The "footprints" of irreversibility

A. Gomez-Marin[1], J.M.R. Parrondo[2] and C. Van den Broeck[3]

[1] *Facultat de Física, Universitat de Barcelona, Diagonal 647, Barcelona, Spain*
[2] *Departamento de Física Atómica, Molecular y Nuclear and* GISC, *Universidad Complutense, 28040 Madrid, Spain*
[3] *Hasselt University, B-3590 Diepenbeek, Belgium*

PACS 05.70.Ln – Nonequilibrium and irreversible thermodynamics
PACS 05.20.-y – Classical statistical mechanics
PACS 05.40.-a – Fluctuation phenomena, random processes, noise, and Brownian motion

Abstract. - We reformulate the result for the entropy production given in Phys. Rev. Lett. **98**, 080602 (2007) in terms of the relative entropy of microscopic trajectories. By a combination with the Crook's theorem, we identify the path variables that are sufficient to fully identify irreversibility. We show that work saturates the relative entropy, and derive the entropy production for stochastic descriptions.

Recent results, known as fluctuation [1–5] or work [6–15] theorems, point to the existence of exact equalities that rule the fluctuating amounts of work or entropy produced during far from equilibrium processes. For example, the Jarzynski equality states that $\langle \exp(-\beta W)\rangle = \exp(-\beta\Delta F)$, where W is the work needed to bring a system, in contact with a heat bath at temperature T ($\beta^{-1} \equiv k_B T$), from one initial state prepared in equilibrium to another one. ΔF is the difference in free energy of these states (see [10] for a more precise discussion). By the application of Jensen's inequality, one finds $\langle W\rangle \geq \Delta F$. Since $(\langle W\rangle - \Delta F)/T$ is the entropy increase in the entire construction, system plus heat bath, this result is in agreement with the second law. While such a result is certainly intriguing and of specific interest for the study of small systems, where the distribution of work is relevant and measurable, it provides no extra information on the actual value of the average work or entropy increase, which is the central quantity in the second law. Recently however, the microscopically exact value of these quantities has been obtained in a set-up similar to that of the work theorem [16]. The purpose of this letter is to investigate some consequences of this result, with special emphasis on the case when the dynamics of the system can be described in terms of a reduced set of variables. To make this connection, we will rewrite the main result from [16] in an alternative form, as an integral over paths. In combination with a microscopic version of Crooks' theorem, this result identifies the "footprints" of irreversibility, namely the path variables whose statistics are sufficient to reproduce the exact total entropy production. This prescription is in agreement with the expressions for entropy production proposed in the literature for stochastic models.

We consider a system described by the Hamiltonian $H(\Gamma, \lambda)$ with $\Gamma = (\{q\}, \{p\})$ a point in phase space, representing all position and momentum variables. λ is an external control parameter, for example the volume or an external field. The system is perturbed away from its initial canonical equilibrium at temperature T by changing this control parameter according to a specific schedule, from an initial to a final value. For simplicity, we will assume that during this time the system is disconnected from the outside world, except for the action of changing λ. This assumption makes the derivation and discussion simpler, even though the result can be shown to have a much wider range of validity [17]. We also consider the time-reversed schedule, in which the system starts in canonical equilibrium at the same temperature T, but at the final value of the control parameter, and the time-reversed perturbation in λ is applied. We will use the superscript "tilde" to refer to corresponding time-reversed quantities (including, by convention, the change of sign for momentum variables).

The quantity of interest is the amount of work W performed during the forward process. Since the system is isolated, W is equal to the energy difference of the system between final and initial state. While the final state is the deterministic outcome, prescribed by Hamiltonian dynamics, of the initial condition, the latter is a random variable in view of the canonical sampling. Therefore W is also a random variable. In the following, it will be useful to regard the work W as a functional of the specific microscopic trajectory followed by the system. As mentioned above, such a trajectory is completely specified by the initial condition, but also by the micro-state Γ of the system at any intermediate time t. In [16], the following explicit expression was derived for the corresponding work $W(\Gamma;t)$:

$$W(\Gamma;t) - \Delta F = k_B T \ \ln \frac{\rho(\Gamma;t)}{\tilde{\rho}(\tilde{\Gamma};t)}. \tag{1}$$

Here $\rho(\Gamma;t)$ and $\tilde{\rho}(\tilde{\Gamma};t)$ are the phase space densities at the same (forward) time t in forward and backward experiment. If the system is reconnected after the perturbation to a (ideal) heat bath at temperature T, the dissipated work $W_{\rm dis} = W(\Gamma;t) - \Delta F$ will be evacuated to the heat bath, resulting in a total entropy production equal to $W_{\rm dis}/T$. The above formula is thus the microscopic analogue of the path-dependent entropy production proposed in various stochastic models [5,18–26]. Our emphasis here however is on the average dissipated work or average entropy production. Starting from the same Eq. (1), we derive for this average two different expressions, the combination of which will lead to a general prescription identifying the "footprints" of irreversibility.

First, we derive from Eq. (1) a symmetry relation for the probability distribution $P(W)$ of the work as follows:

$$\begin{aligned} P(W) &= \langle \delta(W - W(\Gamma;t)) \rangle \\ &= \int \mathrm{d}\Gamma \rho(\Gamma;\mathrm{t}) \delta(\mathrm{W} - \mathrm{W}(\Gamma;\mathrm{t})) \\ &= \int \mathrm{d}\Gamma \mathrm{e}^{\beta(\mathrm{W}(\Gamma;\mathrm{t}) - \Delta \mathrm{F})} \tilde{\rho}(\tilde{\Gamma};\mathrm{t}) \delta(\mathrm{W} - \mathrm{W}(\Gamma;\mathrm{t})) \\ &= e^{\beta(W-\Delta F)} \int \mathrm{d}\tilde{\Gamma} \tilde{\rho}(\tilde{\Gamma};\mathrm{t}) \delta(\mathrm{W} + \tilde{\mathrm{W}}(\tilde{\Gamma};\mathrm{t})) \\ &= e^{\beta(W-\Delta F)} \ \tilde{P}(-W), \end{aligned} \tag{2}$$

since the work in the backward processes verifies $\tilde{W}(\tilde{\Gamma};t) = -W(\Gamma;t)$. This microscopic Crooks' relation was obtained in the context of Markovian stochastic dynamics by Crooks [18], and later extended to Hamiltonian dynamics in [14]. The above result is usually viewed as an interesting relation for the probability distribution of the work. It however also provides a revealing expression for the average work. By solving Eq. (2) for W and averaging over $P(W)$, one finds:

$$\begin{aligned} \langle W \rangle - \Delta F &= k_B T \int \mathrm{dW \ P(W) \ln \frac{P(W)}{\tilde{P}(-W)}} \\ &= k_B T \ D(P(W)||\tilde{P}(-W)). \end{aligned} \tag{3}$$

Here, we introduced the relative entropy, or Kullback-Leibler distance, between two probability distributions $p(x)$ and $q(x)$ [27]:

$$D(p||q) = \int \mathrm{dx \ p(x) \ln \frac{p(x)}{q(x)}}. \tag{4}$$

The relative entropy and its powerful properties will play a central role in the sequel. At first sight, it may appear superfluous to express the average $\langle W \rangle$, which is obviously just an integral of $P(W)$, in terms of a more complicated expression involving the second probability distribution $\tilde{P}$ for the reverse experiment. But the following two important properties of the relative entropy [27] reveal an additional benefit. Firstly, a relative entropy is non-negative. Eq. (3) thus implies that the dissipated work $\langle W \rangle - \Delta F$ is a positive quantity, in agreement with the second law. Secondly, the relative entropy expresses the difficulty for distinguishing samplings from two distributions. The dissipated work is thus equal to the difficulty to distinguish the arrow of time from the statistics of the work involved in forward versus backward experiment. The main interest of Eq. (3) however comes from its comparison with an expression for $\langle W \rangle$ in terms of the micro-dynamics, which we proceed to derive below.

By performing the straightforward average in Eq. (1), we find [16]:

$$\langle W \rangle - \Delta F = k_B T \int \mathrm{d}\Gamma \rho(\Gamma;\mathrm{t}) \ln \frac{\rho(\Gamma;\mathrm{t})}{\tilde{\rho}(\tilde{\Gamma};\mathrm{t})} = \mathrm{k_B T D}(\rho||\tilde{\rho}). \tag{5}$$

In comparison with Eq. (3), the above result fully reveals the microscopic nature of the dissipation, but it may appear to be of little practical interest. Indeed, it requires *full statistical information* on *all* the microscopic degrees of freedom of the system (even though only at one particular time). This stringent requirement is obviously on par with the generality of the above result, which is valid however far the system is perturbed away from equilibrium. The perturbation could therefore imprint its effect on all the degrees of freedom and their complete statistical properties would be required to reproduce the corresponding dissipation.

While Eqs. (3) and (5) provide two different exact expressions for the dissipated work, we note that the formulas for entropy production in coarse grained descriptions are usually in terms of path integrals, on par with the fact that the determinism of Hamiltonian dynamics is then replaced by stochastic dynamics. Eq. (3) can be considered to be a path integral version since the work W will, in a reduced description, indeed depend on the path followed by the coarse grained variables during the perturbation. To derive a path integral version of Eq. (5), we invoke another property of relative entropy [27], known as the chain rule. The relative entropy between probability distributions $p(x,y)$ and $q(x,y)$ of two random variables can be written as follows:

$$\begin{aligned} D\left(p(x,y)||q(x,y)\right) &= D\left(p(x)||q(x)\right) \\ &+ \int \mathrm{dx \, p(x)} \int \mathrm{dy \, p(y|x) \ln \frac{p(y|x)}{q(y|x)}}. \end{aligned} \tag{6}$$

If the random variables are related to each other by a one-to-one function $x = f(y)$, their conditional probabilities become infinitely sharp and the second term in the r.h.s. of Eq. (6) vanishes. One then finds:

$$D\left(p(x,y)||q(x,y)\right) = D\left(p(x)||q(x)\right) = D\left(p(y)||q(y)\right). \tag{7}$$

In words, the addition of dependent variables does not modify the relative entropy. Since Hamiltonian dynamics generates such one-to-one relations between the microstates at different times, one can specify in Eq. (5), without changing the value of the relative entropy, the microstate Γ_i of the system at as many additional measurement points in times t_i, $i = 1, ..., n$, as one likes:

$$\langle W\rangle - \Delta F = k_B T \int \prod_{i=1}^{n} d\Gamma_i \; \rho(\{\Gamma_i; t_i\}) \ln \frac{\rho(\{\Gamma_i; t_i\})}{\tilde{\rho}(\{\tilde{\Gamma}_i; t_i\})}. \tag{8}$$

In the continuum limit covering the entire time interval (with $n \to \infty$), one thus converges to the following result in terms of a path integral, see also [28]:

$$\begin{aligned}\langle W\rangle - \Delta F &= k_B T \int \mathcal{D}(\text{path}) \mathcal{P}(\text{path}) \ln \frac{\mathcal{P}(\text{path})}{\widetilde{\mathcal{P}}(\widetilde{\text{path}})} \\ &= k_B T D(\mathcal{P}(\text{path})||\widetilde{\mathcal{P}}(\widetilde{\text{path}})). \end{aligned} \tag{9}$$

This expression, while containing redundant information from the point of view of Hamiltonian dynamics, has the important advantage that it is also exact and formally identical, as shown below, when the paths are expressed no longer in terms of microscopic variables but in terms of an appropriate set of reduced variables. Furthermore, in the latter case, the path formulation is no longer redundant since the trajectory captures information about the eliminated degrees of freedom. The identification of the minimal set of variables, for which the elimination is valid, follows from the combination of Eq. (9) with the Crooks' result Eq. (3). One finds:

$$D(\mathcal{P}(\text{path})||\widetilde{\mathcal{P}}(\widetilde{\text{path}})) = D(P(W)||\widetilde{P}(-W)). \tag{10}$$

This is a surprising relationship: from the chain rule for the relative entropy, Eq. (6), one would expect that the relative entropy for the paths, which contains full information on all the microscopic variables, would be bigger than that contained in the work, which is a single scalar path-dependent variable. However, both relative entropies are exactly the same. The combination of Eqs. (9) and (10) allows us now to formulate the following main conclusions. First, it is impossible via relative entropy to overestimate the dissipation. Second, the exact dissipation is revealed by any set of variables that contains the statistical information about the work. The fact that the dissipation is underestimated if we do not have this information is also of practical interest, but will be the object of a separate paper [29].

One set of variables that captures the information on the work is now easy to identify: the information is obviously contained in the dynamic variables that are interacting with the (external) work-performing device. More precisely, the work performed along a trajectory $\Gamma(t)$, $t \in [0, \tau]$, is given by:

$$W = \int_0^\tau dt \frac{\partial H(\Gamma(t), \lambda(t))}{\partial \lambda(t)} \dot{\lambda}(t) \tag{11}$$

and can be exactly calculated from the path followed by the variables coupled to λ. Then it is enough to know the (statistical) behavior of these variables to reproduce the statistics of the work, and hence the average dissipation. Trajectory information of these and only these variables, along the whole (both forward and backward) process, is enough to account for the total average dissipation. In particular, if a stochastic model provides the exact description of a system in its interaction with an external device, one needs only the path information of these variables. Eq. (9) is thus valid for a "correct" stochastic model with the path determined in terms of the corresponding stochastic variables. As a corollary, we note that bath variables which are replaced (in some ideal limit) by a stochastic perturbation, will not appear in the "path", which is in terms of the variables of the stochastic system only.

Let us mention another surprising consequence of the above equality (10). By applying the chain rule, Eq. (6), one finds:

$$D(\mathcal{P}(\text{path}|W)||\widetilde{\mathcal{P}}(\widetilde{\text{path}}| - W)) = 0, \tag{12}$$

and hence:

$$\mathcal{P}(\text{path}|W) = \widetilde{\mathcal{P}}(\widetilde{\text{path}}| - W), \tag{13}$$

for all W. Eq. (12) means that, by selecting trajectories corresponding to a given value of work, W and $-W$, in the forward and backward process respectively, it is not possible to detect the arrow of time in them. According to Eq. (13), the sub-ensembles of these trajectories are in fact statistically indistinguishable! As an example, the snapshots of the positions of particles, during the expansion and compression of a gas, will be statistically identical, when the corresponding amounts of work are each other's opposite. For the folding or unfolding of an RNA molecule [30], the trajectories are statistically indistinguishable, again if the amounts of works are opposite. We however also note that, according to the Crooks' relation Eq. (2), the probabilities for such forward and backward trajectories will be very different if the experiment is performed in an irreversible way. If, e.g., the forward set corresponds to typical realizations, the same set of trajectories will be atypical for the backward experiment [28] (except if the overall process is reversible).

We conclude with a brief discussion about the range of applicability of the above result. Recall that, in its derivation, it is assumed that the system starts in canonical equilibrium in both forward and backward scenario, and is disconnected from the heat bath in the intermediate time. However, both the microscopic expression for dissipation given in Eq. (5) and the Crooks' equality can be derived for other equilibrium initial conditions [17] so that Eq. (9) remains valid for these other "transient" nonequilibrium scenarios linking equilibrium states. Furthermore, the formula can also be applied to nonequilibrium steady states according to the following argument. Imagine that the perturbation induces, after an initial transient, a steady state in a sub-part of the system. In the limit that the other degrees of freedom for the remainder of the system have an infinitely fast relaxation to (local) equilibrium, these will not contribute in the formula and both the time-irreversibility and dissipation will be completely captured by the steady state variables. This hand-waving argument explains why the formula (9) is also known to reproduce the correct entropy production in nonequilibrium steady state models [31,32], where ideal heat, work, and/or particle sources are responsible for the generation of the steady state.

* * *

We acknowledge financial help from the Ministerio de Educacion y Ciencia (Spain) under Grants FPU-AP-2004-0770 (A. G-M.) and MOSAICO (JMRP), and from the FWO Vlaanderen.

References

[1] EVANS D. J., COHEN E. G. D. and MORRISS G. P., *Phys. Rev. Lett.*, **71** (1993) 2401.
[2] GALLAVOTTI G. and COHEN E. G. D., *J. Stat. Phys.*, **80** (1995) 931.
[3] KURCHAN J., *J. Phys. A*, **31** (1998) 3719.
[4] LEBOWITZ J. L. and SPOHN H., *J. Stat. Phys.*, **95** (1998) 333.
[5] MAES C., *J. Stat. Phys.*, **95** (1999) 367.
[6] BOCHKOV G. N. and KUZOVLEV Y. E., *Zh. Eksp. Teor. Fiz.*, **72** (1977) 238.
[7] BOCHKOV G. N. and KUZOVLEV Y. E., *Sov. Phys. JETP*, **45** (1977) 125.
[8] BOCHKOV G. N. and KUZOVLEV Y. E., *Physica A*, **106** (1981) 443.
[9] BOCHKOV G. N. and KUZOVLEV Y. E., *Physica A*, **106** (1981) 480.
[10] JARZYNSKI C., *Phys. Rev. Lett.*, **78** (1997) 2690.
[11] JARZYNSKI C., *Phys. Rev. E*, **56** (1997) 5018.
[12] CROOKS G. E., *Phys. Rev. E*, **60** (1999) 2721.
[13] HATANO T. and SASA S. I., *Phys. Rev. Lett.*, **86** (2001) 3463.
[14] CLEUREN B., VAN DEN BROECK C. and KAWAI R., *Phys. Rev. Lett.*, **96** (2006) 050601.
[15] CLEUREN B., VAN DEN BROECK C. and KAWAI R., *C.R. Physique*, **8** (2007) 567.
[16] KAWAI R., PARRONDO J. M. R. and VAN DEN BROECK C., *Phys. Rev. Lett.*, **98** (2007) 080602.
[17] VAN DEN BROECK C., KAWAI R. and PARRONDO J. M. R., unpublished.
[18] CROOKS G. E., *J. Stat. Phys.*, **90** (1998) 1481.
[19] CROOKS G. E., Ph.D. thesis, UC Berkeley (1999).
[20] SEKIMOTO K., *Prog. Theor. Phys. Suppl.*, **130** (1998) 17.
[21] KURCHAN J., *J. Phys. A*, **31** (1998) 3719.
[22] QIAN H., *Phys. Rev. E*, **63** (2001) 042103.
[23] MAES C. and NETOCŸNY K. *J. Stat. Phys.*, **110** (2003) 269.
[24] WU D. and KOFKE D. A., *J. Chem. Phys.*, **123** (2005) 054103.
[25] SEIFERT U., *Phys. Rev. Lett.*, **95** (2005) 040602.
[26] ANDRIEUX D., GASPARD P., CILIBERTO S., GARNIER N., JOUBAUD S. and PETROSYAN A., *Phys. Rev. Lett.*, **98** (2007) 150601.
[27] COVER T. M. and THOMAS J. A., *Elements of Information Theory* 2nd edition (Wiley, Hoboken: New Jersey) 2006.
[28] JARZYNSKI C., *Phys. Rev. E*, **73** (2006) 046105.
[29] GOMEZ-MARIN A., PARRONDO J. M. R. and VAN DEN BROECK C., accepted in Phys. Rev. E.
[30] COLLIN D., RITORT F., JARZYNSKI C., SMITH S. B., TINOCO I. and BUSTAMANTE C., *Nature*, **437** (2005) 231-234.
[31] LUO J. L., VAN DEN BROECK C. and NICOLIS G., *Z. Phys. B*, **56** (1984) 165.
[32] BLYTHE R. A., *Phys. Rev. Lett.*, **100** (2008) 010601.