Thermodynamic derivation of the fluctuation theorem and Jarzynski equality

Article

Published Version

International Scholarly Research Network
ISRN Thermodynamics
Volume 2012, Article ID 528737, 5 pages
doi:10.5402/2012/528737

Research Article

Thermodynamic Derivation of the Fluctuation Theorem and Jarzynski Equality

Maarten H. P. Ambaum

Department of Meteorology, University of Reading, Reading RG6 6BB, UK

Correspondence should be addressed to Maarten H. P. Ambaum, m.h.p.ambaum@reading.ac.uk

Received 14 February 2012; Accepted 15 March 2012

Academic Editors: M. Appell, C. Pierleoni, and Z. Slanina

A thermodynamic expression for the analog of the canonical ensemble for nonequilibrium systems is described based on a purely information theoretical interpretation of entropy. It is shown that this nonequilibrium canonical distribution implies some important results from nonequilibrium thermodynamics, specifically, the fluctuation theorem and the Jarzynski equality. Those results are therefore expected to be more widely applicable, for example, to macroscopic systems.

1. Introduction

The derivations of the fluctuation theorem [1, 2] and the Jarzynski equality [3] appear to depend on the underlying microscopic Hamiltonian dynamics. From this it would follow that these theorems are only relevant to microscopic systems, with their associated definitions of entropy and temperature. In contrast, a statistical mechanical description of macroscopic systems often depends on more general forms of entropy, primarily information entropy [4–6]. Two notable examples from fluid dynamics are the statistical mechanics of point vortices [7] and the statistical mechanics of two-dimensional incompressible flows [8]. In such cases, temperature is defined in terms of the change of entropy with the energy of the system [9] or, equivalently, in terms of the Lagrange multiplier for the energy under the maximization of entropy at a given expectation value of the energy [10].

The question is whether for such macroscopic systems we can derive a fluctuation theorem or Jarzynski equality. This is of particular importance for climate science as there are strong indications that the global state of the climate system and, more generally, other components of the Earth system may be governed by thermodynamic constraints on entropy production [11–15]. The theoretical underpinning of those thermodynamic constraints is still lacking. The presence of a fluctuation theorem for such systems would be of great importance.

Here we demonstrate that the information-theoretical definition of entropy implies the fluctuation theorem and the Jarzynski equality. It is shown that these results are due to the counting properties of entropy rather than the dynamics of the underlying system. As such, both these results are applicable to a much wider class of problems, specifically, macroscopic systems for which we can define an entropy and which are thermostated in some general sense.

The central tenet is that for two states A and B of a system, defined by two sets of macroscopic parameters, the ratio of the probabilities p_B/p_A for the system to be in either state is

$$\frac{p_B}{p_A} = \exp\left(\frac{\Delta_{AB}S}{k}\right), \tag{1}$$

with $\Delta_{AB}S$ being the difference in entropy between the states B and A. This is essentially the Boltzmann definition of entropy: entropy is a counting property of the system. The theoretical background can be found in [10], where it is shown that this information theoretical interpretation reproduces the statistical mechanics based on Gibbs entropy but furthermore gives a justification of the Gibbs formulation as a statistical inference problem under limited knowledge of the system. Of note is that the entropy only has meaning in relation to the macroscopic constraints on the system (indicated by the subscripts A and B), constraints which can be arbitrarily complex and prescriptive, as may be

needed for systems far from equilibrium. In an information-theoretical setting the previous definition of entropy is equivalent to the *principle of indifference*: the absence of any distinguishing information between microscopic states *within* any of the macroscopic states A or B is equivalent to equal prior (prior to obtaining additional macroscopic constraints) probabilities for the microscopic states [16]. Note also that we do not need to specify precisely at this point how the states are counted, or how an invariant measure can be defined on the phase space confined by A or B. The principle of indifference does not imply that all states are assumed equally probable; it is a statement that we cannot *a priori* assume a certain structure in phase space (such as a precisely defined invariant measure) in the absence of further information. The principle of indifference is not a statement about the structure of phase space; it is a principle of statistical inference and it is the only admissible starting point from an information theoretical point of view.

2. A General Form for the Canonical Ensemble

Following Boltzmann, we define the entropy S_A as the logarithm of the number of states accessible to a system under given macroscopic constraints A. For an isolated system, the entropy is related to the size Φ_A of the accessible phase space:

$$S_A = k \ln \Phi_A. \tag{2}$$

For a classical gas system, A is defined by the energy U, volume V and molecule number N, the phase space size Φ_A is the hyperarea of the energy shell, and it defines the usual microcanonical ensemble. For more complicated systems, where A may include several macroscopic order parameters, the energy shell becomes more confined; in the following we will still refer to the accessible phase space under constraints A as the energy shell. The hyperarea Φ_A is nondimensionalised such that $\Phi_A(U)\mathrm{d}U$ is proportional to the number of states between energies U and $U + \mathrm{d}U$. We will not consider other multiplicative factors which make the argument of the logarithm nondimensional; these contribute an additive entropy constant which will not be of interest to us here. Note also that the microcanonical ensemble does not include a notion of equilibrium: the system is assumed to be insulated, so it cannot equilibrate with an external system. It just moves around on the energy shell (defined by A) and the principle of indifference implies that all states, however improbable from a macroscopic point of view, are members of the ensemble. Of course, the number of unusual states (say, with nonuniform macroscopic properties not defined by A) is much lower than the number of regular states (say, with uniform macroscopic density) for macroscopic systems. Only for small systems, the distinction becomes important but it does not invalidate the previous formal definition of entropy. The previous definition of entropy also ensures that entropy is an extensive property such that for two independent systems considered together the total entropy is the sum of the individual entropies, $S = S_1 + S_2$. The Boltzmann constant k ensures dimensional compatibility with the classical thermodynamic entropy when the usual equilibrium assumptions are made [10, 17].

The hyperarea of the energy shell, and thus the entropy, can be a function of several variables which are set as external constraints, such as the total energy U, system volume, V, or particle number N for a simple gas system. For the canonical ensemble we consider a system that can exchange energy with some reservoir. We consider here only a theoretical canonical ensemble in that we consider the coupling between the two systems to be weak such that the interaction energy vanishes compared to the relevant energy fluctuations in the system.

First, we need to define what a reservoir is. Following equilibrium thermodynamics, we formally define an inverse temperature $\beta = (kT)^{-1}$ as

$$\beta = \frac{1}{k}\frac{\partial S}{\partial U} = \frac{1}{\Phi}\frac{\partial \Phi}{\partial U}. \tag{3}$$

We make no claim about the equality of β and the classical equilibrium inverse temperature; β is the expansivity of phase space with energy and as such can be defined for any system, whether it is in thermodynamic equilibrium or not. When an isolated system is prepared far from equilibrium (e.g., when it has a local equilibrium temperature which varies over the system), then β is still uniquely defined for the system as a nonlocal property of the energy shell that the system resides on. Because both energy and entropy in the weak coupling limit are extensive quantities, β must be an intensive quantity.

Now consider a large isolated system R with total (internal) energy U_R. Let this system receive energy U' from the environment. By expanding its entropy S_R in powers of U, we can then write the entropy of this large system as

$$S_R(U_R + U') = S_R(U_R) + kU'\left(\beta + \frac{1}{2}U'\frac{\partial \beta}{\partial U} + \mathcal{O}\left(U'^2\right)\right). \tag{4}$$

We see that for finite U', $(\partial\beta/\partial U)^{-1}$ has to be an extensive quantity. But that means that for a very large system $\partial\beta/\partial U = \mathcal{O}(N^{-1})$, where N is a measure of the size of the system (such as particle number). For a classical thermodynamic system $\partial\beta/\partial U = -k\beta^2/C_V$ with C_V the heat capacity at constant volume. We conclude that for a very large system ($N \to \infty$), the entropy equals

$$S_R(U_R + U') = S_R(U_R) + k\beta U' \tag{5}$$

for all relevant, finite energy exchanges U'. This expression for the entropy defines a reservoir. The size of the energy shell accessible to the reservoir is, for all relevant energy exchanges U', exactly proportional to $\exp(\beta U')$, with β an intensive and constant property of the reservoir. We do not require the reservoir to be in thermodynamic equilibrium. A change of energy in the reservoir pushes the reservoir to a different energy shell A'; the functional dependence of the size of the energy shell with energy defines the inverse temperature β, as in (3). However, it is not assured that a small and fast thermometer would measure an inverse temperature equal to β at some point in the reservoir; only if the reservoir is

allowed to equilibrate, its inverse temperature is everywhere equal to β. Of course, this is how the temperature of a classical reservoir is determined in practice.

Now suppose that a system of interest has energy U_0. We then allow it to exchange heat U with a reservoir. If the system has energy U_0+U, the reservoir must have given up energy U. We can write the hyperarea of the energy shell of the system Φ_0 as a function of U. The total entropy of the system plus reservoir R can then be written as a function of the exchange energy, U, as

$$S = S_0(U) + S_R(U_R) - k\beta U, \tag{6}$$

with $S_0 = k \ln \Phi_0$. The number of states at each level of exchange energy therefore is proportional to

$$\Phi(U) \propto \Phi_0(U)\ \exp(-\beta U), \tag{7}$$

where we omitted proportionality constants related to the additive entropy constants. Nowhere we assume that the system is in equilibrium with the reservoir. This means that $\Phi(U)$ is the relevant measure to construct an ensemble average for the system, even for far-from-equilibrium systems. Even the reservoir can be locally out of equilibrium, as discussed previously. We have also made no reference to the size of the system of interest, as long as it is much smaller than the reservoir. However, in contrast to systems in thermodynamic equilibrium, there is no guarantee that the extensive macroscopic variables, such as U, V, or N, define the state of the system in any reproducible sense. To fully define an out-of-equilibrium system we need to introduce order parameters that can describe the nonequilibrium aspects of the system.

The previous density is an integrated version of the usual canonical distribution. The size of the energy shell of the system of interest, Φ_0, can be written as an integral over states Γ such that

$$\Phi_0(U) = \int_{H_0(\Gamma)=U} \mathrm{d}\Gamma, \tag{8}$$

with H_0 being the Hamiltonian of the system of interest. With this definition, the density in (7) reduces to the usual canonical distribution $\exp(-\beta H_0(\Gamma))$ for states Γ. We will not make further use of this microscopic version of the density.

3. Fluctuation Theorems

The canonical density in (7) can be expanded by parametrizing each energy shell with some continuous coordinate v so that every part of phase space has coordinates (U, v). The coordinate v is again a macroscopic coordinate so that any combination (U, v) can correspond to many microscopic states. At each value of v the differential $\phi(U, v)\mathrm{d}U\,\mathrm{d}v$ is proportional to the number of states between coordinate values U and $U + \mathrm{d}U$, and v and $v + \mathrm{d}v$, and it is normalised such that

$$\int \phi(U, v)\ \mathrm{d}v = \Phi_0(U). \tag{9}$$

The parametrisation is arbitrary at this point and can be chosen such as to divide the phase space in as fine a structure as desired for a given application. We can define an entropy $S_0(U, v)$ again as the logarithm of the number of available states for the system of interest corresponding to the subset of phase space defined by (U, v):

$$S_0(U, v) = k \ln \phi(U, v). \tag{10}$$

Now consider a process that occurs on the energy shell U where some variable changes from $A \rightarrow B$. On the parametrized energy shell this corresponds to a coordinate shift from $v(A) \rightarrow v(B)$. The number of corresponding states changes from $\phi(U, v(A)) \rightarrow \phi(U, v(B))$. We can use detailed balance to express the ratio of the probability of making this transition to the probability of making the reverse transition as the ratio of the number of states at $(U, v(A))$ to the number of states at $(U, v(B))$:

$$\frac{p_{A\to B}}{p_{B\to A}} = \frac{\phi(U, v(B))}{\phi(U, v(A))} = \exp\left(\frac{\Delta_{AB}S}{k}\right), \tag{11}$$

where $\Delta_{AB}S/k = S_0(U, v(B)) - S_0(U, v(A))$. If, in addition, during the process $A \rightarrow B$ the energy of the system of interest changes from $U_A \rightarrow U_B$ through exchange with the reservoir, then the previous ratio of probabilities can still be expressed as $\exp(\Delta_{AB}S/k)$ but now with

$$\Delta_{AB}S = S_0(U_B, v(B)) - S_0(U_A, v(A)) - k\beta(U_B - U_A). \tag{12}$$

We can always write the entropy change of the system of interest as the sum of the entropy change due to heat exchange with the reservoir and an irreversible entropy change associated with uncompensated heat [14, 18], namely, $S_0(U_B, v(B)) - S_0(U_A, v(A)) = k\beta(U_B - U_A) + \Delta_i S_0$. We thus conclude that $\Delta_{AB}S = \Delta_i S_0$; that is, the relevant entropy change in (11) equals the irreversible entropy change of the system of interest. So for processes that occur either on or across energy shells, we have

$$\frac{p_{A\to B}}{p_{B\to A}} = \exp\left(\frac{\Delta_i S_0}{k}\right), \tag{13}$$

with $\Delta_i S_0$ being the irreversible entropy change of the system in a process $A \rightarrow B$. The right-hand side of this equation is only dependent on the irreversible entropy change $\Delta_i S_0$ between the two states of the system of interest. So this equation must be true for any pair of states (A, B) that are related by the same irreversible entropy change. We thus arrive at the *fluctuation theorem* [1, 2]:

$$\frac{p(\Delta_i S)}{p(-\Delta_i S)} = \exp\left(\frac{\Delta_i S}{k}\right), \tag{14}$$

with $p(\Delta_i S)$ being the probability that the system of interest makes a transition with irreversible entropy change of $\Delta_i S$ and $p(-\Delta_i S)$ being the probability for the opposite change.

The fluctuation theorem applies to spontaneous processes that occur in thermostated but otherwise isolated systems. We next consider processes that occur when we modify the system of interest by changing some external

macroscopic parameters. The entropy of the energy shell U is then also a function of some parameter λ, namely, $S = S_\lambda(U, v)$. Without loss of generality we set $\lambda = 0$ at A and $\lambda = 1$ at B. In this case the irreversible entropy change in (13) is

$$\frac{\Delta_i S}{k} = S_1(U_B, v(B)) - S_0(U_A, v(A)) - k\beta(U_B - U_A). \quad (15)$$

Apart from this, there is no change in the considerations leading to the fluctuation theorem. By definition, thermostated systems that receive work W_{AB} from their environment have an irreversible entropy change equal to

$$\frac{\Delta_i S}{k} = \beta(W_{AB} - \Delta_{AB}F), \quad (16)$$

with $\Delta_{AB}F$ being the change in free energy going from A to B. Recognising that the right-hand side is again only a function of the difference between the two states, we arrive at the *Crooks fluctuation theorem* [19]:

$$\frac{p_{01}(W)}{p_{10}(-W)} = \exp(\beta(W - \Delta_{01}F)), \quad (17)$$

with $p_{01}(W)$ being the probability that the system absorbs work W when λ changes from 0 to 1, and $p_{10}(-W)$ being the probability that the system performs work W when λ changes in reverse from 1 to 0. Because the transition probabilities can be normalised with respect to the exchanged work, it is straightforward to use this equation to show that the expectation value of $\exp(-\beta(W - \Delta_{01}F))$ equals unity, or equivalently,

$$\langle \exp(-\beta W) \rangle = \exp(-\beta \Delta_{01}F). \quad (18)$$

This is the Jarzynski equality [3].

The consistency of the previous argument is strengthened by the following independent route to calculate free energy changes. The phase space measure $\Phi(U)$ can be normalised with the partition function Z_λ:

$$Z_\lambda = \int \Phi_\lambda(U)\ \exp(-\beta U)\mathrm{d}U, \quad (19)$$

where $\Phi_\lambda(U)$ is proportional to the number of accessible states of the isolated the system of interest when the external parameter is set to λ. The equilibrium free energy for the thermostated system is

$$F_\lambda = -\beta^{-1} \ln Z_\lambda. \quad (20)$$

Next we consider what happens to the equilibrium free energy of the system when we vary λ from 0 to 1. The partition function at $\lambda = 1$ satisfies

$$\begin{aligned} Z_1 &= \int \Phi_1(U)\ \exp(-\beta U)\mathrm{d}U \\ &= \int \Phi_0(U)\ \exp\left(\frac{\Delta S}{k}\right)\ \exp(-\beta U)\mathrm{d}U \\ &= Z_0\ \left\langle \exp\left(\frac{\Delta S}{k}\right)\right\rangle, \end{aligned} \quad (21)$$

where $\langle . \rangle$ denotes an ensemble average over the initial ensemble, and $\Delta S = k \ln(\Phi_1(U)/\Phi_0(U))$. As before, the entropy change can be written as the sum of the entropy change due to heat exchange with the reservoir and the irreversible entropy change due to uncompensated heat. Because the system plus the reservoir is thermally insulated, any heat given to the reservoir must be compensated by work performed by the external parameter change. The entropy change can therefore be written as $\langle \exp(\Delta S/k) \rangle = \langle \exp(\Delta_i S/k - \beta W) \rangle$ so that we find

$$\frac{Z_1}{Z_0} = \left\langle \exp\left(\frac{\Delta_i S}{k} - \beta W\right)\right\rangle. \quad (22)$$

Because (16) is true for any microscopic realisation of the process, we find that the right-hand side of the previous equation is the same for every realisation and it is equal to $\exp(-\beta \Delta F)$. This is consistent with the equilibrium expression for the free energy, (20), from which follows that $\exp(-\beta \Delta F) = Z_1/Z_0$. The previous equation is only apparently in contradiction to the Jarzynski equality, (18). To arrive at the Jarzynski equality we recognise that (16) implies that $\langle \exp(\beta(\Delta F - W)) \rangle = \langle \exp(-\Delta_i S/k) \rangle = 1$, where the last equality follows from integrating the fluctuation theorem over all values of $\Delta_i S$.

4. Discussion

We have shown that the fluctuation theorem (14) and Jarzynski equality (18) follow from general counting properties of entropy and not from the underlying dynamics. As such we expect both results to be widely applicable to systems that are in some sense thermostated, that is, systems that are able to settle on a given expectation value for the total energy by interaction with a reservoir.

The climate system is potentially a nontrivial example of such a system: the incoming short-wave radiation from the Sun is depleted by long-wave (thermal infrared) radiation from the Earth to space. The corresponding equilibrium temperature is the bolometric temperature of the planet (about 255 K in case of the Earth [14]). (The bolometric radiation temperature of the Earth is substantially lower than the observed average surface temperature of about 288 K, because of the greenhouse effect of the atmosphere.) It is not obvious how to apply the fluctuation theorems to the climate system and how the entropy production in the climate system is related to the actual climate on Earth. For example, most of the entropy production in the climate system is due to degradation of radiation (e.g., [20]); namely, short wavelength visible sunlight is thermalized by molecular absorption into molecular thermal energy corresponding to long wavelength infrared radiation. This degradation of radiative energy is the main source of entropy production in the climate system, but as this entropy production only resides in the photon field, its relation to, for example, kinetic energy dissipation in the atmosphere is not clear. So from this example it appears that we need to select the relevant forms of entropy production before we can use it to make inferences about the climate system.

It remains to be seen whether the fluctuation theorems can be usefully applied to complex systems such as the climate, but we believe that the derivation presented here can pave the way for attempts in that direction.

References

[1] D. J. Evans, E. G. D. Cohen, and G. P. Morriss, "Probability of second law violations in shearing steady states," *Physical Review Letters*, vol. 71, no. 15, pp. 2401–2404, 1993.

[2] D. J. Evans and D. J. Searles, "The fluctuation theorem," *Advances in Physics*, vol. 51, no. 7, pp. 1529–1585, 2002.

[3] C. Jarzynski, "Nonequilibrium equality for free energy differences," *Physical Review Letters*, vol. 78, no. 14, pp. 2690–2693, 1997.

[4] C. E. Shannon and W. Weaver, *A Mathematical Theory of Communication*, The University of Illinois Press, Champaign, Ill, USA, 1963.

[5] R. T. Cox, *The Algebra of Probable Inference*, The Johns Hopkins University Press, Johns Hopkins, NJ, USA, 1961.

[6] E. T. Jaynes, *Probability Theory: The Logic of Science*, Cambridge University Press, 2003.

[7] L. Onsager, "Statistical hydrodynamics," *Nuovo Cimento, Supplemento*, vol. 6, pp. 279–287, 1949.

[8] R. Robert and J. Sommeria, "Statistical equilibrium states for two-dimensional flows," *Journal of Fluid Mechanics*, vol. 229, pp. 291–310, 1991.

[9] R. Baierlein, *Thermal Physics*, Cambridge University Press, 1999.

[10] E. T. Jaynes, "Information theory and statistical mechanics," *The Physical Review*, vol. 106, no. 4, pp. 620–630, 1957.

[11] G. W. Paltridge, "Global dynamics and climate—a system of minimum entropy exchange," *Quarterly Journal of the Royal Meteorological Society*, vol. 101, no. 429, pp. 475–484, 1975.

[12] H. Ozawa, A. Ohmura, R. D. Lorenz, and T. Pujol, "The second law of thermodynamics and the global climate system: a review of the maximum entropy production principle," *Reviews of Geophysics*, vol. 41, no. 4, pp. 1–24, 2003.

[13] A. Kleidon and R. D. Lorenz, *Non-Equilibrium Thermodynamics and the Production of Entropy: Life, Earth, and Beyond*, Understanding Complex Systems, Springer, Berlin, Germany, 2005.

[14] M. H. P. Ambaum, *Thermal Physics of the Atmosphere*, Wiley-Blackwell, Chichester, UK, 2010.

[15] S. Pascale, J. M. Gregory, M. H. P. Ambaum, R. Tailleux, and V. Lucarini, "Vertical and horizontal processes in the global atmosphere and the maximum entropy production conjecture," *Earth System Dynamics*, vol. 3, no. 1, pp. 19–32, 2012.

[16] E. T. Jaynes, "Prior probabilities," *IEEE Transactions on Systems Science and Cybernetics*, vol. 4, no. 3, pp. 227–241, 1968.

[17] E. T. Jaynes, "Gibbs vs boltzmann entropies," *American Journal of Physics*, vol. 5, pp. 391–398, 1965.

[18] D. Kondepudi and I. Prigogine, *Modern Thermodynamics*, John Wiley & Sons, Chichester, UK, 1998.

[19] G. E. Crooks, "Entropy production fluctuation theorem and the nonequilibrium work relation for free energy differences," *Physical Review E*, vol. 60, no. 3, pp. 2721–2726, 1999.

[20] S. Pascale, J. Gregory, M. Ambaum, and R. Tailleux, "Climate entropy budget of the hadcm3 atmosphere-ocean general circulation model and of famous, its low-resolution version," *Climate Dynamics*, vol. 36, pp. 1189–1206, 2011.