# Entropy production and the arrow of time

**J M R Parrondo**[1,4], **C Van den Broeck**[2] **and R Kawai**[3]

[1] Departamento de Física Atómica, Molecular y Nuclear and GISC,
Universidad Complutense de Madrid, 28040-Madrid, Spain
[2] University of Hasselt, B-3590 Diepenbeek, Belgium
[3] Department of Physics, University of Alabama at Birmingham,
Birmingham, AL 35294, USA
E-mail: parrondo@fis.ucm.es

**Abstract.** We present an exact relationship between the entropy production and the distinguishability of a process from its time-reverse, quantified by the relative entropy between forward and backward states. The relationship is shown to remain valid for a wide family of initial conditions, such as canonical, constrained canonical, multi-canonical and grand canonical distributions, as well as both for classical and quantum systems.

**Contents**

[4] Author to whom any correspondence should be addressed.

**IOP** Institute of Physics  Φ DEUTSCHE PHYSIKALISCHE GESELLSCHAFT

## 1. Introduction

Providing a microscopic expression for the entropy production has been one of the grand aims of statistical mechanics, going back to the seminal work of Boltzmann. However, both the range of validity of the second law and of its proposed derivations have, from the very beginning, generated discussion and controversy. The recent discovery of the fluctuation and work theorems [1]–[6] has re-invigorated the debate. These intriguing results provide equalities involving the probability distributions for work or entropy production, and appear to be of specific interest for the study of small systems [7]–[9]. With respect to the average value of work or entropy production, they are consistent with the second law, but provide no additional information beyond it. In a recent work [10] (see also [11]), we have derived, from first principles, the exact value of the average dissipated work $\langle W \rangle_{\text{diss}}$ upon bringing a system from one canonical equilibrium state at a temperature $T$ into another one at the same temperature. The dissipated work is defined as the extra amount of work, on top of the difference of free energy $\Delta F$ that is required for making this transition. Its expression reads

$$\langle W \rangle_{\text{diss}} = \langle W \rangle - \Delta F = kT \left\langle \ln \frac{\rho}{\tilde{\rho}} \right\rangle. \tag{1}$$

Here $k$ is the Boltzmann constant and $\rho = \rho(\Gamma; t)$ is the probability density in phase space to observe the system to be in a micro-state $\Gamma = (q, p)$ specified by a set of positions $q$ and momenta $p$ at an intermediate time $t$. The averaging brackets denote an average with respect to the density $\rho$. The other density $\tilde{\rho} = \tilde{\rho}(\tilde{\Gamma}; t)$ represents the distribution in the time-reversed process observed at a corresponding time-reversed phase point $\tilde{\Gamma} = (q, -p)$ at the same time as the forward process. As an example, consider compression and expansion of a gas in a cylinder, a forward experiment corresponding to the expansion of a gas by moving a piston from an initial to a final position; the backward experiment corresponds to performing the motion of the piston in the time-reversed manner. The statistics of the micro-states are taken at the same intermediate position of the piston. Both forward and backward experiments are assumed to start in canonical equilibrium at the same temperature.

The dissipated work (1) can also be written in terms of a well known and powerful concept from information theory, the relative entropy or Kullback–Leibler distance between two probability densities [12]:

$$\langle W \rangle_{\text{diss}} = kT D(\rho \| \tilde{\rho}), \tag{2}$$

where

$$D(\rho \| \tilde{\rho}) = \int d\Gamma \rho(\Gamma, t) \ln \frac{\rho(\Gamma, t)}{\tilde{\rho}(\tilde{\Gamma}, t)} \tag{3}$$

$$= \int d\Gamma \rho(\Gamma, t) \ln \rho(\Gamma, t) - \int d\Gamma \rho(\Gamma, t) \ln \tilde{\rho}(\tilde{\Gamma}, t). \tag{4}$$

is the relative entropy between $\rho$ and $\tilde{\rho}$. Written in this way, the result reveals its deep meaning, linking the dissipation directly to the irreversibility of the underlying experiment. Indeed the relative entropy between probability densities is expressing the difficulty of distinguishing samplings from these densities [12]. In the present case, it measures the difficulty of distinguishing whether observed data of the micro-state correspond to those from a forward or backward experiment. Therefore, the relative entropy (4) can be considered as a quantitative measure of the arrow of time [13]–[25].

At the end of the forward process, the system is generally speaking in a non-equilibrium state, for which the entropy is not well defined. Hence, to make the connection with the entropy production, we introduce an idealized heat bath at temperature $T$, with which the system is put in contact at the end of the forward experiment, without extra work. The system thus relaxes back to canonical equilibrium, which is also the starting state of the backward process. During the relaxation, the system transfers an average amount of heat $Q$ to the bath. We can now evaluate the total entropy production. On the one hand, the entropy change in the bath, assuming it operates quasi-statically, is given by $Q/T$. On the other hand, the entropy change between the canonical equilibrium state of initial and final states in the system reads: $(\Delta U - \Delta F)/T = (\langle W \rangle - Q - \Delta F)/T$. Here, we used the relation $F = U - TS$, where $U = \langle H \rangle$ is the equilibrium internal energy, calculated with respect to the prevailing canonical distribution and $S$ the thermodynamic entropy of the system. The total entropy production (in the total device, system plus heat bath) in the forward process, $\Delta S$, is thus equal to the dissipated $\langle W \rangle_{\text{diss}}$ divided by temperature, hence

$$\Delta S = k \left\langle \ln \frac{\rho}{\tilde{\rho}} \right\rangle = k D(\rho \| \tilde{\rho}). \tag{5}$$

This result goes right to the core of the second law, as it provides an exact explicit microscopic expression for the total entropy production in the considered scenario.

This scenario is very specific: the system is initially at canonical equilibrium, and disconnected from any heat bath during the perturbation [10]. These assumptions greatly simplify the derivation, but also generate the misleading impression that they are essential. The aim of this paper is to further clarify the status of the basic results (5), and to show that it has a wider range of validity.

Our discussion will be based on a general and exact mathematical identity for the relative entropy, derived in section 2, which is valid for arbitrary initial conditions in the forward process and the backward process. In section 3, this identity is applied to different scenarios, including various types of initial equilibrium states, such as grand canonical ensembles, constrained equilibrium states and multi-canonical distributions at different temperatures. In each of these cases, the entropy production is (aside a factor $k$) equal to the relative entropy of the phase space density between the forward and the backward states, provided some idealized relaxation process is assumed at the end of the forward process. In section 4, we present the quantum analogue of the above result, expressed in terms of the quantum relative entropy [26]–[28].

## 2. General formulation

We consider a system described by the Hamiltonian $H(\Gamma; \lambda)$, with $\lambda$ a control parameter that describes the energy exchange with an external device. For simplicity, we will assume that the Hamiltonian is an even function with respect to inversion of momenta. As is well known, the probability density in phase space $\rho(\Gamma, t)$ to observe the system in a specific micro-state $\Gamma$ at time $t$ obeys the Liouville equation:

$$\frac{\partial \rho(\Gamma, t)}{\partial t} = \mathcal{L}\rho(\Gamma, t). \tag{6}$$

In a classical system, the Liouville operator $\mathcal{L}$ is given by Poisson brackets: $\mathcal{L}\rho = \{H, \rho\} = \partial_q H \partial_p \rho - \partial_p H \partial_q \rho$ (summation convention over all positions and momenta is assumed).

An important implication of the Liouville theorem is that the Shannon entropy associated with $\rho$, namely

$$-\int d\Gamma \rho(\Gamma, t) \ln \rho(\Gamma, t) \tag{7}$$

is constant in time, as can be easily proven by applying the Liouville equation followed by partial integration [16]. This appears to be in contradiction with the second law of thermodynamics, or at least precludes the use of this expression as a microscopic definition for the entropy. In fact, we share the general opinion that entropy is not defined at the *microscopic level* except when the system is at equilibrium. The Shannon entropy (7) reduces to the proper thermodynamic entropy only for systems at equilibrium.

An essential remark for the further discussion is that the above property remains valid for a smooth time-dependent Hamiltonian or Liouvillian. We will henceforth focus on scenarios where the control parameter $\lambda$ is changed from an initial value $\lambda_A$ to a final value $\lambda_B$, according to a specific protocol $\lambda(t), t \in [0, \tau]$. We also consider a time-reversed experiment, starting at the final value $\lambda_B$ of the control parameter, and performing the time-reversed perturbation. The quantities appearing in this setting will be indicated by a superscript tilde. We will always use the forward time $t$ to designate time in both forward and backward experiments, keeping in mind that the corresponding real time in the backward experiment is equal to $\tau - t$. The initial conditions for both forward and backward experiments are for the moment left unspecified. The above set-up is reminiscent of the one introduced by Jarzynski in his derivation of the Jarzynski equality [4], although the scope of our derivation extends, as we will see, well beyond the validity of the Jarzynski equality.

The key observation is that the phase space density of backward evolution with reversed momenta, namely $\tilde{\rho}(\tilde{\Gamma}, t)$, obeys the same Liouville equation (6) with respect to the forward time variable $t$:

$$\frac{\partial \tilde{\rho}(\tilde{\Gamma}, t)}{\partial t} = \mathcal{L}\tilde{\rho}(\tilde{\Gamma}, t). \tag{8}$$

The transition from equation (6) to (8) is based on micro-reversibility: trajectories retrace their steps upon reversing the schedule of $\lambda$ and inverting momenta. Mathematically, the result follows easily by writing the Liouville equation for $\tilde{\rho}$ in its proper time $\tau - t$. When switching to the forward time $t$, the negative sign, due to the time derivative in the lhs of the Liouville equation, is cancelled by another one in the rhs, appearing upon inverting the momenta. As a particular example, for the case of a time-independent Hamiltonian, the above statement reduces to the fact that $\rho(\Gamma, t)$ and $\rho(\tilde{\Gamma}, \tau - t)$, obey the same Liouville equation, when the time-derivative is with respect to $t$.

We can now invoke a remarkable property for probability distributions that obey the same Liouville equation: their relative entropy is invariant in time [16, 29]. In particular, the following quantity

$$\int d\Gamma \rho(\Gamma, t) \ln \tilde{\rho}(\tilde{\Gamma}, t) \tag{9}$$

is time-invariant, as can again be easily verified by applying Liouville's equation followed by partial integration.

The invariance in time of (7) and (9) now allows us to rewrite the quantity of central interest, namely the relative entropy (4), as follows. Since each term in the rhs of equation (4)

is invariant under time translation, we can shift time in the first term back to $t = 0$, and in the second term forward to $t = \tau$:

$$D\left(\rho(\Gamma, t) \| \tilde{\rho}(\tilde{\Gamma}, t)\right) = \int d\Gamma \rho(\Gamma, 0) \ln \rho(\Gamma, 0) - \int d\Gamma \rho(\Gamma, \tau) \ln \tilde{\rho}(\tilde{\Gamma}, \tau). \tag{10}$$

We stress again that this is an exact result, valid for any initial conditions for $\rho$ and $\tilde{\rho}$. For specific choices of these states, we proceed to make the connection with the entropy production in a variety of scenarios. Note also that the above result will only be useful if the support of $\tilde{\rho}$ includes the support of $\rho$, i.e. there are no phase points for which $\tilde{\rho} = 0$ and $\rho \neq 0$. Otherwise $D(\rho \| \tilde{\rho})$ diverges.

## 3. Explicit entropy production in different scenarios

In this section, we apply equation (10) for specific choices of equilibrium initial conditions $\rho(\Gamma, 0)$ and $\tilde{\rho}(\tilde{\Gamma}, \tau)$ in forward and backward processes, respectively. We will see that in all the considered cases, the resulting relative entropy is equal to the entropy production along the forward process. If $\rho(\Gamma, 0)$ is an equilibrium distribution or the product of independent equilibrium distributions, Shannon entropy is equal to the thermodynamic entropy:

$$S(0) = -k \int d\Gamma \rho(\Gamma, 0) \ln \rho(\Gamma, 0). \tag{11}$$

This equality holds for different equilibrium ensembles, such as canonical and grand canonical. With this choice of initial condition for the forward process the first term in equation (10) is minus the initial equilibrium entropy. In the following, we will prove that the second term can be identified with the final entropy.

As mentioned in previous sections, in this paper we only consider as meaningful the entropy of equilibrium states. On the other hand, after the forward process is completed at time $\tau$, the resulting state $\rho(\Gamma, \tau)$ is not at equilibrium. In order to have a well-defined expression for the entropy production, we need some relaxation mechanism to equilibrium *after* the forward process has been completed. The function of this mechanism is twofold: it allows us to have a meaningful definition of entropy production, and also drives the system from the non-equilibrium final state $\rho(\Gamma, \tau)$ to the equilibrium one $\tilde{\rho}(\Gamma, \tau)$, which is the initial condition for the backward process[5].

Note that the relaxation from $\rho(\Gamma, \tau)$ to $\tilde{\rho}(\Gamma, \tau)$ in an isolated system is incompatible with the Liouville theorem. Our derivation does not touch on this old and unresolved problem. Instead, for the relaxation to $\tilde{\rho}(\Gamma, \tau)$, we invoke the presence of one or several ideal (super)baths, to which the system is weakly coupled. This relaxation can, in principle, be performed without any extra energy input, since coupling the superbath with just one degree of freedom of the system is enough to induce thermalization [30]. Moreover, if this degree of freedom is chosen to have zero mean (for instance, one of the momenta of the bath), then the average work done by switching on the coupling is strictly zero. Finally, we will assume that the entropy production in the superbaths, due to exchange of energy and/or particles with the system, is given by the standard expressions from statistical mechanics and thermodynamics.

---

[5] This relaxation will take some time $\tau_{\rm rel}$. However, recall that our time variable $t$ in $\tilde{\rho}(\Gamma, t)$ denotes the stage of the process, using the forward one as reference, rather than the real physical time $t_{\rm phys}$. In a complete forward–backward cycle, the time variable in $\tilde{\rho}(\Gamma, t)$ will be $t = 2\tau + \tau_{\rm rel} - t_{\rm phys}$, so $\tilde{\rho}(\Gamma, \tau)$ is the state of the system at $t_{\rm phys} = \tau + \tau_{\rm rel}$.

**IOP** Institute of Physics $\Phi$ DEUTSCHE PHYSIKALISCHE GESELLSCHAFT

Note that such or similar assumptions also appear in the various derivations of both fluctuation and work theorems, and in the study of non-equilibrium steady states. They are also intrinsically present in the formulations via mesoscopic stochastic descriptions.

### 3.1. Canonical distributions

In our first example, we start the forward process with a canonical distribution at temperature $T$ and, after the process is completed at time $\tau$, we connect the system with a thermal bath at a different temperature $T'$. The initial condition for the backward process is in this case a canonical distribution at temperature $T'$:

$$\rho(\Gamma, 0) = \frac{e^{-\beta H(\Gamma; \lambda_A)}}{Z(T, \lambda_A)}, \tag{12}$$

$$\tilde{\rho}(\Gamma, \tau) = \frac{e^{-\beta' H(\Gamma; \lambda_B)}}{Z(T', \lambda_B)}, \tag{13}$$

where $Z(T, \lambda) = \int d\Gamma e^{-\beta H(\Gamma, \lambda)}$ being the partition function. Recalling that we consider Hamiltonians even functions of the momenta, we obtain from equations (10) and (11)

$$k D(\rho || \tilde{\rho}) = -S(0) + \frac{\langle H \rangle_\tau - F(T', \lambda_B)}{T'}, \tag{14}$$

where averages $\langle \cdot \rangle_t$ are taken over the forward time density $\rho(\Gamma, \tau)$ and

$$F(T, \lambda) = -kT \ln Z(T, \lambda) \tag{15}$$

is the usual equilibrium free energy. After relaxation to equilibrium with the bath at temperature $T'$, the average energy of the system becomes $\langle H \rangle_{eq,\tau}$, which is only a function of $\lambda_B$ and $T'$. In the relaxation, the system transfers an amount of energy to the thermal bath $Q = \langle H \rangle_\tau - \langle H \rangle_{eq,\tau}$. Therefore

$$\begin{aligned} k D(\rho || \tilde{\rho}) =& -S(0) + \frac{\langle H \rangle_{eq,\tau} - F(T', \lambda_B)}{T'} + \frac{Q}{T'} \\ =& -S(0) + S(\tau) + \Delta S_{bath} = \Delta S, \end{aligned} \tag{16}$$

where $S(\tau)$ is the final entropy of the system (in equilibrium at temperature $T'$) and $\Delta S_{bath}$ is the increase of the entropy in the bath.

Therefore, the relative entropy is equal to the total entropy production along the forward process plus the final relaxation: $\Delta S = k D(\rho || \tilde{\rho})$. Note that *all* entropies have been calculated only for equilibrium states. In fact, this example illustrates that it makes no sense to talk about the entropy produced along the process if one does not specify the final equilibrium state and how it is reached. In our case, the total entropy production depends on the final temperature $T'$. Hence, the same forward process $\lambda(t)$ can have different entropy production depending on the final temperature $T'$.

### 3.2. General equilibrium distribution

The above example can be easily generalized to an arbitrary equilibrium distribution defined over some conserved quantities $A_i(\Gamma)$ ($i = 1, \ldots, r$), with conjugated variables $\alpha_i = T \partial S / \partial A_i$ and the corresponding thermodynamic potential $\mathcal{U}(T, \vec{\alpha}) = \langle H \rangle + \sum_i \alpha_i \langle A_i \rangle - T S$. After the

forward process, the system is coupled to a reservoir characterized by temperature $T'$ and certain values $\alpha_i'$ of the conjugated variables. The initial condition for the backward process reads

$$\tilde{\rho}(\Gamma, \tau) = \exp\left[-\beta'\left(H(\Gamma, \lambda_B) + \sum_{i=1}^{r} \alpha_i' A_i(\Gamma) - \mathcal{U}(T', \vec{\alpha}')\right)\right]. \tag{17}$$

Hence

$$-k \int d\Gamma \rho(\Gamma, \tau) \ln \tilde{\rho}(\Gamma, \tau) = \frac{\langle H \rangle_\tau + \sum_{i=1}^{r} \alpha_i' \langle A_i \rangle_\tau - \mathcal{U}(T', \vec{\alpha}')}{T'}. \tag{18}$$

During relaxation, there is a transfer of energy and of the conserved quantities $A_i$ from the system to the reservoir given, respectively, by $Q = \langle H \rangle_\tau - \langle H \rangle_{\text{eq},\tau}$ and $\Delta A_i = \langle A_i \rangle_\tau - \langle A_i \rangle_{\text{eq},\tau}$. The final entropy of the system is

$$S(\tau) = \frac{\langle H \rangle_{\text{eq},\tau} + \sum_{i=1}^{r} \alpha_i' \langle A_i(\Gamma) \rangle_{\text{eq},\tau} - \mathcal{U}(T', \vec{\alpha}')}{T'} \tag{19}$$

and the entropy increase in the reservoir is $\Delta S_{\text{bath}} = (Q + \sum_i \alpha_i' \Delta A_i)/T'$. The cross term (18) is then equal to $S(\tau) + \Delta S_{\text{bath}}$. If, along the forward process, the external agent moving $\lambda(t)$ does not undergo any change of entropy, we again find equation (5): $\Delta S = kD(\rho||\tilde{\rho})$.

Note that the cross entropy (18) contains precisely the non-equilibrium average values, $\langle H \rangle_\tau$ and $\langle A_i(\Gamma) \rangle_\tau$, yielding the correct expression for the final equilibrium entropy of the system $S(\tau)$ *plus* the entropy increase in the bath $\Delta S_{\text{bath}}$.

As a particular case, consider $A = N$ the number of particles. Then, $-\alpha = \mu$, the chemical potential. In equilibrium, the thermodynamic potential

$$\mathcal{U}(T, \mu) = \langle H \rangle_{\text{eq}} - \mu \langle N \rangle_{\text{eq}} - TS \tag{20}$$

is the so-called grand canonical potential, where the averages $\langle \cdot \rangle_{\text{eq}}$ are taken with respect to the grand canonical ensemble

$$\rho(\Gamma, N) = e^{-\beta[H(\Gamma, \lambda) - \mu N - \mathcal{U}(T, \mu)]}. \tag{21}$$

We now choose as initial condition for the forward process a grand canonical ensemble characterized by temperature $T$ and chemical potential $\mu$. We proceed with the forward process by changing $\lambda(t)$ ($N$ remaining constant to avoid change of entropy in the external agent moving $\lambda$). After the process is completed we connect the system with a particle and energy reservoir at temperature $T'$ and chemical potential $\mu'$. In the relaxation, there will be an exchange of energy and particles between the system and the bath, as explained above. Finally, the total entropy production along the forward process plus the relaxation is given by the relative entropy between forward and backward states, equation (5).

### 3.3. Multicanonical distribution

The above examples can be extended to the product of several equilibrium states. We consider the special case where systems at different temperatures are brought into contact with each other. This case is of particular interest since thermal exchange is another standard example of entropy production. Also, the present scenario produces in the appropriate limit, e.g., the limit of infinitely large thermal reservoirs, a stationary non-equilibrium state. We will thus show that our formula of entropy production as a relative phase space entropy also applies in this example of a non-equilibrium steady state.

Consider $n$ systems $i = 1, \ldots, n$, at the initial time decoupled, and each one at canonical equilibrium at temperature $T_i$:

$$\rho(\Gamma, 0) = \prod_{i=1}^{n} \frac{\mathrm{e}^{-\beta_i H_i(\Gamma_i, \lambda_{\mathrm{A}})}}{Z_i(T_i, \lambda_{\mathrm{A}})}. \tag{22}$$

We will assume a similar initial state for the backward process, decoupled systems but with arbitrary temperatures $T_i'$, i.e.

$$\tilde{\rho}(\Gamma, \tau) = \prod_{i=1}^{n} \frac{\mathrm{e}^{-\beta_i' H_i(\Gamma_i, \lambda_{\mathrm{B}})}}{Z_i(T_i', \lambda_{\mathrm{B}})}. \tag{23}$$

While the systems are decoupled both at initial and final times, thermal contact with each other can be established during intermediate stages of the process by an appropriate choice of the time-dependent Hamiltonian. One possibility is adding to the original decoupled Hamiltonians a coupling term inducing energy exchange and multiplied by the control parameter $\lambda(t)$, set equal to zero at initial and final times. Equation (10) gives us the following result:

$$kD(\rho\|\tilde{\rho}) = \sum_{i=1}^{n} \left[ -\frac{\langle H_i(\lambda_{\mathrm{A}})\rangle_0}{T_i} - k \ln Z_i(T_i, \lambda_{\mathrm{A}}) \right]$$

$$+ \sum_{i=1}^{n} \left[ \frac{\langle H_i(\lambda_{\mathrm{B}})\rangle_\tau}{T_i'} + k \ln Z_i(T', \lambda_{\mathrm{B}}) \right]. \tag{24}$$

As in our previous examples, each subsystem transfers to their corresponding thermal baths an energy $Q_i = \langle H_i(\lambda_{\mathrm{B}})\rangle_\tau - \langle H_i(\lambda_{\mathrm{B}})\rangle_{\mathrm{eq},\tau}$, $\langle H_i(\lambda_{\mathrm{B}})\rangle_{\mathrm{eq},\tau}$ being the equilibrium energy of each subsystem after the relaxation to $\tilde{\rho}(\Gamma, \tau)$. Therefore

$$kD(\rho\|\tilde{\rho}) = \sum_{i=1}^{n} \left[ -\frac{\langle H_i(\lambda_{\mathrm{A}})\rangle_0 - F(T_i, \lambda_{\mathrm{A}})}{T_i} + \frac{\langle H_i(\lambda_{\mathrm{B}})\rangle_{\mathrm{eq},\tau} - F(T_i', \lambda_{\mathrm{B}}) + Q_i}{T_i'} \right]$$

$$= \sum_{i=1}^{n} \left[ -S_i(0) + S_i(\tau) + \Delta S_{\mathrm{bath},i} \right] = \Delta S. \tag{25}$$

Hence, the total entropy produced in the entire forward process (plus relaxation) is again given by the relative entropy, equation (5). Similar arguments can be applied to multi-grand canonical distributions, allowing chemical reactions and exchange of particles between the system and the final reservoirs.

### 3.4. Constrained canonical distributions and filters

Finally, we now apply the general result, equation (10), to the case of canonical states restricted to subdomains of the phase space. One can imagine two scenarios leading to such restricted initial states, namely: (a) the system is subject to constraints in the initial and/or final states, or (b) one selects (or restricts the observation to) that specific set of trajectories that lie within a prescribed subdomain of phase space; this selection can, in principle, be performed at any intermediate time of the process.

**IOP** Institute of Physics ⬧ DEUTSCHE PHYSIKALISCHE GESELLSCHAFT

In both cases, the relevant initial condition for the forward process can be written as follows:

$$\rho(\Gamma, 0) = \frac{e^{-\beta H(\Gamma; \lambda_A)}}{Z(\beta, \lambda_A; D_A)} \chi_{D_A}(\Gamma), \tag{26}$$

where the characteristic function is defined as

$$\chi_D(\Gamma) = \begin{cases} 1, & \Gamma \in D, \\ 0, & \text{otherwise}, \end{cases} \tag{27}$$

for any subset $D$ in the phase space. The partition function is then defined as $Z(\beta, \lambda; D) = \int_D d\Gamma e^{-\beta H(\Gamma; \lambda)}$.

As initial condition for the backward process we consider

$$\tilde{\rho}(\Gamma, \tau) = \frac{e^{-\beta H(\Gamma; \lambda_B)}}{Z(\beta, \lambda_B; D_B)} \chi_{D_B}(\Gamma). \tag{28}$$

The prescriptions for $D_A$ and $D_B$ are as follows. For case (a), the subsets $D_A$ and $D_B$ reproduce the imposed initial constraints for the forward and backward processes, respectively. For case (b), the set of filtered forward trajectories equals $D_A$ at the initial time $t = 0$ and $D_B$ at the final time $\tau$. For simplicity, we will assume that $D_B$ is invariant under inversion of momenta $p \rightarrow -p$.

The relative entropy (10) now reads

$$kT\, D(\rho \| \tilde{\rho}) = \langle H(\Gamma; \lambda_B) \rangle_\tau - \langle H(\Gamma; \lambda_A) \rangle_0 - kT \ln \frac{Z(\beta, \lambda_A; D_A)}{Z(\beta, \lambda_B; D_B)}$$

$$= \langle W \rangle - kT \ln \frac{Z(\beta, \lambda_A; D_A)}{Z(\beta, \lambda_B; D_B)}. \tag{29}$$

There are two possible interpretations of this result. If the restriction of microstates to $D_A$ and $D_B$ is due to a filter, as in scenario (b), the initial and final free energies must be calculated over the whole phase space. In this case, we have

$$kT\, D(\rho \| \tilde{\rho}) = \langle W \rangle - \Delta F - kT \ln \left[ \frac{Z(\beta, \lambda_A; D_A) Z(\beta, \lambda_B)}{Z(\beta, \lambda_A) Z(\beta, \lambda_B; D_B)} \right]. \tag{30}$$

The ratios between the restricted and the non-restricted partition functions are precisely the probability that an arbitrary trajectory passes the filter in the forward ($p_{D_A}$) and backward processes ($\tilde{p}_{D_B}$), respectively. Therefore, we obtain the following exact equality:

$$kT\, D(\rho \| \tilde{\rho}) + kT \ln \frac{p_{D_A}}{\tilde{p}_{D_B}} = \langle W \rangle - \Delta F. \tag{31}$$

In particular, the non-negativeness of the relative entropy yields the following inequality:

$$\langle W \rangle - \Delta F \geqslant kT \ln \frac{p_{D_A}}{\tilde{p}_{D_B}}, \tag{32}$$

which was derived in [10] using the Liouville theorem and applied to a Brownian version of the Szilard engine and the restore-to-zero process introduced by Landauer.

On the other hand, if the restriction is due to constraints in the system at $t = 0$ and $\tau$, the logarithm of the restricted partition functions can be considered as the actual initial and final free energies. Hence, we recover the same result as in the canonical case:

$$kT\, D(\rho \| \tilde{\rho}) = \langle W \rangle - \Delta F = \langle W \rangle_{\text{diss}}. \tag{33}$$

**IOP** Institute of Physics  **Φ** DEUTSCHE PHYSIKALISCHE GESELLSCHAFT

It is worthwhile to illustrate the last case on a concrete example, especially since the Jarzynksi and Crooks equalities are failing for this kind of scenario [31]. We consider the case of free expansion of a gas from a volume $V/2$ to a final volume $V$. The Hamiltonian in this case is time-independent, but we can still apply our theory to this process consisting of a free relaxation from a restricted initial condition. In the forward process, the gas is initially in canonical equilibrium, but confined to one-half of the container, with vacuum in the other half. It is subsequently allowed to expand and fill the entire container. Since the Hamiltonian is time independent, the backward process starts with the gas at equilibrium filling the whole volume $V$ of the container, and obviously remaining so for all times. We thus have

$$\rho(\Gamma, 0) = \frac{e^{-\beta H}}{Z(\beta; V/2)} \chi_{V/2}(\Gamma),$$

$$\tilde{\rho}(\tilde{\Gamma}, \tau) = \frac{e^{-\beta H}}{Z(\beta; V)} \chi_V(\Gamma),$$

where the characteristic function $\chi_V(\Gamma)$ is 1 for microstates with all the particles in the volume $V$ and 0 otherwise. Now the relative entropy is

$$kT\, D(\rho\|\tilde{\rho}) = \langle H \rangle_\tau - \langle H \rangle_0 - kT \ln \frac{Z(\beta; V/2)}{Z(\beta; V)} = -\Delta F, \tag{34}$$

since the energy is conserved along the evolution ($H$ has no explicit time-dependence). There is no work performed in this experiment, $\langle W \rangle = 0$, and thus $-\Delta F$ is the dissipative work (the work that we waste compared to a reversible expansion). Therefore, equations (1) and (5) remain valid, even though the Jarzynski and Crooks relations no longer hold.

## 4. Quantum version

Just as the Jarzynski equality has been shown to remain valid for quantum processes [32]–[42], we now show that our approach can easily be extended to the quantum. Time evolution in quantum mechanics is ruled by unitary transformations. In particular, the density operator obeys the following quantum version of the Liouville equation:

$$\frac{\partial}{\partial t} \rho = \mathcal{L}\rho = -\frac{i}{\hbar}[H, \rho]. \tag{35}$$

Under this unitary evolution, the Von Neumann entropy $S_N(\rho) = -k\mathrm{Tr}(\rho \ln \rho)$ is preserved. For the time-reversed process, the time-reversed density operator $\tilde{\rho} = \Theta \rho \Theta^{-1}$ satisfies the same quantum Liouville equation if the forward time $t$ is used (the actual time is $\tau - t$ in the same sense as in the classical case). Here, $\Theta$ is the anti-linear time-reversal operator and a fundamental symmetry of the Hamiltonian $\Theta H(t) = H(t)\Theta$ is assumed [45]. Noting that $\Theta p \Theta^{-1} = -p$, the commutation of $H$ and $\Theta$ is, in the present context, equivalent to the assumption that the classical Hamiltonian is symmetric with respect to the inversion of momenta. Then, we can prove that $\mathrm{Tr}(\rho \ln \tilde{\rho})$ is also time invariant. In general, both von Neumann entropy and the quantum relative entropy are invariant under any unitary transformation [28]. Hence, each term in the quantum version of the relative entropy [26]

$$D_Q(\rho(t)\|\tilde{\rho}(t)) = \mathrm{Tr}(\rho(t) \ln \rho(t)) - \mathrm{Tr}(\rho(t) \ln \tilde{\rho}(t)) \tag{36}$$

　　　　　**IOP** Institute of Physics　Ⓓ DEUTSCHE PHYSIKALISCHE GESELLSCHAFT

can again be shifted to the initial and final time, respectively:

$$D_Q(\rho\|\tilde{\rho}) = \mathrm{Tr}[\rho(0)\ln\rho(0)] - \mathrm{Tr}[\rho(\tau)\ln\tilde{\rho}(\tau)], \tag{37}$$

which is analogous to the classical result (10).

By using the appropriate choices of initial and final equilibrium density matrices, one can easily derive the quantum versions of all the aforementioned classical results. As an illustration, we present the quantum version of equation (1). Substituting canonical equilibrium density operators

$$\rho(0) = \frac{\mathrm{e}^{-\beta H(\lambda_A)}}{Z(T,\lambda_A)}, \quad \tilde{\rho}(\tau) = \frac{\mathrm{e}^{-\beta H(\lambda_B)}}{Z(T,\lambda_B)}, \tag{38}$$

where the partition function is $Z(T,\lambda) = \mathrm{Tr}\,\mathrm{e}^{-\beta H(\lambda)}$ into equation (37), we again obtain

$$kT\,D_Q(\rho\|\tilde{\rho}) = \langle W\rangle - \Delta F = \langle W\rangle_{\mathrm{diss}}. \tag{39}$$


## 5. Measuring the relative entropy

While our general formula for entropy production is exact and may be useful as the starting point for other theoretical investigations, the experimental or even numerical measurement of the phase space density may practically not be feasible. As shown in [10], one can still have good estimate of the relative entropy through the measurement of a limited number of physical quantities. With this partial information, it is clear that the distinction between the forward and backward processes becomes harder. Mathematically, this is tantamount to saying that the relative entropy decreases when only partial information of the system is used:

$$\Delta S = kD(\rho\|\tilde{\rho}) \geqslant kD(p\|\tilde{p}), \tag{40}$$

where $p$ and $\tilde{p}$ are the probability distributions of the quantities used to describe the system in the forward and backward processes, respectively. We have checked the accuracy of this inequality in different situations [10, 24, 43, 44]. It is remarkable that, in the canonical scenario with $T = T'$, the variables coupled with $\lambda(t)$ are enough to saturate the inequality [24].

The quantum case is more involved. One can lose information about the state of the system in different ways: by performing specific measurements or by tracing out the degrees of freedom of a subsystem, such as a thermal bath. In any of these two cases, we can derive an inequality similar to (40).

Consider first a measurement of the observable $\hat{\Omega} = \sum_i \omega_i \mathcal{P}_i$, where the projection operator satisfies the closure relation: $\sum_i \mathcal{P}_i = I$. The probability to obtain $\omega_i$ is given by $p_i = \mathrm{Tr}(\rho\mathcal{P}_i)$ and $\tilde{p}_i = \mathrm{Tr}(\tilde{\rho}\mathcal{P}_i)$ for the forward and backward processes, respectively. Then, from the joint convexity of the relative entropy [28], we have

$$\Delta S = kD_Q(\rho\|\tilde{\rho}) \geqslant kD(p\|\tilde{p}), \tag{41}$$

where the discrete version of classical relative entropy is defined as

$$D(p\|\tilde{p}) = \sum_i p_i \ln\frac{p_i}{\tilde{p}_i}. \tag{42}$$

It should be stressed that the corresponding von Neumann entropy shows the opposite inequality, namely,

$$S(p) \geqslant S_N(\rho). \tag{43}$$

**IOP** Institute of Physics $\Phi$ DEUTSCHE PHYSIKALISCHE GESELLSCHAFT

Since some information is lost, the Shannon information entropy increases. However, this result should not be interpreted as the increase of thermodynamic entropy due to the quantum measurement since the quantum states after measurement are not in thermal equilibrium and the Shannon entropy does not correspond to thermodynamic entropy. The inequality merely indicates that an estimate of von Neumann entropy using the probability distribution of measured quantities provides an upper bound.

Next we consider compound systems. For simplicity, we consider a bipartite system consisting of two subsystems $a$ and $b$. The total Hilbert space is the product space $\mathcal{H}_{a+b} = \mathcal{H}_a \otimes \mathcal{H}_b$. If we are interested in the state of the subsystem $a$, the other subsystem can be traced out to obtain a reduced density operator $\rho_a = \mathrm{Tr}_b \, \rho_{a+b}$. The monotonicity of the relative entropy states [26]:

$$\Delta S = k D_{\mathrm{Q}}(\rho_{a+b} \| \tilde{\rho}_{a+b}) \geqslant k D_{\mathrm{Q}}(\rho_a \| \tilde{\rho}_a). \tag{44}$$

Since we lose some information about the system, again we cannot find the exact amount of the dissipation. However, the relative entropy of reduced densities provides its lower bound.

While the quantum version of the relative entropy shows the same properties as the classical version, it is, again, worth mentioning that the von Neumann entropy satisfies an unusual inequality:

$$|S_{\mathrm{N}}(\rho_a) - S_{\mathrm{N}}(\rho_b)| \leqslant S_{\mathrm{N}}(\rho_{a+b}) \leqslant S_{\mathrm{N}}(\rho_a) + S_{\mathrm{N}}(\rho_b). \tag{45}$$

This property of sub-additivity is quite remarkable since, unlike the classical Shannon entropy, the entropy of a subsystem can be larger than that of the whole system. This observation plays an important role in quantum information theory. Furthermore, the role of such specific quantum features on the foundations of statistical mechanics is the object of an ongoing debate [46]–[49].

## 6. Conclusions

We have shown that the relative entropy or Kullback–Leibler distance between the forward and backward state is equal to the entropy production along the forward processes in different scenarios, both classical and quantum. Stein's lemma [12] gives a precise meaning to the relative entropy $D(\rho_{\mathrm{A}} \| \rho_{\mathrm{B}})$ between two distributions $\rho_{\mathrm{A}}$ and $\rho_{\mathrm{B}}$: if $n$ data from $\rho_{\mathrm{B}}$ are given, the probability of guessing incorrectly that the data come from $\rho_{\mathrm{A}}$ is bounded by $2^{-nD(\rho_{\mathrm{A}} \| \rho_{\mathrm{B}})}$, for $n$ large. Therefore, the relative entropy is a quantitative measure of the distinguishability between the two distributions. In the case of the forward $\rho(t)$ and backward $\tilde{\rho}(t)$ states of a system, the relative entropy $D(\rho(t) \| \tilde{\rho}(t))$ measures the distinguishability of the direction of time. Therefore, our main result $\Delta S = k D(\rho(t) \| \tilde{\rho}(t))$, is a quantitative relationship between entropy production and the arrow of time.

We have proven this identity in a number of scenarios, by choosing the appropriate initial conditions for the forward and the backward process. The identity is valid whenever $\rho(0)$ and $\tilde{\rho}(\tau)$ are given by equilibrium states or factorized equilibrium states (such as in the multicanonical case).

From a fundamental point of view, one can of course argue, as we already mentioned in the introduction that we have merely shifted the issues of relaxation and entropy production to the purported properties of the bath. While this is, strictly speaking, indeed the case, the resulting expression of the entropy production in terms of relative entropy, equation (5), incorporates two fundamental properties, namely time-translational invariance (Liouville equation) and time-reversibility of the laws of physics. Hence, our explicit expression for the entropy production

quantified as the statistical time asymmetry, is very much in the spirit of Onsager's work on the symmetry of the Onsager coefficients, resulting from the same ingredients, Liouville's theorem and micro-reversibility.

Finally, in the quantum case, while the extension of our results to quantum processes looks straightforward, at least mathematically, there are intriguing fundamental questions. For example, how do decoherence, quantum measurement and quantum entanglement contribute to the fluctuation theorems and ultimately to the second law? On the other hand, the von Neumann entropy and the quantum relative entropy can detect such processes through the change in the density operators. In fact, these entropies are used to quantify the degree of entanglement in quantum information theory [27, 28]. Therefore, our novel approach for irreversibility in quantum processes allows one to investigate quantum non-equilibrium thermodynamics involving entanglement and decoherence from a novel perspective.

## Acknowledgments

## References

[1] Bochkov G N and Kuzovlev Y E 1981 *Physica* A **106** 443
     Bochkov G N and Kuzovlev Y E 1981 *Physica* A **106** 480
[2] Evans D, Cohen E G D and Morris G P 1993 *Phys. Rev. Lett.* **71** 2401
[3] Gallavotti G and Cohen E G D 1995 *Phys. Rev. Lett.* **74** 2694
[4] Jarzynski C 1997 *Phys. Rev. Lett.* **78** 2690
     Jarzynski C 1997 *Phys. Rev.* E **56** 5018
[5] Crooks G E 1998 *J. Stat. Phys.* **90** 1481
     Crooks G E 1999 *Phys. Rev.* E **60** 2721
[6] Derrida B, Gaspard P and Van den Broeck C (ed) 2007 Work, dissipation and fluctuations in nonequilibrium physics *C.R. Physique* **8**
[7] Ritort F 2003 *Semin. Poincare* **2** 195
[8] Liphardt J, Dumont S, Smith S B, Tinoco I Jr and Bustamante C 2002 *Science* **296** 1832
[9] Collin D, Ritort F, Jarzynski C, Smith S B, Tinoco I Jr and Bustamante C 2005 *Nature* **437** 231
[10] Kawai R, Parrondo J M R and Van den Broeck C 2007 *Phys. Rev. Lett.* **98** 080602
[11] Jarzynski C 2006 *Phys. Rev.* E **73** 046105
[12] Cover T M and Thomas J A 2006 *Elements of Information Theory* 2nd edn (Hoboken, NJ: Wiley)
[13] Schnakenberg J 1976 *Rev. Mod. Phys.* **48** 571
[14] Pomeau Y 1982 *J. Physique* **43** 859
[15] Luo J L, Van den Broeck C and Nicolis G 1984 *Z. Phys.* B **56** 165
[16] Mackey M C 1989 *Rev. Mod. Phys.* **61** 981
     Mackey M C 1992 *Time's Arrow: The Origins of Thermodynamic Behavior* (New York: Springer)
[17] Qian H 2001 *Phys. Rev.* E **63** 042103
[18] Jiang D Q, Qian M and Zhang F-X 2003 *J. Math. Phys.* **44** 4176
[19] Maes C and Netocyny K 2003 *J. Stat. Phys.* **110** 269
[20] Gaspard P 2004 *J. Stat. Phys.* **117** 599
[21] Costa M, Goldberg A and Peng C-K 2005 *Phys. Rev. Lett.* **95** 198102
[22] Seifert U 2005 *Phys. Rev. Lett.* **95** 040602
[23] Porporato A, Rigby J R and Daly E 2007 *Phys. Rev. Lett.* **98** 094101

**IOP** Institute of Physics  ⦿ DEUTSCHE PHYSIKALISCHE GESELLSCHAFT

[24] Gomez-Marin A, Parrondo J M R and Van den Broeck C 2008 *Europhys. Lett.* **82** 50002

[25] Blythe R A 2008 *Phys. Rev. Lett.* **100** 010601

[26] Vedral V 2002 *Rev. Mod. Phys.* **74** 197

[27] Vedral V 2006 *Introduction to Quantum Information Science* (Oxford: Oxford University Press)

[28] Petz D 2008 *Quantum Information Theory and Quantum Statistics* (Berlin: Springer)

[29] Plastino A R and Daffershofer A 2004 *Phys. Rev. Lett.* **93** 138701

[30] Parrondo J M R and Español P 1996 *Am. J. Phys.* **64** 1125

[31] Gross D H E 2005 arXiv:cond-mat/0508721
Jarzynski C 2005 arXiv:cond-mat/0509344
Bena I, Van den Broeck C and Kawai R 2005 *Europhys. Lett.* **71** 879

[32] Tasaki H 2000 arXiv:cond-mat/0009244

[33] Kurchan J 2000 arXiv:cond-mat/0007360

[34] Yukawai S 2000 *Japan. J. Phys. Soc.* **69** 2367

[35] Mukamel S 2003 *Phys. Rev. Lett.* **90** 170604

[36] Chernyak V and Mukamel S 2004 *Phys. Rev. Lett.* **93** 048302

[37] De Roeck W and Maes C 2004 *Phys. Rev.* E **69** 026115

[38] Monnai T 2005 *Phys. Rev.* E **72** 027102

[39] Esposito M and Mukamel S 2006 *Phys. Rev.* E **73** 046129

[40] Engel A and Nolte R 2007 *Europhys. Lett.* **79** 10003

[41] Talkner P, Lutz E and Hanggi P 2007 *Phys. Rev.* E **75** 050102

[42] Andrieux D and Gaspard P 2008 *Phys. Rev. Lett.* **100** 230404

[43] Gomez-Marin A, Parrondo J M R and Van den Broeck C 2008 *Phys. Rev.* E **78** 011107

[44] Horowitz J and Jarzynski C 2009 *Phys. Rev.* E **79** 021106

[45] Sisterna P D 2000 *Found. Phys. Lett.* **13** 205

[46] Germmer J, Michel M and Mahler G 2004 *Quantum Thermodynamics* (Berlin: Springer)

[47] Popescu S, Short A J and Winter A 2006 *Nat. Phys.* **2** 758

[48] Goldstein S, Lebowitz J L, Tumulka R and Zanghi N 2006 *Phys. Rev. Lett.* **96** 050403

[49] Reimann P 2008 *Phys. Rev. Lett.* **101** 190403