On the fluctuation theorem for the dissipation function and its connection with response theory.

Author
J. Evans, Denis, Bernhardt, Debra, R. Williams, Stephen

Published
2008

Journal Title
Journal of Chemical Physics

DOI
https://doi.org/10.1063/1.2812241

Copyright Statement
Copyright 2008 American Institute of Physics. This article may be downloaded for personal use only. Any other use requires prior permission of the author and the American Institute of Physics. along with the following message: The following article appeared in J. Chem. Phys. 128, 014504 (2008) and may be found at http://dx.doi.org/10.1063/1.2812241

Downloaded from
http://hdl.handle.net/10072/21822

Link to published version
http://jcp.aip.org/
Nonequilibrium Free Energy Relations For Thermal Changes

Stephen R. Williams,¹ Debra J. Searles,² and Denis J. Evans¹

¹Research School of Chemistry, The Australian National University, Canberra, ACT 0200, Australia
²Nanoscale Science and Technology Centre, School of Biomolecular and Physical Sciences, Griffith University, Brisbane, Qld 4111, Australia

(Dated: April 12, 2008)

Abstract

The Jarzynski Equality and the Crooks Fluctuation Theorem enable the calculation of the change in a system's free energy from nonequilibrium path integrals. These relations consider processes where the system is driven out of equilibrium by a mechanical external agent while remaining in contact with a thermal reservoir at a fixed temperature. We generalise these relations to describe processes driven by any type of external agent, be it thermal or mechanical. Attention is given to the case of a system, initially in equilibrium, that is driven through a temperature change by a heat reservoir. The results are cast in a form applicable to experiments.
The Jarzynski equality (JE) [1, 2], the Crooks fluctuation theorem (CFT) [3, 4] and the Evans-Searles fluctuation theorem [5–7] are important new results in nonequilibrium statistical mechanics. To obtain experimentally applicable forms of these theorems which are valid arbitrarily far from equilibrium, it is necessary to introduce a thermal reservoir that is large and remote enough from the system of interest, to effectively remain in equilibrium. To this end models involving synthetic thermostats, which only act on particles in the reservoir, have been developed [7–9]. The system of interest cannot possibly “know” the details of how the thermostat operates due to its remoteness from the system of interest. The details of the synthetic thermostat are then irrelevant to the final result [8]. This approach works well for mechanically driven systems in contact with a fixed temperature reservoir. However, matters are more complicated if we wish to consider a realistic model of a system that is driven away from equilibrium by a reservoir whose temperature is changing, e.g. see [10].

Here we address this issue by considering a system of interest, containing some very slowly relaxing constituents such as soft matter or pitch [11], in contact with a rapidly relaxing reservoir. The reservoir may be formed from a copper block or another highly thermally conductive material. Changing the temperature of the reservoir (say with a thermostatically controlled heat exchanger) then drives the system of interest out of equilibrium. The change in temperature is slow enough that the reservoir may be treated to high accuracy, as undergoing a quasistatic temperature change. The slowly relaxing system of interest is far from equilibrium. We develop generalised versions of the CFT and the JE applied to this system. Importantly the quantities that appear in our theory are physically measurable variables.

We first state the CFT and the JE for the canonical ensemble. The thermodynamic potential is the Helmholtz free energy, A, which is related to the phase space \( \Gamma = (q_1, .., q_N, p_1, .., p_N) \) integral of the negative exponential of the Hamiltonian \( H_0(\Gamma) \), of the system [12],

\[
A(\lambda) = -k_B T \ln Z(\lambda) = -k_B T \ln \left( \int d\Gamma \exp (-\beta H_0(\Gamma, \lambda)) \right) .
\]  

(1)

In this equation \( k_B \) is Boltzmann’s constant, \( T \) is the absolute temperature and \( \beta = 1/(k_B T) \). The functional form of the system’s Hamiltonian may vary parametrically, over the period
$0 < t < \tau$, i.e. $H_0(\Gamma, \lambda(t)) = \sum_{i=1}^{N} p_i \cdot p_i + \Phi(q, \lambda(t))$ where $\Phi$ is the interparticle potential and the system is initially at equilibrium and $\lambda = \lambda_0$. The system may also be driven by an external dissipative field [13–15]. For $t > \tau$ the Hamiltonian’s parametric dependence is fixed at $H_0(\Gamma, \lambda_\tau)$ and the external field is turned off. Over the times $0 < t \leq \tau$, the ensemble is driven away from equilibrium, and if the transformation is halted at $t = \tau$, the system will eventually relax to a new equilibrium state. The work done on the system as a function of time, for a given phase space trajectory is

$$\Delta W_t = H_E(\Gamma(t), \lambda(t)) - H_E(\Gamma(0), \lambda(0)) - k_B T \int_0^t ds \Lambda(\Gamma(s))$$

(2)

where usually $H_E = H_0$ (but not in the case of Nosé-Hoover dynamics [16], as detailed below), $\Lambda = \partial / \partial \Gamma \cdot \dot{\Gamma}$ is the phase space compression factor, determined from the equations of motion [7, 17] and (2) reduces to $W_t = \dot{\lambda} \partial H_0(\Gamma(t), \lambda(t)) / \partial \lambda$ when the external agent is purely parametric [15]. The CFT is then given as

$$\frac{P_f(\Delta W_\tau = B)}{P_r(\Delta W_\tau = -B)} = \exp[-\beta (\Delta A - B)]$$

(3)

where $\Delta A = A(\lambda_\tau) - A(\lambda_0)$, $P_f$ is the probability density of observing $\Delta W_\tau = B$ in a forward trajectory starting from the initial equilibrium given by $\lambda = \lambda_0$ and $P_r$ is the probability density for observing $W_\tau = -B$ in a reverse trajectory but starting from the equilibrium given by $\lambda = \lambda_\tau$. From the CFT it is trivial to obtain the JE (e.g. Ref. [4, 18]) which is,

$$\langle \exp(-\beta \Delta W_\tau) \rangle = \exp(-\beta \Delta A).$$

(4)

We now consider a much more general case for the nonequilibrium free energy theorems (3) & (4). We consider two closed $N$-particle systems: 1, 2. These systems may have the same or different Hamiltonians; it does not matter. They may have the same or different temperatures or volumes; again it does not matter. We also do not care what ensemble the systems are in: microcanonical, canonical or isothermal isobaric, etc. We define a generalised dimensionless “work” $\Delta X_\tau(\Gamma'; 0, \tau)$, for a trajectory of duration $\tau$, as [19]

$$\exp[\Delta X_\tau(\Gamma')] \equiv \frac{P_1(\Gamma'_0, \delta \Gamma'_0) Z(\lambda_1)}{P_2(\Gamma'_\tau, \delta \Gamma'_\tau) Z(\lambda_2)}$$

$$= \frac{f_1(\Gamma'_0) \delta \Gamma'_0 Z(\lambda_1)}{f_2(\Gamma'_\tau) \delta \Gamma'_\tau Z(\lambda_2)}$$

(5)

3
where $Z(\lambda_i)$ is the partition function for system $i$ and $\Gamma'$ is the extended phase space vector which includes the phase space vector $\Gamma$ and may include additional dynamical variables such as the volume or those associated with the thermostat. If the dynamics is deterministic, a phase space trajectory evolves an initial phase vector $\Gamma_0$ at $t = 0$ to $\Gamma'$ at $t = \tau$ [20]. If the system is canonically distributed then $Z(\lambda_i)$ is as given in (1). For other ensembles the partition functions are well known. In (5) $P_i(\Gamma', \delta \Gamma') = f_i(\Gamma')\delta \Gamma'$ is the probability of observing the infinitesimal phase volume $\delta \Gamma'$, centred on the phase vector $\Gamma'$, according to the initial equilibrium distribution function, $f_i$. Although the significance of the variable $X$ might seem obscure at this point, we will show that for particular choices of dynamics and ensemble, it is related to important physical properties. We identify $\|\partial \Gamma' / \partial \Gamma_0\|$ as the Jacobian and note that

$$\|\partial \Gamma'/\partial \Gamma_0\| = \frac{\delta \Gamma'(0,0)}{\delta \Gamma'(t,t)}$$

(6)

due to conservation of the number of ensemble members with respect to time. Since the distribution function is normalized, $\int d\Gamma' f(\Gamma') = 1$, using (5) and the first equality in (6) it is obvious that:

$$\langle \exp(-\Delta X) \rangle_1 = \int d\Gamma' f(\Gamma'_0) \frac{f(\Gamma'_0)\delta \Gamma'_0 Z(\lambda_2)}{f(\Gamma'(t),t)} = \frac{Z(\lambda_2)}{Z(\lambda_1)}$$

(7)

where the brackets $\langle \ldots \rangle_1$ denote an equilibrium ensemble average over the initial (i.e. $f_1(\Gamma')$) distribution. This relationship is very general [19]. It relates the ensemble average of the exponential of a nonequilibrium path integral to equilibrium thermodynamic free energy differences. The paths do not need to be quasistatic paths as in traditional thermodynamics. In common with (3) & (4) other nonequilibrium (even dissipative) processes can be carried out during the period $0 < t < \tau$. It is also a simple matter to prove the following generalised CFT,

$$\frac{P_\tau(\Delta X = B)}{P_\tau(\Delta X = -B)} = \frac{Z(\lambda_2)}{Z(\lambda_1)} \exp(B).$$

(8)

Equations (7) & (8) can be applied to thermal and mechanical changes.

To enable further development we now introduce the equations of motion for our model. We consider thermostatted dynamics with the thermostat only acting on the remote equilibrium reservoir as discussed in the introduction. Although other equations of motion, such
as Gaussian or constant pressure Nosé-Hoover [9] could just as easily be used to the same effect, we will concentrate on the Nosé-Hoover equations of motion [16] for the canonical ensemble,

\[
\dot{q}_i = \frac{p_i}{m_i}, \quad \dot{p}_i = F_i - S_i \alpha p_i
\]

\[
\dot{\alpha} = \left( \frac{\sum_{i=1}^{N} S_i \frac{p_i \cdot p_i}{dN_R k_B T}}{dN_R k_B T} - 1 \right) \frac{1}{\tau^2}, \quad (9)
\]

where \( S_i \) is a switch that is set to unity for particles in the reservoir that the thermostat acts upon and zero otherwise, \( d \) is the Cartesian dimension, \( \tau_\alpha \) is the Nosé-Hoover time constant and \( N_R = \sum_{i=0}^{N} S_i \). For (9) the extended Hamiltonian is \( H_E = H_0 + \frac{d}{2} N_R k_B T \alpha^2 \tau^2 \) and the extended phase space of the system is \( \Gamma' = (\Gamma, \alpha) \). The Liouville equation states:

\[
df/dt = -\Lambda f \quad [17]
\]

and using (9) it is easy to show that, \( k_B T \Lambda = -dN_R k_B T \alpha \equiv \dot{Q} \), where \( \dot{Q} \) is the rate of increase in \( H_E \) due to thermostating alone. The equilibrium distribution function for this system is easily shown to be

\[
f_{eq}(\Gamma, \lambda_i, \alpha) = \frac{\tau_\alpha \sqrt{dN_R/(2\pi)}}{Z(\lambda_i)} \exp(-\beta H_E(\Gamma, \lambda_i, \alpha)) \quad (10)
\]

where \( Z(\lambda_i) \) is given in (1). The streaming solution of the Liouville equation is,

\[
\frac{f(\Gamma(0), \alpha(0), 0)}{f(\Gamma(t), \alpha(t), t)} = \exp\left[ \int_0^t dt' \beta(t') \dot{Q}(\Gamma(t'), \alpha(t')) \right], \quad (11)
\]

which is valid for both equilibrium and nonequilibrium processes, see [7, 15, 17] for details.

Now we consider the special case of transformations using thermostatted dynamics between canonical equilibrium states with the same temperature. Using (10) for the equilibrium distributions \( f_{eq}(\Gamma, \lambda_i, \alpha) \), along with (5), (6) & (11) one sees that \( \Delta X_\tau/\beta \) is the total energy change in the system’s Hamiltonian \( H_E \) minus the energy (i.e. the heat) gained by the system from the thermostat (usually a negative quantity), \( \Delta Q(\Gamma; 0, \tau) = \int_0^\tau dt \dot{Q}(\Gamma'(t)) \):

\[
\Delta X_\tau(\Gamma; 0, \tau) = \beta \left( H_E(\Gamma'(t)) - H_E(\Gamma'(0)) \right) - \beta \Delta Q(\Gamma; 0, \tau)
\]

\[
= \beta \Delta W_\tau(\Gamma; 0, \tau). \quad (12)
\]

The final equality can be obtained by consideration of the First Law of Thermodynamics that states that the total energy change is the sum of contributions from both the heat and the work. Equation (12) shows that in this case (thermostatted dynamics with canonical initial
and final distributions), \( \Delta X_r \) is just the work performed on the system in the transformation multiplied by \( \beta \): \( \Delta X_r(\Gamma; 0, \tau) = \beta \Delta W_r(\Gamma; 0, \tau) \) as can readily be seen by referring to (2) [2, 13-15, 18, 19]. Equation (7) then reduces to the well known JE (4) and equation (8) reduces to the CFT (3). The same result is obtained for the canonical distribution when the dynamics are either thermostatted by a Gaussian thermostat or the dynamics are adiabatic (i.e. unthermostatted). For other ensembles and transformations (7) does not necessarily refer to a work (see [10, 19, 21]).

In deriving (7) & (8) it is assumed that every point in the initial distribution \( f_1 \), which occurs with finite probability at time \( t = 0 \), must evolve to a point at time \( t = \tau \) which occurs with finite probability in the distribution \( f_2 \) and vice versa for the reverse trajectories which are used to compute \( P_r \) in (8), i.e. \( f_1(\Gamma_0') \neq 0 \) and \( f_2(\Gamma_r') \neq 0 \). This requirement is a form of ergodic consistency [7, 15, 17]. For the forward trajectories the system must initially be in the equilibrium state given by \( f_1 \) at time \( t = 0 \) but at time \( t = \tau \) the state need not be given by \( f_2 \). The converse holds for the reverse trajectories which must initially sample the equilibrium state given by \( f_2 \) at time \( t = 0 \) but need not be in state \( f_1 \) at time \( t = \tau \). In the case of the JE (4), if at the end of the protocol \( t = \tau \), the system is not in equilibrium, it does not matter. Any subsequent relaxation processes will have no effect on the work and can only affect the heat. Furthermore at the end of the protocol the system cannot “know” how long that final relaxation process takes or even if the final relaxation is so slow that effectively this relaxation never occurs [22]. Analogous statements hold for (7), with \( \Delta X \) being defined in terms of the ratio of partition functions of the two the equilibrium states regardless of the relaxation that takes place after the protocol has ceased \( (t > \tau) \).

A necessary condition for (7) or (8) to converge is that in the ensemble averaging process the time reversed path of the most probable path, must be observed. If the averaging process is not sufficiently exhaustive for these possibly extremely rare events to be observed, (7) & (8) will give incorrect results. This observation has an immediate impact on the calculation of free energy differences in the thermodynamic limit. This difference must be calculated in finite systems for a series of system sizes and then extrapolation must be employed to obtain the thermodynamic limit.

We now consider applying (7) to an example of a thermal rather than a mechanical process. Consider a system of \( N \) particles whose temperature is changed from \( T_1 \) to \( T_2 \). We do not change the Hamiltonian during this process. For simplicity we consider a canonical
ensemble for the two equilibrium states (10), and use the equations of motion (9). The temperature dependence of the reservoir is achieved by making the Nose-Hoover target temperature \( T(t) \) a time dependent variable \( T(t) \),

\[
\dot{\alpha} = \left( \sum_{i=1}^{N} \frac{S_{i} p_{i} \Delta p_{i}}{dN_R k_B T(t)} - 1 \right) \frac{1}{\tau_{\alpha}^2}.
\]  

The change in temperature is slow enough that the reservoir may be treated as changing quasistatically, at the target temperature \( T(t) \) while the slowly relaxing system of interest is driven out of equilibrium, i.e. it changes irreversibly. Of course if one is just interested in the synthetic dynamics this restriction may be lifted and the temperature can be changed at arbitrary rate. Either way the system of interest will approach the temperature \( T_2 \) in the long time limit \( t/\tau \to \infty \). We use (5) with \( f_1 \) and \( f_2 \) given by (10) at the two different temperatures to obtain

\[
\Delta X_{\tau}(\Gamma'; 0, \tau) = \int_{0}^{\tau} ds \dot{X}(\Gamma'(s)) = \beta_2(H_E(\Gamma'(\tau)) - \beta_1H_E(\Gamma'(0)) - \int_{0}^{\tau} dt \beta(t)\dot{Q}(\Gamma'(t)),
\]  

where \( \beta(t) = 1/k_B T(t) \) is the inverse target temperature and \( \dot{X}(\Gamma'(t)) \) is the generalized “power”. Now if we take the derivative of the extended Hamiltonian while the temperature is changing, but with no other external agent acting on the system, we obtain using (9),

\[
\frac{d}{dt} H_E(\Gamma'(t)) = \dot{Q}(\Gamma'(t)) + \frac{d}{2} N_R k_B T \alpha^2 \tau_{\alpha}^2.
\]  

We then obtain

\[
\frac{d}{dt} \beta(t)H_E(\Gamma'(t)) = -\beta(t) \left[ H_0(\Gamma(t)) \frac{\dot{T}(t)}{T(t)} - \dot{Q}(\Gamma'(t)) \right]
\]  

and the generalized “power” for a change in the target temperature with time is,

\[
\dot{X}(\Gamma(t)) = \dot{\beta}(s)H_0(\Gamma(t), \lambda(t)).
\]  

Note that (17) only depends upon \( \Gamma \) and not the thermostat multiplier \( \alpha \). Equation (7) then becomes
\[ \left< \exp \left( - \int_0^\tau ds \, \dot{\beta}(s) H_0(\Gamma(s)) \right) \right>_1 = \frac{Z^{(2)}}{Z^{(1)}} = \exp [ - \beta_2 A_2 + \beta_1 A_1 ] \]  

where \( A_i \) is the equilibrium Helmholtz free energy of ensemble \( i \). One can see that this equation is consistent with thermodynamics because in the quasistatic limit, equilibrium thermodynamics gives us the relation,

\[ \dot{\beta}(t) \left< H_0(\Gamma(t)) \right>_\text{eq} = \dot{T} \frac{d}{dT} \beta(t) A(t). \]  

The Hamiltonian of the total system may be split in parts representing the system of interest, \( H_{si} \), the reservoir \( H_r \) and the interaction between the reservoir particles and the system of interest particles \( H_{sir} \), giving \( H_0(\Gamma) = H_{si}(\Gamma_{si}) + H_r(\Gamma_r) + H_{sir}(\Gamma) \). Here \( \Gamma_{si} \) is the phase space vector for all the particles in the system of interest, \( \Gamma_r \) is for all the particles in the reservoir. Now by construction we have set up our system such that the changes to \( \left< H_r \right> \) and \( \left< H_{sir} \right> \) are quasistatic. This allows us to take the contributions of these parts of the Hamiltonian through the average appearing in (18),

\[ \left< \exp \left( - \int_0^\tau ds \, \dot{\beta}(s) H_{si}(\Gamma_{si}) \right) \right>_1 \times \exp \left( - \int_0^\tau ds \, \dot{\beta}(s) \left< H_r + H_{sir} \right>_\text{eq} \right) = \exp [ \beta_1 A_1 - \beta_2 A_2 ] \]  

and obtain,

\[ \left< \exp \left( - \int_0^\tau ds \, \dot{\beta}(s) H_{si}(\Gamma_{si}) \right) \right>_1 = \exp [ \beta_1 A_{si,1} - \beta_2 A_{si,2} ], \]  

where \( \beta \) and \( T \) are given by the temperature of the reservoir, and

\[ \beta_2 A_{si,2} - \beta_1 A_{si,1} = \int_0^\tau ds \, \dot{\beta}(s) \left< H_{si}(\lambda(s)) \right>_\text{eq}. \]  

For temperature changes at finite rates, the thermodynamic temperature of the system of interest cannot be defined and the kinetic temperature of the system of interest may not be equal to the temperature of the thermal reservoir. Nonetheless (21) can still be used to compute changes in the free energy of the system of interest as specified by (22) because the reservoir is being changed approximately quasistatically.
From the above, one observes that the function appearing in the quasistatic thermodynamic path integral (22) is the same as that which appears in the nonequilibrium free energy relation. One could conjecture that any correct microscopic expression for the thermodynamic path integral derived using classical thermodynamics would yield a correct Nonequilibrium Free Energy Relation, for some protocol. All that is required is sufficient ingenuity to design a protocol consistent with the microscopic expression for the generalized work. To be absolutely sure that your microscopic expression and protocol are consistent one should simply check that when substituted into (5) that the protocol generates the required generalized “work”. However if the Nonequilibrium Free Energy Relation is to be used beyond the synthetically thermostatted dynamics care is required. It must be ensured that the system is controlled by a thermal reservoir which remains in equilibrium.

If one constructs an algorithm (9) & (13) to accomplish some thermal transformation \((N_1, V_1, T_1) \rightarrow (N_1, V_1, T_2)\) then (5) gives a precise microscopic form for the generalized “work” appearing in the classical thermodynamic path integral for the free energy change. Although the quasistatic path integral expression is unique, the nonequilibrium expression is certainly not. This is because there are infinitely many protocols that accomplish the required change. Nonetheless each of these expressions give identical values for the free energy difference.

We thank the Australian Research Council for supporting this work.

[12] We assume that the system’s centre of mass motion is zero.
[20] It is trivial to extend this equation to the stochastic case - see ref. [19].
[21] It may be tempting to try an arbitrary expression for ∆W in the JE. In general this would be erroneous. ∆W, and more generally ∆X, are well defined, depending on the ensemble & dynamics.