



A simple proof of the Jarzynski equality?

Frédéric Douarche

► **To cite this version:**

| Frédéric Douarche. A simple proof of the Jarzynski equality?. 2005. <hal-00015478>

HAL Id: hal-00015478

<https://hal.archives-ouvertes.fr/hal-00015478>

Submitted on 8 Dec 2005

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

A simple proof of the Jarzynski equality?

F. Douarche

*Laboratoire de Physique de l'ENS Lyon — CNRS UMR 5672
46, Allée d'Italie — 69364 Lyon Cedex 07, France*

8th December 2005

1 Introduction

In this short communication, I give a very simple derivation of the Jarzynski equality [1]

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

which allows to compute the free energy difference $\Delta F = F_B - F_A$ of a body, which is driven between two equilibrium states A and B by an external (time-dependent) force λ , from the probability distribution function $P(W)$ of the work done on the system

$$W = \int_{t_A}^{t_B} \frac{\partial H_t(x, \lambda)}{\partial t} dt = \int_{t_A}^{t_B} \frac{\partial H_t(x, \lambda)}{\partial \lambda} \dot{\lambda} dt, \quad (2)$$

regardless of the nature of the transformation (reversible or irreversible) between the states A and B . Here $H_t(x, \lambda)$ denotes the full (time-dependent) Hamiltonian of the system, which is supposed to be in thermal contact with a heat bath of temperature $T = 1/k_B\beta$, the whole ensemble being isolated, $x = (p, q) = (p_1, \dots, p_n, q^1, \dots, q^n)$ denotes the phase-space point of the system (which is supposed to have $2n$ degrees of freedom), and $\langle f(W) \rangle = \int f(W) P(W) dW$ denotes an average over the ensemble of realizations of the work W done on the system (for any “smooth” function f).

There are two parts in this derivation. The first is mathematical and will be discussed later. The result is the following: let $E(x)$ be the (time-independent) energy of a body in thermal equilibrium with a heat bath of temperature T . If its energy can be written in the additive manner $E(x) = E_0(x) + V(x)$, which can always be achieved by making certain very general assumptions, then one can compute the free energy F of the body (and all the relevant thermodynamic quantities) from the only knowledge of the equilibrium Gibbsian distribution $\rho_0(x) = Z_0^{-1} e^{-\beta E_0(x)}$, where $Z_0 = \int e^{-\beta E_0(x)} dx$ is the partition function corresponding to the “unperturbed” energy $E_0(x)$ of the body (and to its “unperturbed” free energy $F_0 = -\beta^{-1} \ln Z_0$).

The second part of the argument, which relates the perturbation energy V to the thermodynamical work W done on the system, first in the time-independent case, and then in the time-dependent case, is physical.

2 Thermodynamic perturbation theory

The fundamental need for perturbation theory runs as follows: for example, in quantum mechanics, the goal is to compute the eigenvalues λ (the energy spectrum) and the eigenvectors u (the wave functions) of the (time-dependent or not) Hamiltonian $H(x)$ of a system, from the eigenvalue equation $H(u) = \lambda u$, where H is generally a very complicated differential operator, occasionally depending on time via some time-dependent energy coupling (due, e.g., to some external time-dependent force).

However, for evident analytical reasons, the calculus of the eigenvalues λ and of the eigenvectors u is generally intractable. In order to circumvent this problem, one generally write (when this is possible) the energy of the system in the additive manner $H(x) = H_0(x) + V(x)$, where $H_0(x)$ denotes the “unperturbed” energy of the system (i.e. the quantity from which one can compute the eigenvalues λ_0 and the eigenvectors u_0), and $V(x)$ the perturbation energy. Then, one computes by iteration λ and u from the only knowledge of the “unperturbed” eigenvalues λ_0 and eigenfunctions u_0 , which are computable [2]. Eventually, the perturbation energy $V(x)$ can depend on some external force λ , time-dependent or not (e.g., λ can be an electrical or magnetic field).

The situation is very similar in statistical mechanics: once ones has set up the Hamiltonian $E(x)$ of the system (which, again, is in thermal contact with a heat bath of temperature T , the whole ensemble being isolated), the fundamental problem is to compute the partition function Z of the system

$$Z = \int e^{-\beta E(x)} dx,$$

where $dx = dp dq = \prod dp_i dq^i$, from which one can compute the free energy F of the system

$$F = -\beta^{-1} \ln Z,$$

and, e.g., the thermal average energy of the system $\langle E \rangle = -\frac{\partial \ln Z}{\partial \beta}$ [3].

Again, these well-known formulas are purely formal, and inapplicable in realistic cases, because the energy $E(x)$ of the body is generally a very complicated function of the coordinates q and momenta p of the “particles” constituting the body, and the integration over the phase-space cannot be carried out in order to compute Z [in the quantum case, this calculus is of combinatoric nature, and requires a counting of the (occasionally degenerate) states $|n\rangle$ of the system via the formula $Z = \sum e^{-\beta E_n}$, which also becomes quickly an impossible task]. However, as we will show below, much more can be said about Z , F and $\langle E \rangle$, at least formally.

Let us first assume that the energy of the body (which, again, is in thermal contact with a heat bath of temperature T , the whole ensemble being isolated) is time-independent, and that it can be written in the additive manner

$$E(x) = E_0(x) + V(x), \quad (3)$$

where “everything” can be calculated from the “unperturbed” energy $E_0(x)$ of the body [the partition function $Z_0 = \int e^{-\beta E_0(x)} dx$, the free energy $F_0 = -\beta^{-1} \ln Z_0$, etc.], but not from its “full” energy $E(x)$. Therefore, the fundamental problem is to devise a method to compute, even approximatively, the partition function Z , the free energy F , etc., corresponding to the “full” energy $E(x)$ of the system. Indeed, we will show that if Eq. (3) holds, then one can compute exactly Z and F , without anymore assumptions than Eq. (3), which relies on the separability of the Hamiltonian of the system, and on thermal equilibrium.

From these very few assumptions, the algebra is very simple. Let us first compute the partition function $Z = \int e^{-\beta E(x)} dx$ of the system. From Eq. (3), one gets

$$Z = Z_0 \int \rho_0(x) e^{-\beta V(x)} dx = Z_0 \langle e^{-\beta V} \rangle_0, \quad (4)$$

where $\rho_0(x) = Z_0^{-1} e^{-\beta E_0(x)}$ is the “unperturbed” equilibrium Gibbsian distribution of the system. Therefore, the free energy of the system is given by

$$F = -\beta^{-1} \ln Z = F_0 - \beta^{-1} \ln \langle e^{-\beta V} \rangle_0. \quad (5)$$

This result was derived up to the second order [i.e. only with the first two cumulants of V in Eq. (5)] by Peierls and Bogoliubov, who gave an upper bound for the free energy F of the system [3] [indeed, Eq. (5) indicates that the “full” free energy F is merely the cumulants generating function of the perturbation energy V]. Our result is much stronger, and looks amazingly similar to the Jarzynski equality (1).

3 Dynamical interpretation

The preceding result holds for a time-independent and equilibrium system. However, provided that the system is initially (say, at time $t = 0$) at equilibrium, these results are also true if the energy of the system depend on time through a time-dependent perturbation energy, which can be due to an external force λ which is switched on at times $t > 0$. Therefore, if the energy of the system reads

$$E_t(x, \lambda) = E_0(x) + V_t(x, \lambda), \quad (6)$$

where the time-dependent perturbation energy V_t is only non-zero for times $t > 0$, then one trivially has

$$F_t = F_0 - \beta^{-1} \ln \langle e^{-\beta V_t} \rangle_0 \quad \text{for all } t > 0, \quad (7)$$

which has exactly the same meaning as Eqs. (4)-(5).

In fact, the perturbation energy V_t can be identified to the work W_t done on the system on the time interval $(0, t)$ in a very simple way [one simply has to put $t_A = 0$ and $t_B = t$ in Eq. (2), but here B does not necessarily denotes an equilibrium state]. Since the energy E_t of the system depends explicitly on time only through the external force $\lambda(t)$, one has by applying the chain rule

$$\frac{dE_t(x, \lambda)}{dt} = \frac{\partial E_t(x, \lambda)}{\partial t} = \frac{\partial E(x, \lambda)}{\partial \lambda} \frac{d\lambda}{dt}.$$

Therefore, by Eq. (2) one has $W_t = E_t - E_0 = V_t$, and one finally gets

$$F_t = F_0 - \beta^{-1} \ln \langle e^{-\beta W_t} \rangle_0, \quad (8)$$

which resembles amazingly the Jarzynski equality (1).

4 Discussion and conclusion

The formula (8) obtained above is not really the Jarzynski equality (1), since the Jarzynski equality computes the free energy difference of the system from the work distribution $P(W)$, which takes into account all the (in principle) nonequilibrium fluctuations of the work W done on the system during the transformation between the (equilibrium) states A and B . In our derivation, this is obviously not the case, since the average $\langle \cdot \rangle$ is performed from the equilibrium Gibbsian distribution $\rho_0(x) = Z_0^{-1} e^{-\beta E_0(x)}$.

However, this slight detail perhaps throws some light on the debate started by Cohen and coworkers [4], and perhaps explains why all the experiments which has been performed in order to test the Jarzynski equality are successful, even in very defavorable (irreversible) cases (and more strikingly, when the state B is not an equilibrium one) [5].

To conclude, we should add that among the hypotheses needed to the result (8), the ergodicity of the system is of course required. However, even unrealistic, this is a rather weak statement, because we only need ergodicity at times $t \leq 0$, that is to say without external (time-dependent) driving force λ .

Acknowledgments — The author thank S. Ciliberto and K. Gawędzki for useful discussions.

References

- [1] C. Jarzynski, Phys. Rev. Lett. **78** (14), 2690 (1997); see also G. E. Crooks, J. Stat. Phys. **90** (5-6), 1481 (1997)
- [2] L. D. Landau, E. M. Lifchitz, *Mécanique Quantique*, 3rd ed. (Editions Mir Moscou, 1975); see also E. Merzbacher, *Quantum Mechanics*, 3rd ed. (John Wiley & Sons, 1998), and A. Messiah, *Mécanique Quantique*, 2 vols. (Dunod, 1995, republication of the 1959 ed.)

- [3] L. D. Landau, E. M. Lifshitz, *Statistical Physics* (Pergamon Press, 1970); see also R. Peierls, Phys. Rev. **54**, 918 (1938), H. B. Callen, *Thermodynamics and an Introduction to Thermostatistics*, 2nd ed. (John Wiley & Sons, 1985), and R. P. Feynman, *Statistical Mechanics: a Set of Lectures* (Addison-Wesley, 1972)
- [4] E. G. D. Cohen, D. Mauzerall, J. Stat. Mech.: Theor. Exp. P07006 (2004)
- [5] F. Douarche, S. Ciliberto, A. Petrosyan, and I. Rabbiosi, Europhys. Lett. **70** (5), 593 (2005); F. Douarche, S. Ciliberto, and A. Petrosyan, J. Stat. Mech.: Theor. Exp. P09011 (2005)