \rightarrow \lambda_{i+1}$, performing a certain amount of work, and then the system evolves for fixed λ_{i+1} , $x_i \rightarrow x_{i+1}$, exchanging heat with the thermal environment. As a consequence of Markovian dynamics the probability $P(x(t)|\lambda)$ of following a path through phase space under the work process can be expressed as a product of transition probabilities for the discretized quantities $\{x_0, \dots, x_t\}$ and $\{\lambda_1, \dots, \lambda_t\}$ [9]:

$$P(x(t)|\lambda(t)) = \prod_{i=0}^{t-1} P(x_i \rightarrow x_{i+1} | \lambda_{i+1}). \quad (7)$$

Then the ratio of probabilities of a forward path and its corresponding time-reversed path is

$$\begin{aligned} \frac{P(x(t)|\lambda(t))}{P(\bar{x}(-t)|\bar{\lambda}(-t))} &= \prod_{i=0}^{t-1} \frac{P(x_i \rightarrow x_{i+1} | \lambda_{i+1})}{P(x_i \leftarrow x_{i+1} | \lambda_{i+1})} \\ &= e^{-\beta \sum_{i=0}^{t-1} (E(x_{i+1}, \lambda_{i+1}) - E(x_i, \lambda_{i+1}))} \\ &= e^{-\beta Q(x(t), \lambda(t))}. \end{aligned} \quad (8)$$

In the second line it is assumed that, for each time step, a 'detailed balance like' condition holds for the ratios of forward and reverse probabilities analogous to Eq. (6) for fixed values of the work parameter λ . The third line is due to the first law of thermodynamics: the difference in

energy between two successive states is completely supplied by the heat bath if no work can be performed (i.e. for constant λ), $\Delta Q = E(x_{i+1}, \lambda_{i+1}) - E(x_i, \lambda_{i+1})$. Similarly the work performed by the system originates only from a change in λ : $\Delta W = E(x_i, \lambda_{i+1}) - E(x_i, \lambda_i)$. This decoupling is crucial for the derivation of Eq. (5). By considering only transitions $x_i \rightarrow x_{i+1}$ for constant work parameter $\{\lambda_1, \dots, \lambda_t\}$ in Eq. (8), the evolution of the system is effectively reduced to a sequence of static states that obey detailed balance. Crucial to this is the assumption that the transition rates are independent of the rate $\dot{\lambda}$ at which the system is disturbed. One might therefore question whether Eq. (5) is indeed valid away from equilibrium, where detailed balance is not generally expected to hold [12, 13]. If detailed balance is violated under particular conditions, then Eq. (5) also fails. It is thus reasonable to state that the derivation of Eq. (5) is correct under the given assumptions, but that these assumptions do not properly take into account a non-equilibrium evolution of the system.

In the next section we will further discuss microscopic reversibility with regard to the isothermal ideal gas model.

III. THE IDEAL GAS WITH A PISTON: AN ISOTHERMAL MODEL

We consider a one-dimensional classical non-interacting ideal gas in a container with a moving piston (see Fig. 1). Both the end wall of the container and the piston are connected to a heat reservoir which keeps the gas at constant temperature. This heat reservoir is assumed to be in thermal equilibrium such that its degrees of freedom are distributed according to a canonical Boltzmann distribution. Interactions between the system and its thermal environment are modeled by stochastic scattering at the boundaries: when the gas particle reaches either side of the container a completely inelastic collision takes place and the particle loses all its kinetic energy. It receives a new stochastic velocity which is sampled from the probability distribution of the heat reservoir, independent of its former velocity. We refer to this situation as a strong thermal coupling between system and environment.

Due to flux conservation, the probability distribution of the new particle velocity v_{out} after a particular collision with the fixed (left-hand) boundary (proportional to the flux of particles leaving the boundary) therefore takes the form [16]:

$$\rho_B(v_{out}) = \frac{|v_{out}|}{k_B T} \exp\left(-\frac{v_{out}^2}{2k_B T}\right). \quad (9)$$

Under equilibrium conditions, this boundary condition yields the Boltzmann distribution for the velocities of particles in the container volume. A similar expression applies to the distribution of velocities assigned at the

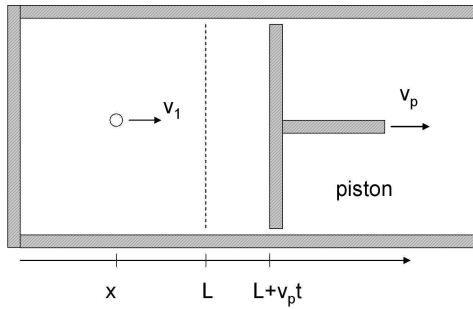


FIG. 1: The ideal gas confined in a container with a piston. The initial position and velocity of a single gas particle are denoted x and v_1 , and the piston has velocity v_p . The (one-dimensional) gas is initially confined to a length L .

moving boundary (the piston), but with the important distinction that this is the distribution of out-going velocities *in the frame of the piston*. Movement of the piston therefore results in a non-zero mean streaming velocity in the laboratory frame.

In contrast to the commonly used Gaussian or Nosé-Hoover thermostating schemes, this stochastic boundary thermostat is non-deterministic and non-time-reversible (in the lab frame). Nevertheless it provides a valid physical model of the heat bath interaction. Furthermore, since no potential acts on the ideal gas particles, their energy is purely kinetic and completely determined by the canonical probability distribution of the heat reservoir.

An important property of the heat reservoir in this model is its dependence on the work parameter λ , which is more precisely a λ -dependence of its center of mass. In the counter-intuitive context of the Jarzynski equation this should not be considered trivial. As has been mentioned in the previous section, the derivation of the non-equilibrium work theorem assumes λ -dependence only for the Hamiltonian of the system, not for the heat reservoir. Here, that assumption is violated by the moving piston, which, by definition, changes its position as a function of λ and, at the same time, thermostats the system. Although a rigorous and general treatment of this issue would require a Hamiltonian description of the heat reservoir, the stochastic model provides important insight.

The isothermal ideal gas is subjected to the following thermodynamic process (see Fig. 2). In the initial state the gas is in equilibrium with the heat bath, confined to the (one-dimensional) ‘volume’ L by a fixed position of the piston. An individual gas particle samples its velocity from a Maxwell-Boltzmann distribution $\rho_{MB}(v)$ and its position from the uniform distribution $1/L$. Then the external work process begins by moving the piston outwards at constant speed v_p for a time period τ . Work is performed when gas particles collide with the piston during its movement. (The retreating piston then does a

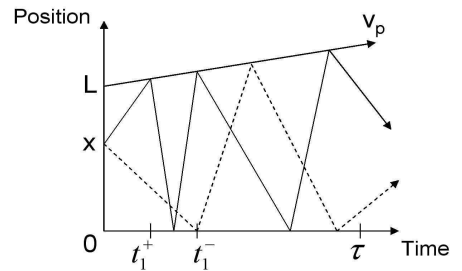


FIG. 2: Position-time-diagram for the work process. The particle performs work by hitting the piston during the time of its movement τ . Example trajectory for a positive (negative) initial velocity is denoted by a solid (dotted) line.

negative amount of work on the gas.) When the piston stops the gas thermalizes to the final equilibrium state at volume $L + v_p\tau$.

With regard to the previous discussion we identify the work parameter λ as the position of the piston and the work rate as dependent on $\dot{\lambda} \equiv v_p$. Since the speed of the heat-transfer mechanism remains fixed in this model, the piston velocity determines the reversibility or irreversibility of the process. In the limit $v_p \rightarrow 0$ with $\tau \rightarrow \infty$, we perform a quasistatic reversible expansion of the gas. In the converse case of large piston speed and short τ , the bulk of the gas remains in the initial part of the container after the piston has stopped, although the volume is extended. Subsequent equilibration is only completed sometime after the full volume has been explored by the gas.

A. Microscopic reversibility

It has been pointed out in Sec. II B that the validity of the non-equilibrium work theorem for stochastic dynamics relies on the condition of microscopic reversibility. The isothermal ideal gas model offers the opportunity to check Crooks’s assumption (Eq. (5)) against a physically motivated, realistic case. In order to determine the transition rates of Eq. (6) we have to discretize the phase space of the system into intervals of size Δx , Δv . As the internal energy is only changed by inelastic collisions with the boundaries, the discussion can be reduced to a single collision event in the vicinity of the piston.

First we consider the equilibrium case, where the work parameter (i.e. the position of the piston) is fixed at L , and the probability of finding a gas particle at x is x/L . The particle is in state a if it occupies the phase space element $[v_a, v_a + \Delta v] \times [L - \Delta x, L]$ and in state b if it is found in $[v_b, v_b + \Delta v] \times [L - \Delta x, L]$. Let a and b refer to the same position interval (but different velocities) before and after the collision.

A transition rate $\omega(a \rightarrow b) = p(b|a)/T$ is given in terms of a conditional probability $p(b|a)$, which is the probability of the system being found in state b , given that it was in state a some short time T earlier. Due to the statistical independence of v_a and v_b , the transition probability for the velocity is simply $p(v_b|v_a) = \rho_B(v_b)$. We find, for the ratio of forward and reverse transitions, upon cancellation of the identical position dependence,

$$\frac{\omega(a \rightarrow b)}{\omega(b \rightarrow a)} = \frac{\rho_B(v_b)\Delta v/T_a}{\rho_B(v_a)\Delta v/T_b}, \quad (10)$$

where the time for the transition $T_i = \Delta x/|v_i|$ depends on the incoming velocity. It then follows from Eq. (9) that the ratio of transition rates between these states is equal to the Boltzmann factor of their energy difference. We are therefore reassured that, in the absence of external work, the model respects the detailed balance condition, Eq. (6).

In the non-equilibrium case, the movement of the piston has to be taken into account. In the context of the discussion in Sec. IIB, the derivation of Crook's microscopic reversibility condition, Eq. (5), is tantamount to fixing the work parameter (the piston position) at a succession of different values $L(t) = \{L + v_p t_1, \dots, L + v_p \tau\}$, but disregarding the momentum exchange resulting from the movement. With that omission, the detailed balance condition follows trivially for each fixed piston position $L + v_p t_i$ since we can simply repeat the calculations above with the new position $L = L + v_p t_i$. Equation (5) would then follow.

However, the transition rates in the non-equilibrium case have to take into account the dynamic change of the work parameter. It is easy to see that this leads to a violation of detailed balance. As in the equilibrium case the initial state a refers to the phase space element $[v_a, v_a + \Delta v] \times [L - \Delta x, L]$. The transition time for the forward process is now $T_a = \Delta x/(v_a - v_p)$ since the particle needs longer to catch the moving piston which is at position $L + v_p T_a$ when the collision takes place. Accordingly state b occupies the phase space element $[v_b, v_b + \Delta v] \times [L + v_p T_a - \Delta x, L + v_p T_a]$. For the reverse transition we consider the initial state b for an inward-moving piston with speed $-v_p$. The transition time is now $T_b = \Delta x/(v_b + v_p)$. Assuming that the reverse collision takes place with outgoing velocity v_a , the new particle position is now in the interval $[L + v_p(T_a - T_b) - \Delta x, L + v_p(T_a - T_b)]$, different from state a . Indeed the reverse transition $b \rightarrow a$ only takes place if $T_a = T_b$ or consequently $v_b = v_a - 2v_p$, which would be the case if the collision were elastic. Otherwise, state a cannot be reached in the reverse process, hence the transition rate $\omega(b \rightarrow a)$ is zero. One immediately realizes that Eq. (6) has to fail under these conditions. The dependence of the particle position after the collision on the transition time is crucial for the breakdown of detailed balance in the dynamic case. As a result we observe that the isothermal ideal gas model violates microscopic reversibility for a moving piston.

B. The exponential work average

In order to verify the non-equilibrium work theorem for the ideal gas with stochastic boundary conditions, the main task is to evaluate the exponential work average of the isothermal expansion process. The free energy difference on the other hand can be calculated from simple thermodynamic considerations:

$$\Delta F = \ln \frac{Z_A}{Z_B} = N \ln \left(\frac{L}{L + v_p \tau} \right). \quad (11)$$

Throughout the calculations we set the temperature parameter $\beta = 1$ without loss of generality. In the case of a non-interacting gas the partition functions Z_A (initial state) and Z_B (final equilibrated state) both factorise and one can effectively reduce the calculations to a single particle $N = 1$. This gas molecule performs work if it can hit the piston during its movement, i.e. during the time period τ . Since it obtains a new randomly chosen velocity upon reaching either the wall or the piston, the time t_k for k bounces to occur depends on all realized velocities v_1, \dots, v_k such that $t_k = t_k(\{v_1, \dots, v_k\}, x; L, v_p)$. Throughout this article we shall use the convention that the variables v_i always refer to the velocity of the particle in the reference frame of the particular wall, where this velocity has been obtained. Since the new velocity of the particle after collision with the piston is thus given in the frame of the piston, the following recursion relation holds, as can easily be verified by inspecting the position-time diagram (Fig. 2):

$$t_k = \frac{L}{v_k - v_p} + \frac{v_k}{v_k - v_p} t_{k-1}. \quad (12)$$

Here we consider always the modulus of velocities $v_i = |v_i|$. As a consequence we have to distinguish between a positive or negative initial velocity v_1 since, in each case, different bounces contribute to the work average. For a positive initial velocity the odd-numbered bounces yield the work contribution and the time for the first bounce is $t_1^+ = (L - x)/(v_1 - v_p)$; for a negative sign the even bounces contribute with $t_1^- = x/v_1$. Hence the exponential work average can be written as:

$$\langle e^{-W} \rangle = \langle e^0 \rangle_+ + \langle e^0 \rangle_- + \langle e^{-W} \rangle_+ + \langle e^{-W} \rangle_-, \quad (13)$$

where $+/-$ refers to a positive or negative initial velocity and the average is to be taken over the uniformly distributed initial position x and all velocities v_1, \dots, v_k . The zero-work contributions to the exponential work average are then determined according to:

$$\begin{aligned} \langle e^0 \rangle_+ &= \frac{1}{L} \int_0^L dx \int_0^\infty dv_1 \rho_{MB}(v_1) \\ &\quad \times [\Theta(t_1^+ - \tau)\Theta(t_1^+) + \Theta(-t_1^+)], \\ \langle e^0 \rangle_- &= \frac{1}{L} \int_0^L dx \int_0^\infty dv_1 \rho_{MB}(v_1) \int_0^\infty dv_2 \rho_B(v_2) \\ &\quad \times [\Theta(t_2^- - \tau)\Theta(t_2^-) + \Theta(-t_2^-)]. \end{aligned} \quad (14)$$

Here $\Theta(x)$ denotes the Heaviside theta function. Recall that a negative amount of work is performed on the gas by the piston, since we consider an expanding volume. The total work W is given by summation of the contributions due to the individual bounces against the piston. The single bounce contributions,

$$w_i = -\Delta\mathbf{p} \cdot \mathbf{v}_p = -(v_i + v_{i+1})v_p + v_p^2, \quad (15)$$

are determined by the momentum transfer $\Delta\mathbf{p} = \mathbf{v}_i - (\mathbf{v}_{i+1} - \mathbf{v}_p)$, where we have set the particle mass to unity without loss of generality. Here, \mathbf{v}_{i+1} is always the new velocity of the particle after collision with the piston and is therefore given in the frame of the piston. The shift $-\mathbf{v}_p$ takes this into account when considering the momentum transfer in the laboratory frame, leading to the term $+v_p^2$ in the work contribution Eq. (15). The w_i are statistically independent random variables.

The average of the exponential work can be expressed as a series in the number of bounces n with the piston:

$$\begin{aligned} \langle e^{-W} \rangle_+ &= \frac{1}{L} \sum_{n=1}^{\infty} \int_0^L dx \int_0^{\infty} dv_1 \dots \int_0^{\infty} dv_{2n+1} \times \\ &P_+(\{v_1, \dots, v_{2n+1}\}, 2n-1; \tau) e^{-\sum_{i=1}^n w_{2i-1}} \end{aligned} \quad (16)$$

We introduce the joint probability distribution function $P_+(\{v_1, \dots, v_{2n+1}\}, 2n-1; \tau)$ which determines the probability of a particular realization of velocities $\{v_1, \dots, v_{2n+1}\}$ and of exactly n bounces with the piston resulting, within the time period τ :

$$\begin{aligned} P_+(\{v_1, \dots, v_{2n+1}\}, 2n-1; \tau) &= \rho_{MB}(v_1)\rho_B(v_2)\dots\rho_B(v_{2n+1})[\Theta(\tau - t_{2n-1}^+) \times \\ &\Theta(t_1^+)\dots\Theta(t_{2n-1}^+) - \Theta(\tau - t_{2n+1}^+)\Theta(t_1^+)\dots\Theta(t_{2n+1}^+)] \end{aligned} \quad (17)$$

An analogous expression holds for $\langle e^{-W} \rangle_-$:

$$\begin{aligned} \langle e^{-W} \rangle_- &= \frac{1}{L} \sum_{n=1}^{\infty} \int_0^L dx \int_0^{\infty} dv_1 \dots \int_0^{\infty} dv_{2n+2} \times \\ &P_-(\{v_1, \dots, v_{2n+2}\}, 2n; \tau) e^{-\sum_{i=1}^n w_{2i}} \end{aligned} \quad (18)$$

with the joint probability distribution

$$\begin{aligned} P_-(\{v_1, \dots, v_{2n+2}\}, 2n; \tau) &= \rho_{MB}(v_1)\rho_B(v_2)\dots\rho_B(v_{2n+2})[\Theta(\tau - t_{2n}^-) \times \\ &\Theta(t_1^-)\dots\Theta(t_{2n}^-) - \Theta(\tau - t_{2n+2}^-)\Theta(t_1^-)\dots\Theta(t_{2n+2}^-)] \end{aligned} \quad (19)$$

The difficulties in evaluating the exponential work average originate primarily from the integration over the theta function $\Theta(\tau - t_k(\{v_1, \dots, v_k\}, x))$ where t_k is given by the recursion relation (12). Therefore we resort to a numerical investigation of the average, Eq. (13), in Sec. IV and in Sec. III C below to an analytical but approximate evaluation, which tends to the exact answer in a well controlled limit.

C. The limit of a fast moving piston

The assumption that at most one bounce takes place between particle and piston yields an approximation that is analytically tractable and becomes exact in the limit of a fast moving piston while the volume extension is small compared with the original volume. In this case, $L \gg v_p \tau \gg \tau$ where velocities are measured in units of the thermal velocity since $\beta \equiv m \equiv 1$. In Ref. [6] the one-bounce approximation validated the non-equilibrium work theorem for the adiabatic ideal gas expansion by considering the $n = 1$ approximation of the work distribution. Here we establish this result by calculating the one-bounce approximation of the exponential work average Eq. (13). According to the Jarzynski relation this average should yield the exponential of the free energy difference Eq. (11):

$$e^{-\Delta F} = 1 + \frac{v_p \tau}{L}. \quad (20)$$

The $n = 1$ approximation of the exponential work average consists of four contributions:

$$\langle e^{-W} \rangle_{n=1} = \langle e^0 \rangle_+ + \langle e^0 \rangle_- + \langle e^{-w_1} \rangle_+ + \langle e^{-w_2} \rangle_- \quad (21)$$

Here $\langle e^0 \rangle_+$ and $\langle e^0 \rangle_-$ are given as in (14). The two one-bounce contributions, from Eqs. (16) and (18), are:

$$\begin{aligned} \langle e^{-w_1} \rangle_+ &= \frac{1}{L} \int_0^L dx \int_0^{\infty} dv_1 \rho_{MB}(v_1) \int_0^{\infty} dv_2 \rho_B(v_2) \\ &\times \Theta(\tau - t_1^+) \Theta(t_1^+) e^{-w_1}, \\ \langle e^{-w_2} \rangle_- &= \frac{1}{L} \int_0^L dx \int_0^{\infty} dv_1 \rho_{MB}(v_1) \int_0^{\infty} dv_2 \rho_B(v_2) \\ &\times \int_0^{\infty} dv_3 \rho_B(v_3) \Theta(\tau - t_2^-) \Theta(t_2^-) e^{-w_2} \end{aligned} \quad (22)$$

In the appendix we derive the following results in the limit $L \gg v_p \tau \gg \tau$:

$$\begin{aligned} \langle e^0 \rangle_+ + \langle e^0 \rangle_- &\rightarrow 1, \\ \langle e^{-w_2} \rangle_- &\rightarrow 0. \end{aligned} \quad (23)$$

In this section we calculate the dominant contribution $\langle e^{-w_1} \rangle_+$. This average takes the explicit form:

$$\begin{aligned} \langle e^{-w_1} \rangle_+ &= \frac{1}{\sqrt{2\pi}L} \int_0^{\infty} dv_2 v_2 e^{-(v_2 - v_p)^2/2} \int_0^L dx \\ &\times \int_{v_p}^{\infty} dv_1 e^{-(v_1 - v_p)^2/2} \Theta\left(\tau - \frac{L - x}{v_1 - v_p}\right). \end{aligned} \quad (24)$$

The integrations over the v_1 and v_2 variables are independent. For $v_p \gg 1$ the integral over v_2 yields $\sqrt{2\pi}v_p$. Making the substitution $v = v_1 - v_p$ and changing the

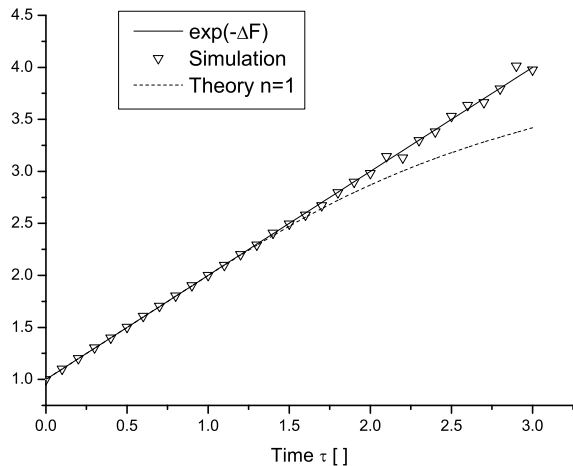


FIG. 3: Comparison of the $n = 1$ approximation Eq. (21) (dotted line) with the exponential free energy difference $1 + v_p\tau/L$ (solid line) for $v_p = 1$. The triangles show the numerical results for the same parameter values.

argument of the theta function, we can calculate the remaining integral in a straightforward way:

$$\begin{aligned}
 \langle e^{-w_1} \rangle_+ &= \frac{v_p}{L} \int_0^\infty dv e^{-v^2/2} \int_0^L dx \Theta(x - (L - v\tau)) \\
 &= \frac{v_p\tau}{L} \int_0^{L/\tau} dv v e^{-v^2/2} + v_p \int_{L/\tau}^\infty dv e^{-v^2/2} \\
 &\rightarrow \frac{v_p\tau}{L}.
 \end{aligned} \tag{25}$$

The last limit holds for $L \gg v_p\tau$ such that, in combination with Eq. (23), we obtain the expected result Eq. (20). Fig. 3 shows a plot of $\langle e^{-W} \rangle_{n=1}$ for $L = 1$, $v_p = 1$ together with Eq. (20). According to our numerical results presented below a piston velocity in this regime drives the system sufficiently out of equilibrium. One notes that here the one-bounce approximation reproduces Jarzynski's relation for times up to $\tau \approx 1$, so that this approximation actually holds in a broader regime than the limits considered above. For larger times τ the single bounce is no longer sufficient to approximate the averaged exponential and higher order terms have to be taken into account.

IV. NUMERICAL METHOD

We have performed a numerical study of the isothermal ideal gas model in order to evaluate the exponential work average, Eq. (13), for arbitrary numbers of bounces n . This thermostated ‘Molecular Dynamics’ simulation essentially consists of picking random numbers from the velocity distributions $\rho_{MB}(v_1)$, $\rho_B(v_i)$ and calculating the

bouncing times t_k (Eq. (12)). When the velocity configuration allows the particle to hit the piston one or more times within τ , the corresponding work values (Eq. (15)) are recorded.

A general problem of the applicability of the non-equilibrium work theorem is the convergence of the average. Since the exponential $\exp[-\beta W]$ emphasizes small work values, one effectively has to sample the far left tail of the work distribution. In our model, the work performed on the system is always negative and the average is dominated by those events that lead to a large negative work contribution. A general discussion of the convergence problem can be found in [14] with particular focus on the ideal gas and piston, but for an adiabatic work process (elastic collisions). It was shown that, in this particular case, the number of realizations needed in order to sample the dominant part of the average grows exponentially with the system size [6] or, more generally, this number is proportional to the exponential of the averaged work that is dissipated during the reverse process [14]. It is not obvious whether these results can be directly applied to the isothermal ideal gas model under consideration. In contrast to the adiabatic case, the occurrence of many collisions in a particular realization does not necessarily imply a large work contribution (a dominant event), as the work depends not only on the incoming velocity but also on the statistically independent outgoing velocity.

The numerical results below have been obtained with typically 10^6 realizations per data point, which yields such excellent convergence of the exponential average that error bars have been omitted in the figures. All less relevant parameters are set to unity for simplicity and units are non-dimensional. This includes the length of the initial volume, $L = 1$, and the inverse temperature, $\beta = 1$, which sets the width of ρ_{MB} and ρ_B . Accordingly the thermal velocity is unity as well so, for piston speeds $v_p > 1$, we are in a regime where the tail of the initial velocity distribution contributes the work. The plots show the average exponential work for a given v_p when τ is varied from zero up to the extended volume $v_p\tau = L$ (i.e. the volume is doubled).

A. Numerical results

We report the following results. In the limit of a very slow (quasistatic) expansion we obtain the isothermal free energy difference, Eq. (11), from both the work average $\langle W \rangle$, in accordance with the Second Law for a reversible process, and the exponential work average (see Fig. 4). The quasistatic regime is found at $v_p \leq 0.001$. If we pull the piston at a higher speed the work average deviates noticeably from the free energy difference, indicating the onset of dissipation. The dissipated work is $W_d = \langle W \rangle - \Delta F$ and we are effectively performing an irreversible non-equilibrium experiment. On the other

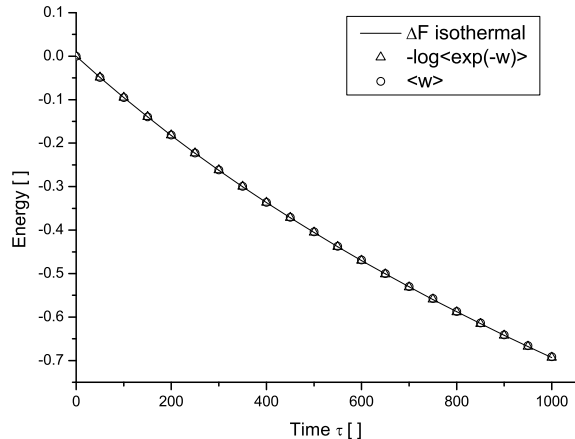


FIG. 4: Reversible isothermal expansion of the ideal gas for $v_p = 0.001$.

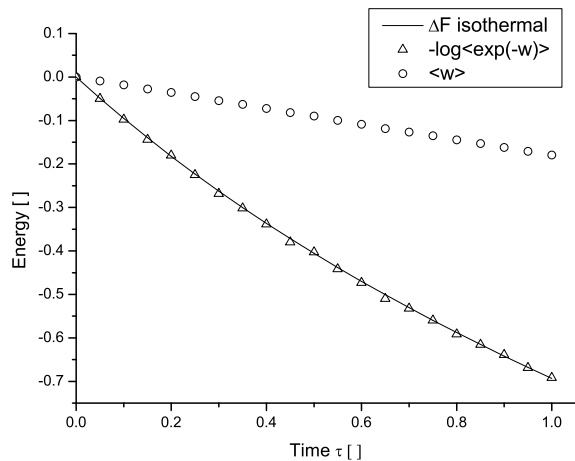


FIG. 5: Irreversible isothermal expansion of the ideal gas for $v_p = 1$. One notices the dissipated work as the difference between the average work and ΔF .

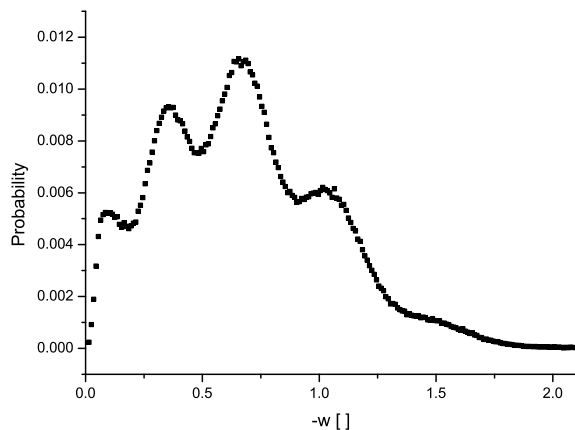


FIG. 6: Work distribution of the isothermal expansion process for $v_p = 0.1$.

hand the negative logarithm of the exponential work average still agrees with the isothermal free energy difference as predicted by the non-equilibrium work theorem (see Fig. 5). This is the main result of our numerical investigation: that the Jarzynski relation holds in the non-equilibrium regime of our model ($v_p > 0.001$) despite the fact that the model does not belong to the class of systems for which the theorem was derived, but has a more physically motivated coupling to its heat bath. The simulation shows excellent convergence up to $v_p \approx 1$. For higher piston speeds it becomes increasingly difficult to sample the tails of the velocity distributions.

In Fig. 6, we present the full distribution of work values determined numerically for $v_p = 0.1$. The distribution exhibits a multi-peak structure, demonstrating that, in imposing only one constraint on the distribution, the non-equilibrium work theorem does not confine it to adopt a simple shape.

B. The adiabatic piston model revisited

An ideal gas with a piston was previously investigated by Lua and Grosberg [6] for the case of adiabatic expansion, i.e. perfectly elastic collisions at the boundaries. We can reproduce their adiabatic model by considering elastic, energy-conserving (and therefore deterministic) collisions instead of completely inelastic, stochastic ones. Consequently the probability distribution, Eq. (9), is substituted by

$$\rho_B(v_{out}) = \delta(v_{out} - (v_{in} - v_p)), \quad (26)$$

when the incoming velocity is v_{in} . The shift $-v_p$ is due to our convention for the velocity variables v_i (see Sec. III B) and is explained as follows. When the particle collides with the piston, v_{in} is given in the lab frame, whereas v_{out} refers to the velocity in the piston frame, therefore $v_{out} = v_{in} - v_p$. For the subsequent bounce against the fixed wall, v_{in} is given in the piston frame, but v_{out} refers to the velocity in the lab frame, hence again $v_{out} = v_{in} - v_p$ if the collision is elastic. Overall the distribution Eq. (26) is valid for bounces on the wall and on the piston side. However, when the initial velocity is negative, the first collision takes place with the fixed wall such that both incoming and out-going velocities refer to the lab frame. In this case only, $\rho_B(v_2) = \delta(v_2 - v_1)$. Using these distributions, the bouncing times $t_k(\{v_1, \dots, v_k\}, x)$ in Eq. (12) are reduced to the correct elastic counterpart $t_k(v_1, x)$ and the averages, Eqs. (16) and (18), are evaluated by integrating over the initial velocity v_1 and initial position x . In this case the non-equilibrium work theorem has been proven to hold exactly for all parameter values [6].

We observe a particularly interesting feature of the non-equilibrium work theorem for the adiabatic expansion of the ideal gas, that is worth highlighting. In the quasistatic limit of a very slow moving piston, the work

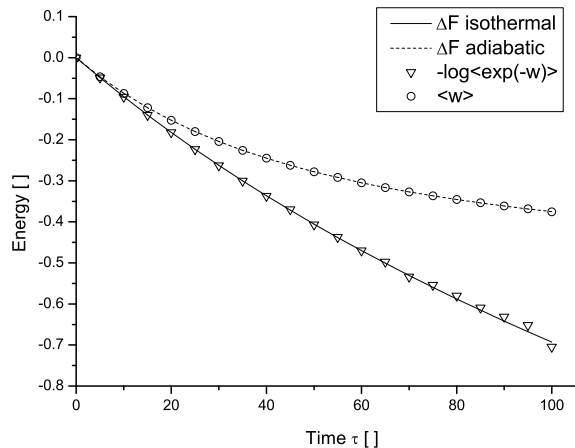


FIG. 7: Reversible adiabatic expansion of the ideal gas for $v_p = 0.01$. Both the free energy difference for isothermal expansion and that for adiabatic expansion are obtained from the same work values.

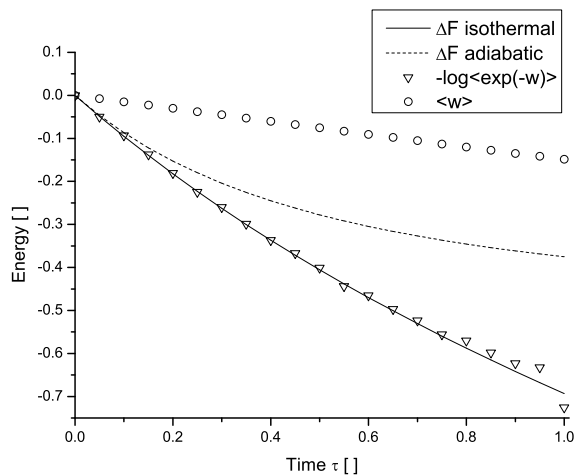


FIG. 8: Irreversible adiabatic expansion of the ideal gas for $v_p = 1$. The dissipated work is observed.

average $\langle W \rangle$ yields the free energy difference of the adiabatic expansion of the gas as one would expect. The exponential work average, on the other hand, yields the free energy difference of the *isothermal* expansion of the gas. This can clearly be seen in Fig. 7. The averages $\langle W \rangle$ and $-\log \langle \exp(-W) \rangle$ respectively show excellent agreement with the isothermal free energy difference Eq. (11) (solid line) and the adiabatic free energy difference (dotted line). The latter reads explicitly:

$$\Delta F_{\text{ad}} = \frac{1}{2} \left(\left(\frac{L}{L + v_p \tau} \right)^2 - 1 \right). \quad (27)$$

Thus, by performing an adiabatic experiment, we ob-

tain information about both an adiabatic and an isothermal system. If we consider the irreversible case for this adiabatic process, we observe again the dissipated work W_d (see Fig. 8) which is responsible for the deviation of $\langle W \rangle$ from ΔF_{ad} . On the other hand ΔF (the isothermal result) can still be determined by evaluating the exponential work average. By application of the non-equilibrium work theorem we can perform an adiabatic or an isothermal experiment and obtain the same result, the isothermal free energy difference. From a numerical point of view the isothermal simulation proves slightly more advantageous because more sampling takes place during each realization of the protocol leading to a faster convergence of the exponential average.

V. CONCLUSION

The main result of our investigation is the validation, both numerically and analytically, of the non-equilibrium work theorem for the isothermal expansion of an ideal gas against a piston. Although the analytical calculation is restricted to the limit of a fast-moving piston and small volume extension, the simulation confirms the result for a wide range of parameter values. The two main characteristics of the model under consideration should be emphasized again. First, the isothermal model exhibits strong thermal coupling between system and heat reservoir, an important and more physically relevant extension to the ideal gas models previously discussed in the literature [6, 7]. Second, both the system and the heat reservoir depend on the work parameter λ , violating the assumptions in Jarzynski's original derivation [8]. Furthermore it has been shown that microscopic reversibility is broken due to the moving and thermostating piston, such that Crooks' derivation [9] (which assumes transition rates independent of λ) does not hold either. We have thus identified a regime where Jarzynski's relation might have been expected to fail, however it appears that the validity of the non-equilibrium work theorem is not affected.

It is a pleasure to acknowledge helpful discussions with C. Jarzynski. This work was funded by EPSRC Grant GR/T24593/01. RMLE is funded by the Royal Society.

Appendix

We show the convergence of the averages $\langle e^0 \rangle_+ + \langle e^0 \rangle_- \rightarrow 1$ and $\langle e^{-w_2} \rangle_- \rightarrow 0$ in the limit of a fast moving piston and small extended volume. In order to calculate the averages, we resolve the theta functions with respect to v_1 and obtain multiple dependent integrals.

First the zero-work contribution for positive initial ve-

locity:

$$\langle e^0 \rangle_+ = \frac{1}{\sqrt{2\pi L}} \int_0^L dx \int_0^{(L-x)/\tau+v_p} dv_1 e^{-v_1^2/2}. \quad (28)$$

The zero-work contribution for negative initial velocity reads

$$\begin{aligned} \langle e^0 \rangle_- &= \frac{1}{\sqrt{2\pi L}} \int_0^L dx \int_{L/\tau+v_p}^\infty dv_2 v_2 e^{-v_2^2/2} \\ &\quad \times \int_0^{(v_2 x / (v_2 - v_p) \tau - L)} dv_1 e^{-v_1^2/2} \\ &\quad + \frac{1}{2} \int_0^{L/\tau+v_p} dv_2 v_2 e^{-v_2^2/2}, \end{aligned} \quad (29)$$

and the one-bounce contribution for negative initial velocity is given as

$$\begin{aligned} \langle e^{-w_2} \rangle_- &= \frac{1}{\sqrt{2\pi L}} \int_0^L dx \int_{L/\tau+v_p}^\infty dv_2 v_2 e^{-(v_2 - v_p)^2/2} \\ &\quad \times \int_{v_2 x / ((v_2 - v_p) \tau - L)}^\infty dv_1 e^{-v_1^2/2} \\ &\quad \times \int_0^\infty dv_3 v_3 e^{-(v_3 - v_p)^2/2}. \end{aligned} \quad (30)$$

We discuss the limiting behaviour as follows. The integration over v_1 always leads to the error function (Eq. (28) and Eq. (29)) or to a sum of a constant and the error function (Eq. (30)). These expressions are always bounded by a constant for arbitrary values of the argument. For Eq. (28) we immediately obtain, in the limit of small τ and large v_p : $\langle e^0 \rangle_+ \rightarrow 1/2$.

For the two other cases we see that the integrand of the second integration is multiplied by a bounded term. Thus the convergence of both averages depends only on the v_2 -integration. If we note that $x \exp(-x^2)$ is already small for $x > 1$, we see that the first term in Eq. (29) is clearly vanishing in the considered limit and the result is $\langle e^0 \rangle_- \rightarrow 1/2$. In the case of Eq. (30) we simply observe that there is an additional shift $v_2 - v_p$ in the argument of the exponential such that this integrand decays to zero even faster than the first term in Eq. (29). As a result the relations Eq. (23) are valid and the single bounce approximation analytically confirms the non-equilibrium work-theorem in the regime $L \gg v_p \tau \gg \tau$.

-
- [1] F. Ritort, *Poincaré Seminar* 2, 195 (2003)
 - [2] S. Park and K. Schulten, *J. Chem. Phys.* 120, 5946 (2004)
 - [3] C. Jarzynski, *Phys. Rev. Lett.* 78, 2690 (1997)
 - [4] J. Liphardt et al., *Science* 296, 1832 (2002)
 - [5] E.G.D. Cohen and D. Mauzerall, *J. Stat. Mech.*, P07006 (2004)
 - [6] R.C. Lua and A.Y. Grosberg, *J. Phys. Chem. B* 109, 6805 (2005)
 - [7] I. Bena, C. Van den Broeck and R. Kawai, *Europhys. Lett.* 71, 879 (2005)
 - [8] C. Jarzynski, *J. Stat. Mech.*, P09005 (2004)
 - [9] G.E. Crooks, *J. Stat. Phys.* 90, 1481 (1998)
 - [10] G.E. Crooks, *Phys. Rev. E* 60, 2721 (1999)
 - [11] N.G. Van Kampen, *Stochastic Processes in Physics and Chemistry* (North-Holland, Amsterdam, 1985)
 - [12] M.J. Klein, *Phys. Rev.* 97, 1446 (1955)
 - [13] R.M.L. Evans, *Phys. Rev. Lett.* 92, 150601 (2004)
 - [14] C. Jarzynski, cond-mat/0603185
 - [15] S. Pressé and R. Silbey, *J. Chem. Phys.* 124, 054117 (2006)
 - [16] J.L. Lebowitz and H. Spohn, *J. Stat. Phys.* 19, 633 (1978)